國立臺灣大學海洋研究所

碩士論文

Institute of Oceanography

National Taiwan University

Master Thesis

# 南海北部不同深度的菌相與多樣性研究

# The diversity and compositions of stratified microbial assemblages in the interior of the northern South China Sea

賴虹君

Hung-Chun Lai

指導教授：謝文陽 博士、湯森林 博士

Advisor: Wung-Yang Shieh, Ph. D., Sen-Lin Tang, Ph. D.

中華民國 99 年 7 月

July, 2010

# 國立臺灣大學碩士學位論文
# 口試委員會審定書

## 南海北部不同深度的菌相與多樣性研究
## The diversity and compositions of stratified microbial assemblages in the interior of the northern South China Sea

本論文係 賴虹君 君（R97241208）在國立臺灣大學海洋研究所完成之碩士學位論文，於民國 九十九 年 七 月 二 日承下列考試委員審查通過及口試及格，特此證明

口試委員：謝文陽（簽名）

（指導教授）

湯森林

_____ _____

_____ _____

_____ _____

_____ _____

# 誌謝

感謝雙親對我的包容與照顧。感謝湯森林老師與謝文陽老師兩年多來的諄諄教誨與悉心指導。感謝台灣大學貴重儀器中心的諸位技術員、海研一號研究船上的船員們，還有夏復國老師、曾鈞懋老師及他們兩位實驗室的成員們於採樣期間給予的協助。感謝培汶學姊於實驗部分的重要貢獻，也要謝謝景鴻學長於分析方面的建議與指導。感謝崇斌學長與秀慧學姊總是很有耐心地回答我的問題。感謝美珠學姊、思妤學姊、嘉合學姊、凱迪學長對我的愛護與支持。感謝星儒與我並肩作戰為我打氣加油，還要謝謝承諭這一段時間的協助。完成這篇論文，絕對不是我一個人的成就，一直以來受到許多人的鼓勵與包容，包括我的家人：妹妹品融、弟弟俏暄和士韋，也包括我的高中、大學和研究所期間認識的好朋友們，由衷的感謝大家。

# 中文摘要

原核微生物是海洋中數量最多的生物，這些海洋原核微生物具有非常高的初級和次級生產力，對海洋甚至整個地球的元素循環扮演著重要的角色，因此為了了解這些微生物在海洋中的生態角色，探討海洋微生物群集的結構、分布和多樣性是很重要的基礎研究之一。目前大多數海洋原核微生物群集的研究集中於溫帶或寒帶海域表水或某一深度的調查，對內部的海洋微生物分布卻所知不多。相對地，西太平洋熱帶海域的海洋微生物群集研究極為缺乏，更遑論內部微生物群集。本研究即針對熱帶海洋內部微生物群集進行調查，以期能增加對於熱帶海洋內部微生物了解。

我們利用高輸出定序技術（massively-parallel pyrosequencing）對南海北部的海洋菌相與多樣性作了垂直剖面的調查。從四個不同深度（10、 100、 1000、 3000 公尺）的樣本中，獲得 38,704 條源自於細菌的核糖體小次單元之高度變異區的基因序列（the V6 hypervariable region of 16S rRNA gene）以及 5,269 條的古菌序列。經比較分析發現：（1）細菌群集多樣性遠高於過去用傳統方法所估計。（2）四個深度中，細菌皆高於古菌多樣性。（3）表水和中水層的菌相較深海的菌相複雜。

除群集多樣性的分析外，在細菌組成的分析結果指出 *α-proteobacteria* 和 *γ-proteobacteria* 是南海各水層佔有最高比例的兩種菌群，而 *Cyanobacteria* 則是在表水層佔較高比例。*β-proteobacteria* 菌群則在 10m 和 100m 的水層中佔有較高的比例；相反地，*δ-proteobacteria* 菌群則是在越深的水樣中佔有越高的比例。在古菌組成分析結果，發現 *Euryarchaeota* 在南海淺層 10m 到 3000m 共四個深度的水樣中都是古菌群集中的最優勢菌群，這和其他海域的深海以 *Crenarchaeota* 菌群為主不同。

在獨特序列（即相同序列集結為一獨立序列單位）間數目比較和親緣相關分析結果顯示，發現這些序列的比例多寡和深度有關。針對相對佔多數的序列進行親源關係的分析發現除了發現有些序列（種類）具有廣深性（在各個深度的水樣中都佔不少比例），也發現不少菌群有「依深度分群」的現象，例如四個分支的 *Oceanospirillales* 中，有兩個分支都是在 3000m 佔最高比例，其他三個深度都很少；

另一支則屬於淺層序列的分支，以及第四分支的序列則在中間兩個水層（100m 和 1000m）的較多。另外 *Alteromonadales*、*Pseudomonadales*、marine euryarchaeota group II 和 marine euryarchaeota group III 的序列也都有「依深度分群」的現象，這些結果顯示在南海北部的原核生物分布具有生物地理區隔的特徵。

我們的結果是目前熱帶海域的微生物多樣性中最完整的研究資料，可對南海微生物群集在垂直分布及組成有更深入的了解。

關鍵字：南海、微生物群集、細菌、古菌、多樣性

# ABSTRACT

Prokaryotic microbes are the largest population in oceans; they are evidently responsible for the major primary and secondary productions and, as a result, play a key role in significant element cycles in marine ecosystems and even global. However, the ecology, such as composition, distribution and diversity, of these microbes are still mainly unknown. To date, most of the studies of the marine prokaryotic community are only focused on the surface water or a certain depth in the mesopelagic zone of the polar or temperate oceans. Hence, the goal of this study is to survey the composition and diversity of the stratified microbial assemblages in the interior of the western tropical Pacific Ocean. We anticipate to extending the current understanding for the marine community in tropical water through this study.

By adopting a massively-parallel pyrosequencing strategy, we performed analyses of the diversity and composition of planktonic microbial communities isolated from the northern South China Sea (SCS; 18°15' N, 115°30' E), from the ocean's surface to near the sea bottom. Through examining 38,704 bacterial and 5,269 archaeal V6 amplicons of 16S ribosomal RNA genes, we observed several findings, including that (1) the bacterial diversity in northern SCS is greater than previous estimates based on conventional methods; (2) the bacterial community in the bathypelagic layer is less diverse than that in other three upper layers; (3) our data show the less diversity of the archaeal communities compared to the bacterial communities in four depths. In the composition analysis, our results show that the most abundant bacterial groups at four depths in the northern SCS are *α-* and *γ- proteobacteria*, while *Cyanobacteria* is mainly found from the sea surface. *β-proteobacteria* is more abundant in the euphotic zone than

in the aphotic zone; in contrast to *β-proteobacteria*, the relative abundance of sequences from *δ-proteobacteria* increases with increasing depth in northern SCS. Moreover, we revealed unique assemblages of Archaea in the northern SCS, with *Euryarchaeota* dominating throughout the water column, which is contrast to most of previous studies reported that *Crenarchaeota* are more abundant than *Euryarchaeota* in the deep sea. Furthermore, our results of profiles of the unique sequences show the presences of depth-specific inhabiting groups. Through the phylogenetic analysis of the dominant unique sequences, we observed many eurybathyal sequences (i.e., those sequences were detected in four depths) and the depth-related subdivision in many groups. For example, at least four clades are found in the phylogenetic tree of *Oceanospirillales* V6 sequences, and sequences in two clades are represented at the highest frequencies in the greast depth; one is the mainly upper layers clade, while sequences in the other one clade are highly abundant in the middle two layers. The depth-related subdivision was observed in other groups, including *Alteromonadales*, *Pseudomonadales*, and the marine euryarchaeota groups II and III, indicating biogeographical traits for prokaryotic communities exist in the norther SCS.

Our work presents the most comprehensive examination of microbial diversity in the tropical ocean and provides a significant insight into the marine microbial community structures along the vertical profiles in the SCS.

Index Terms — South China Sea, pyrosequencing, bacteria, archaea, community, diversity

# CONTENTS

口試委員會審定書 .....#

誌謝 .....i

中文摘要 ..... ii

ABSTRACT .....iv

CONTENTS .....vi

LIST OF FIGURES ..... viii

LIST OF TABLES .....x

**Chapter 1 Introduction.....1**

1.1 A brief history .....1

1.2 The importance of marine microbial diversity and community structures.....2

1.3 Microbial composition surveys in the South China Sea.....5

1.4 Massively-parallel pyrosequencing .....9

1.5 Aim of this study .....10

**Chapter 2 Materials and Methods.....12**

2.1 Sample collecting, processing, and sequencing.....12

2.1.1 Sample collection .....12

2.1.2 DNA extraction .....15

2.1.3 Microbe counting .....15

2.1.4 Mutliplexing sample preparation: Amplification of the V6 region DNA of 16S ribosomal RNA genes .....16

2.1.5 Massively-parallel pyrosequencing.....19

2.2 Analysis of V6 sequence .....21

2.2.1 Biodiversity analysis by using DOTUR....21
2.2.2 The taxonomic identification of V6 sequences....22
2.2.3 Profiles of the V6 unique sequences....22
2.2.4 Phylogenetic analysis of the V6 unique sequences....23
**Chapter 3 Results and Discussion....24**
3.1 Sampling information....24
3.1.1 The study area....24
3.1.2 Vertical profiles of CTD data, the T-S diagram, and oceanographic data....24
3.2 Direct cell counts....26
3.3 Analysis of V6 sequence....30
3.3.1 V6 sequencing information....30
3.3.2 Bacterial and archaeal diversity and rarefaction curves....31
3.3.3 Microbial community structures....39
3.3.4 Profiles of unique sequences....45
3.3.5 Phylogenetic analysis of the V6 unique sequences....51
**Chapter 4 General discussion and conclusions....79**
4.1 Diversity of the prokaryotes in the South China Sea....79
4.2 Depth-related distribution....80
4.3 Ambiguous taxonomically classified groups....84
4.4 Unique archaeal assemblages....85
4.5 Conclusions/significance....88
**Chapter 5 Future work....91**
REFERENCE....93

# LIST OF FIGURES


Figure 1. Map of the study area .....8

Figure 2. Flowchart for our studying methods .....13

Figure 3. Tangential Flow Filtration system .....13

Figure 4. Chemical duty vacuum pump and 0.2-μm cellulose acetate filter .....14

Figure 5. The 0.2-μm cellulose acetate filter was used to collect microbes .....14

Figure 6. Vertical profiles of temperature, salinity, and potential density .....27

Figure 7. Temperature-salinity diagram .....28

Figure 8. Analysis for species complexity .....38

Figure 9. Taxonomic distributions of microbes identified in the V6 libraries .....40

Figure 10. Profiles of the bacterial V6 unique sequences .....47

Figure 11. Profiles of the archaeal V6 unique sequences .....49

Figure 12. Phylogenetic tree of *α-proteobacteria* V6 sequences .....53

Figure 13. Phylogenetic tree of *β*- and *δ*- *proteobacteria* V6 sequences .....58

Figure 14. Phylogenetic tree of *Alteromonadales* (*γ-proteobacteria*) V6 sequences .....61

Figure 15. Phylogenetic tree of *Oceanospirillales* (*γ-proteobacteria*) V6 sequences .....64

Figure 16. Phylogenetic tree of *Pseudomonadales* (*γ-proteobacteria*) V6 sequences .....65

Figure 17. Phylogenetic tree of *Vibrionales* and the unclassified

gammaproteobacteria V6 sequences .....67

Figure 18. Phylogenetic tree of other non-proteobacteria V6 sequences .....69

Figure 19. Phylogenetic tree of *Crenarchaeota* V6 sequences .....70

Figure 20. Phylogenetic tree of the marine euryarchaeota group II V6 sequences .....72

Figure 21. Phylogenetic tree of the marine euryarchaeota group III V6 sequences .....76

Figure 22. Phylogenetic tree of the marine euryarchaeota group IV V6 sequences .....78

# LIST OF TABLES

Table 1. List of 5' tagged PCR primers .....20

Table 2. SEATS samples and oceanographic data .....29

Table 3. Sequencing data summary statistics and diversity indices for the bacterial communities .....33

Table 4. Sequencing data summary statistics and diversity indices for the archaeal communities .....34

# Chapter 1 Introduction

## 1.1 A brief history

The extremely tiny, invisible to human naked eyes individually, unicellular microorganisms are found in all environments (Madigan *et al*., 2009). Activities of the microbes drive the biogeochemical fluxes on the Earth, influence the climate globally, and occupy a peculiar place, no matter as pathogens or symbioses, in macroorgansims like humans (Arrigo, 2005; Strom, 2008; Falkowski *et al*., 2008; Madigan *et al*., 2009). Although the study of microorganisms is so important and closely related to us, microbiology is still a "young" scientific field because of the technical limitation. In the more than 6000-year human historic times, the records about microbes are sporadically found only in the last approximately 300 years after the day that single-lens microscope invented by Antonie van Leeuwenhoek. The direct observation initiated the knowledge about microbial morphologies. Thanks to the development of cultivation in the 19th century, the study on individual types and classification by nutritional criteria of microbes is moved on. It is not until the 1960's, the establishment era of molecular biology methods, and the work of Whittaker, the aspects of universal evolutionary relations between lives were revealed (Whittaker, 1969). However, the study on microbial evolution and diversity is hampered by that most of microbes could not be cultivated using current techniques (Pace, 1997). In the mid-1970s, the concept of the three domains (*Eukaryote*, *Bacteria*, and *Archaea*) was unveiled by Carl Richard Woese, who first realized that the small-subunit ribosomal RNA sequences could offer a way to relate all organisms (Woese and Fox, 1978). The molecular sequenced-based technique pioneered by Woese is a breakthrough in the study on uncultivated prokaryotes (Woese

and Fox, 1978; Woese, 1987; Pace, 1997). After that the culture-independent techniques are developed and used broadly, and a considerable number of microbial community structures and diversities in various habitats was revealed and reported (Muyzer *et al*., 1993; Clement *et al*., 1998; Moeseneder *et al*., 1999; Schauer *et al*., 2003; Castle and Kirchman, 2004; Brodie *et al*., 2006; Alonso-Sáez *et al*., 2007; Huber *et al*., 2007; Galand *et al*., 2009a). Now the 16S ribosomal RNA gene as a species-specific signature sequence has therefore become the most abundant single-gene sequence deposited in public databases.

## 1.2 The importance of marine microbial diversity and community structures

Many reports evidently revealed that the majority is free bacterioplankton in oceans (Hobbie *et al*., 1972; Azam and Hodson, 1977) and these bacteria are extremely important in primary and secondary productions in marine ecosystems. In the contemporary ocean, photosynthetic carbon fixation by marine phytoplankton leads to formation of ~45 gigatons of organic carbon per annum (Falkowski *et al*., 1998). Among these marine phytoplanktons, for instance, cyanobacteria numerically dominate most phytoplankton assemblages in the contemporary ocean and produce massive quantities of organic carbon (Falkowski *et al*., 1998). The marine autotrophic bacterioplanktons including cyanobacteria occupy largely in the oceans and are also known as the most important primary producers in the ocean ecosystems because they are able to synthesize new organic materials from inorganic molecules such as $H_2O$ and $CO_2$. These bacteria are responsible for 5 to 30 % of primary production for both coastal and offshore waters (Fuhrman *et al*., 1980; Azam *et al*., 1983; Fenchel, 1988; Falkowski

*et al*., 1998). In addition, other heterotrophic bacterioplanktons contribute the secondary production as significant consumers in the marine ecosystems with bacteria consuming 10 to 50 % of total fixed carbon and a carbon conversion efficiency of 50 %. (Fuhrman and Azam, 1980, 1982; Azam *et al*., 1983). Besides affecting the carbon cycle significantly, marine bacteria also play key roles in other main element cycles, e.g., nitrogen and phosphorus, and minor and trace element cycles, e.g., iron, arsenic and bismuth in marine ecosystems (Azam *et al*., 1983; Bentley and Chasteen, 2002; Arrigo, 2005).

Therefore, obtaining the information of distribution, composition and diversity of microbial community becomes crucial for getting understandings in correlation between microbe and their roles in local marine ecosystems. To detect the distribution, composition and diversity of microbial community, culture-independent molecular methods have been helpful and widely employed for such studies. These methods include clone library, T-RFLP (**T**erminal **R**estriction **F**ragment **L**ength **P**olymorphism), DGGE (**D**enaturing **G**radient **G**el **E**lectrophoresis), FISH (**F**luorescent ***I****n* ***S****itu* **H**ybridization), whole genome shotgun sequencing and massively-parallel pyrosequencing (Giovannoni *et al*., 1990; Schmidt *et al*., 1991; Britschgi and Giovannoni, 1991; Lee and Fuhrman, 1991; Fuhrman *et al*., 1993; Field *et al*., 1997; McInerney *et al*., 1997; Venter *et al*., 2004; Baldwin *et al*., 2005; DeLong *et al*., 2006; Alonso-Sáez *et al*., 2007; Huber *et al*., 2007; Santelli *et al*., 2008; Galand *et al*., 2009a; Feingersch *et al*., 2010).

The first study about the biodiversity of marine bacterioplanktons via the cultured-independent method was reported by Giovannoni *et al*. in 1990, and they phylogenetically analyzed the 16S ribosomal RNA gene sequences of the clone library from the Sargasso Sea. Giovannoni and his coworkers (1990) suggested 16S rRNA gene

is a good molecular tool for studying marine prokaryotic populations. After that, many unique marine bacterial and archaeal lineages were sequently discovered through the analysis of 16S rRNA gene sequences (Giovannoni *et al*., 1990; Schmidt *et al*., 1991; Britschgi and Giovannoni, 1991; Lee and Fuhrman, 1991; Fuhrman *et al*., 1993; Ma *et al*., 2009). Through analysis of 16S rRNA genes, species-richness of marine microbial communities ranged from a few hundred phylotypes per microliter in the seawater samples (Schloss and Handelsman, 2005) to over a thousand from marine sediments (Ravenschlag *et al*., 1999; Schauer *et al*., 2010). The presence of 516 ribotypes (unique rRNA sequences) was detected in one of the largest water column surveys (1,000 PCR amplicons), and the authors estimated occurrence of at least 1,633 co-existing ribotypes in the coastal bacterioplankton community (Acinas *et al*., 2004). Venter and his coworkers (2004) revealed that microbial diversity is very high in the Sargasso Sea through whole-genome shotgun sequencing, and estimated over 1800 species there including 148 previously unknown bacterial phylotypes. In 2006, the magnitude of the number of sequences which Sogin and his coworkers retrieved through the pyrosequencing technique is up to the fifth power of ten, and they showed that the diversity of bacterial communities in the deep North Atlantic Ocean are far more complex than previously reported for any microbial environment (Sogin *et al*., 2006). Manifestly, oceanic microbial communities are highly complex than as previously thought (Huber *et al*., 2007; Brown *et al.*, 2009).

In the past two decades, many reports have revealed that marine microbial community is influenced by environmental factors, such as latitude, depth, temperature, nutrient, salinity and light intensity (Allen *et al*., 2002; Baldwin *et al*., 2005; Giovannoni and Stingl, 2005; Schwalbach and Fuhrman, 2005; Hughes-Martiny *et al*., 2006; Pommier *et al*., 2007; Fuhrman *et al*., 2008). Some studies have reported that the

marine free-living microbial community structures alter in different locations; for example, the depth-related distribution of SAR202 bacteria was observed by Giovannoni *et al*. (1996). García-Martínez & Rodríguez-Valera (2000) showed that the genetic differences among SAR11 bacteria and among marine archaea varied with depth. Perceivably, in oceanic systems, depth is one of the most relevant gradients affecting the community structures in water column.

However, to date, most of the studies of marine prokaryotic community are focused on the surface layer, a certain depth in the mesopelagic zone, sediments, and hydrothermal vents. Few are the large-scale, vertical surveys of microbial community with the exception of reports from the site of the Hawaii Ocean Time-Series program (HOT; DeLong *et al*., 2006) and from the site in the Eastern Tropical North Pacific Ocean (Ma *et al*., 2009). The relevant information is so little that it's hard to provide the comprehensive view to the vertical distribution, composition and diversity of microbial community in the interior of oceans.

## 1.3 Microbial composition surveys in the South China Sea

Since the South China Sea Monsoon Experiment (SCSMEX) was launched, scientists in Taiwan investigate South China Sea (SCS) by many oceanographic cruises and by deploying several moorings. Those studies have revealed that the tropical SCS contains extensive unique characteristics; to monitor the biogeochemical processes and the response of the tropical oceanic system to climate change during various time scales, a time-series study was established in the northern SCS as known the SouthEast Asian Time-series Study (SEATS). The SEATS station is located at 18°15' N and 115°30' E in the northern SCS, and was recognized as a part of the international Joint Global Ocean

Flux Study (JGOFS) program in 1999 (Shiah *et al*., 1999; Wong *et al*., 2007; Figure 1).

South China Sea is one of the largest marginal sea on Earth (Sverdrup *et al*., 1942; Chen *et al*., 2001; Wong *et al*., 2007). The SCS covers a surface area of about 3.5 x $10^6 km^2$ (Wong *et al*., 2007; Hung *et al*., 2007) with a mean depth of 1350m, while its deepest basin is over 5000m deep (Wong *et al*., 2007; Hung *et al*., 2007). The Luzon Strait having about 3000m depth is the only deep, wide channel to mainly connect the SCS to the Pacific Ocean; therefore, the SCS is a semi-enclosed basin (Liu *et al*., 2007; Chang *et al*., 2010). Several large rivers such as the Pearl River and the Mekong River flow through densely populated areas and drain into the SCS (Wong *et al*., 2007; Hung *et al*., 2007; Liu *et al*., 2007). Climatic variations in the atmosphere and in the upper layer of the SCS are under persistent influence of the East Asian monsoon. In addition, the SCS is a common region having typhoons and energetic internal waves passing (Alford, 2003; Duda *et al*., 2004; Ramp *et al*., 2004; Chang *et al*., 2010). All of these physical disturbances could influence water-column stability, and lead to the vertical mixing of different time and spatial scales. The vertical mixing thus impacts on the exchange of nutrients between the cold deep water and the warm surface water, and sequential biological processes in the SCS. Therefore, SCS is a very good site to monitor biogeochemical processes and responses of a oceanic system to climate change during various time scales.

As a part of the JGOFS program, the SEATS station is mainly used for a long-term and regular measurement of the biogeochemical factors of the SCS. The hydrographic data measured by sensors and the oceanographic data of discrete water samples are included in the field sampling at SEATS station. In addition, the biological surveys including the determination of primary production, bacterial production, and the enumeration of phytoplankton and zooplankton have been taken at SEATS station

(Wong *et al*., 2007). The focuses of studies at SEATS station are carbon cycle, nutrient dynamics, and biological community structure (Wong *et al*., 2007). Most studies of microbial community structure at SEATS station are focused on *Prochlorococcus*, *Synechococcus*, picoeukaryotes, nanoplanktons and zooplanktons in the euphotic zone (Liu *et al*., 2007; Lee Chen *et al*., 2007). Although the diversity and composition of microbial community are very important as illustrated in previous sections, surveys about the composition and diversity of marine bacterioplankton are long in lack in the western tropical Pacific Ocean and in the northern SCS. Therefore, we selected the SEATS station to study the vertical distribution of prokaryotic community in the SCS, and furthermore compared these microbial profiles between sites at SEATS and in other oceans.

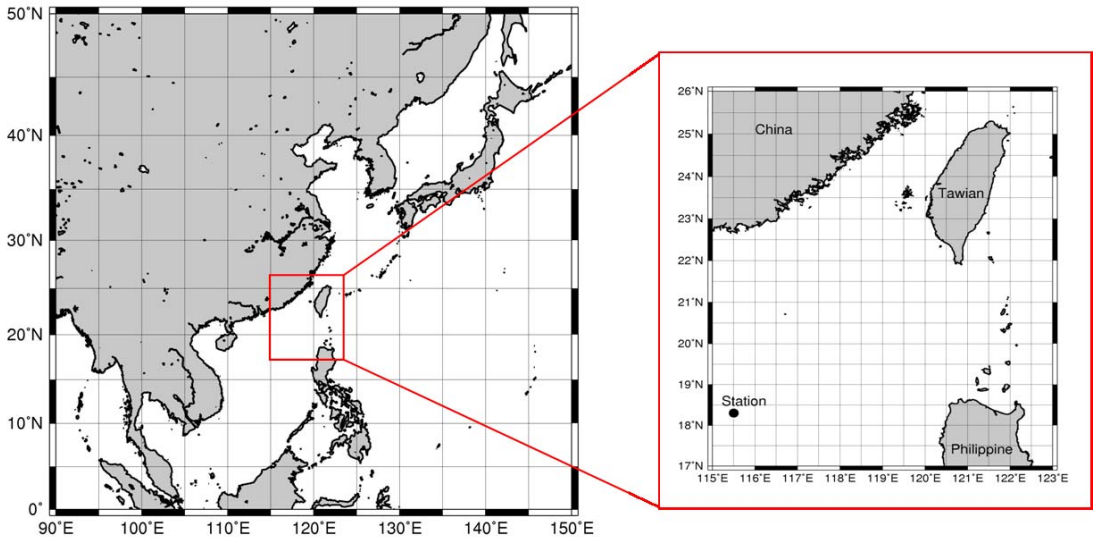


Figure 1. Map of the study area during the 845 cruise of R/V *Ocean Research I* from 21st to 30th Oct. 2007. The black circle indicates the sample collecting site, the SEATS station (18°15' N, 115°30' E).

## 1.4 Massively-parallel pyrosequencing

The development of pyrosequencing in the turn of the 21st century initiated a new sequencing era (Margulies *et al*., 2005). The advantages of this technique are high throughput, time saving, bias minimizing, and more cost-effective than traditional Sanger sequencing; therefore, the technique is also called next generation sequencing technique (Jarvie *et al*., 2005; Edwards *et al*., 2006). This robust technique has been broadly applied to study prokaryotic ecology in various environments to obtain more precise pictures of different prokaryotic community structures (Schuster *et al*., 2010; Roh *et al*., 2010; Manter *et al*., 2010; Andersson *et al*., 2010; Kirchman *et al*., 2010). The study of deep mine microbial ecology reported by Edwards *et al.* (2006) was the first application of the pyrosequencing to environmental samples. They obtained more than 70 Mbp of sequence data from over 700,000 sequences, and the cost was approximately 10 to 30 times cheaper than Sanger sequencing in that time (Edwards *et al*., 2006). Edwards and his coworkers (2006) revealed microbial community structures and metabolic potentials prevalent such as their mechanisms of iron acquisition and respiration. For another example, Sogin and his coworkers (2006) adopted this technique to investigate the bacterial community structures nearby the deep sea hydrothermal vents and they sequenced ~118,000 PCR amplicons that span the V6 hypervariable region of ribosomal RNAs from environmental DNA preparations. Their analyses showed that bacterial diversity in the diffuse flow vents of Axial Seamount and the deep water masses of the North Atlantic are much greater than any published description of marine microbial diversity, and they suggested that microbial diversity in the oceans (and likely elsewhere) should be much greater than previous estimates that were based on conventional molecular techniques (Sogin *et al*., 2006). In 2007, an

investigation about microbial community structures at two neighboring hydrothermal vents by examining the sequences of more than 900,000 microbial small-subunit ribosomal RNA amplicons was reported by Huber *et al*. (2007). They observed different population structures in the two vent communities that reflect local geochemical regimes, and revealed that the bacterial community is more diverse than the archaeal community (Huber *et al*., 2007). Obviously, the massively-parallel pyrosequencing is one of the most powerful techniques applied to study the diversity and compositions of marine microbial community (Galand *et al*., 2009a; Kirchman *et al*., 2010).

## 1.5 Aim of this study

In this study, we proceeded to a large spatial scale survey of microbial community throughout the seawater column in the SCS with several reasons as followed. First, the relationship between microbial groups and marine ecosystem function are very important but still much remain unknown. Second, the microbial composition and diversity surveys are long in lack in the western tropical Pacific Ocean. Third, in oceanic systems, depth is one of the most relevant gradients affecting the community structures in water column; few are the vertical surveys of microbial community with only few exceptions (DeLong *et al*., 2006; Ma *et al*., 2009). Fourth, the site of SEATS is an oceanic measurement site in the SCS, a part of western tropical Pacific Ocean; it enables offering comprehensive information of biological, chemical and physical measurements of the ocean. Furthermore, this study adopted the massively-parallel pyrosequencing to maximize our sample sizes for analysis of microbial community in SCS. Based on the reasons as described, we aimed to gain an in-depth perspective into microbial diversity and to better characterize the vertical distribution pattern of

prokaryotic composition through the ocean's interior by using the massively-parallel pyrosequencing technique to analyze the microbial V6 region of 16S rRNA genes from the seawater samples collected from four different depths at the SEATS station in the SCS (Figure 1).

# Chapter 2 Materials and Methods

## 2.1 Sample collecting, processing, and sequencing

Figure 2 shows the flowchart for our studying methods, and the details are described in the following sections.

### 2.1.1 Sample collection

Sample was collected during a cruise of R/V *Ocean Research I* (Cruise 845) of the Oceanography Institute of National Taiwan University from 21st October to 30th October 2007. Our sample site, the SouthEast Asia Time-series Study (SEATS), is located at 18°15' N and 115°30' E (Figure 1). We used the rosette (General Oceanics, Inc. Model 1015) which was carried the CTD package (Sea-Bird Electronics, Inc) and 10 to 12 twenty-liter Go-Flo bottles and obtained 140L seawater at three depth (10m, 100m, and 3000m) and 80L seawater at 1000m. Total seawater sample weighted over 500 kg were stored at -20˚C in the dark and transported to our lab by low-temperature delivery as soon as R/V *Ocean Research I* landed. Filtration for microbes in the laboratory was conducted under dim light as follows. Seawater (500L in total) was pre-filtered through a filter with pore size 10-μm (Nitex nylon net, Wildlife Supply Company) and sequentially used 0.22-μm Pellicon cassette TFF (**T**angential **F**low **F**iltration) systems using a peristaltic pump (Millipore, MODE: XX80EL) to collect retentate (Figure 3). The retentate was concentrated onto 0.2-μm cellulose acetate filter (ADVANTEC) using a chemical duty vacuum pump (Millipore, MODE: WP6111560, 80-100psi, Figure 4).

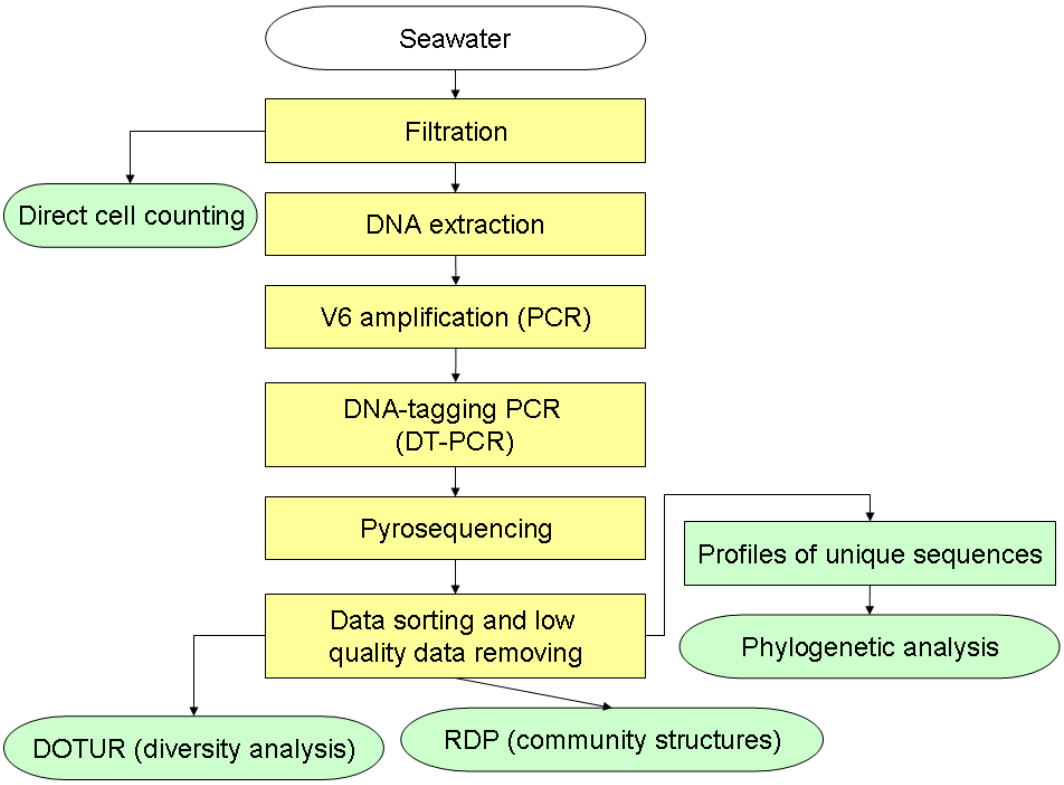


Figure 2. Flowchart for our studying methods.

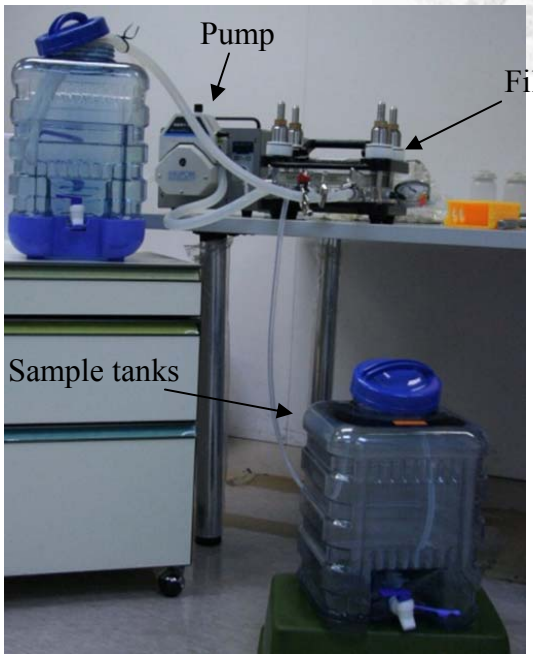


Figure 3. Tangential Flow Filtration (TFF) system using a peristaltic pump (Millipore, MODE: XX80EL).


Figure 4. The 0.2-μm cellulose acetate filter (ADVANTEC) and a chemical duty vacuum pump (Millipore, MODE: WP6111560, 80-100psi).

Figure 5. The 0.2-μm cellulose acetate filter was used to collect microbes (ADVANTEC).

### 2.1.2 DNA extraction

The 0.2-μm cellulose acetate filters were washed by 567μl of TE buffer (10mM Tris•HCl, pH7.5, and 1mM EDTA, pH8.0, Figure 5) and the suspension was placed in 1.5 ml eppendorf tubes respectively. Sodium dodecyl sulfate (SDS, 30μl; 10%) and RNase A (5μl; 100mg/ml) was added to each tube. The tubes were incubated at 37˚C for one hour. The lysate was then treated with Proteinase K (3μl; 20mg/ml) for half hour at 37˚C followed by the addition of NaCl (100μl; 5M) and 80μl CTAB (**C**etyl **T**rimethyl **A**mmonium **B**romide) for 10 minutes at 65˚C. The same volume of chloroform:isoamyl alcohol (24:1) was added to the lysate and the tubes were vortexed to mix and centrifuged at 13,000x*g* for 5 minutes. The supernatant was transferred to another eppendorf tubes respectively, followed by the addition of equal volume of phenol:chloroform:isoamyl alcohol (25:24:1 by v/v). The tubes were then vortexed to mix and centrifuged at 13,000x*g* for 5 minutes, and the supernatant was transferred to another eppendorf tubes respectively. The 0.6 volume of isopropanol was added to one volume of the supernatant. The tubes were gently inverted to mix and centrifuged at 13,000x*g* for 5 minutes. After discarding the supernatant, ethanol (30μl; 70%) was added to each tube. The tubes were centrifuged at 7,000x*g* for 5 minutes before removing the supernatant. The DNA pellets were dried in air and then resuspended in TE buffer and stored at -20˚C.

### 2.1.3 Microbe counting

The concentrated samples were diluted with appropriate volume of water (1ml at least). Total 0.8ml aliquots of the diluted samples were collected onto 0.02-μm pore sized membrane filters overlaid on the washed 0.45-μm pore sized membrane filters.

The pressure of the filtration was controlled under 15kPa. The staining solution contained 10% SYBR Gold (Invitrogen, dd$H_2O$:SYBR Gold (9:1)) and Antifade Mounting Medium (2μl of 10% p-phenylenediamine dihydrochloride to 198μl of glycerol:PBS (1:1 by v/v). Those 0.02-μm pore sized membrane filters with microbes were dried in a laminar flow hood. The filters were then placed on Petri dishes precontaining with 80μl of SYBR Gold working solution and stained for 15 to 20 minutes in dark and subsequently fixed by the Antifade Mounting Medium. The filters were examined by epifluorescence microscopy (ECLIPSE 90i, Nikon Corporation). Twenty fields were counted on each filter using the imaging software NIS-Elements (Nikon Corporation).

### 2.1.4 Mutliplexing sample preparation: Amplification of the V6 region DNA of 16S ribosomal RNA genes

To investigate into the composition of prokaryotes in our seawater samples, we used an updated uncultured-independent molecular method to approach. The 16S ribosomal RNA genes were used as an identification molecular marker in the method.

In contrast of using the clone library method (Ma *et al*., 2009), we adopted the Next-Generation Sequencing technique (i.e., massively-parallel pyrosequencing) to increase the number of sequences and greatly reduce the cost.

To apply this technique for sequencing multiplexing samples in a single run, adding coded primers (unique tags) of polymerase chain reaction (PCR) to individual DNA samples has been a common method to obtain multiple homolog amplification sequencing reads from mixing sequencing reads of different DNA samples in a single plate (Binladen *et al*., 2007; Roh *et al*., 2010). For this strategy, however, a trade-off

problem existing affects the length of code adding at the 5'end of the universal primers of 16S rRNA genes. If the length of code is two-bases, the maximum kind of different codes is sixteen so that only 16 samples can be simultaneously run in a single plate. If more samples are sequenced in a single run, the length of codes should be longer; however, the long length of codes will affect the efficiency of PCR and bring new biases in the DNA amplification (Binladen *et al*., 2007).

Considering a suitable code number for our samples, the length of codes we adopted is four-bases. Therefore, the arrangements of the four nucleotides in the code could produce 4×4×4×4=256 different tags. Because some 5'-tagged PCR primers might be more preferable for the DNA templates in PCR (Binladen *et al.*, 2007), we thus developed a new strategy to minimize the bias. The strategy was to amplify the DNA of 16S rRNA gene by using the traditional PCR with universal primers and then to add unique tags to the PCR product by 5 PCR cycles (i.e., DNA tagging PCR). Details are described as follows.

**The traditional PCR.** Because the secure sequencing length for each read by the Genome Sequencer FLX System is about 150 bases, we selected the hypervariable V6 region of 16S rRNA genes (abbreviated as V6) for detecting microbial diversity and microbial community composition (Huse *et al.*, 2008). A 50-μl reaction mixture for bacterial V6 rRNA gene amplification contained 1x *EX taq* buffer (Takara Bio Inc.), 0.2mM each deoxynucleoside triphosphate (dNTPs), 0.2μM forward primer 967F (5'-CAACGCGAAGAACCTTACC), 0.2μM reverse primer 1046R (5'-CGACAGCCATGCANCACCT), 5U *TaKaRa Ex Taq*™ HS, and 2-5ng environmental DNA template. Polymerase chain reaction cycling were performed using a PxE Thermal Cycler (Thermal Electron Corporation, USA) with the following cycle program: an initial denaturation of 3 minutes at 94˚C; 30 seconds at 94˚C, 20 seconds at

55˚C, 20 seconds at 72˚C for a total of 30 amplification cycles, and followed by a final extension at 72˚C for 2 minutes. In the case of archaea, a 50-µl reaction mixture consisted of 1x *EX taq* buffer (Takara Bio Inc.), 0.2mM of dNTPs each, 0.5µM forward primer 958arcF (5'-AATTGGANTCAACGCCGG), 0.5µM reverse primer 1048arcR (5'-CGRCGGCCATGCACCWC), 5U *TaKaRa Ex Taq*™ HS, and 2-5ng environmental DNA template. Reaction mixtures for archaea were carried out under the following conditions: 30 cycles (30 seconds at 94˚C, 10 seconds at 56˚C, 20 seconds at 72˚C) preceded by 3 minutes initial denaturation at 94˚C and followed by 2 minutes final extension at 72˚C. PCR products from bacterial and archaeal primer sets were visualized by using agarose gel electrophoresis and purified by QIAEX II Gel Extraction Kit as described by manufacturer's protocol (QIAGEN, http://www.ebiotrade.com/buyf/productsf/qiagen/20021.htm).

**The DNA tagging PCR (DT-PCR).** The DNA tagging PCR (DT-PCR) technique is a method to tag the previous purified V6 PCR products from different samples by PCR with various codes, separately. We modified the original primers into 16 unique primers by adding 4 nucleotides codes at 5' end of the primers (Table 1).

The DT-PCR products from four different depth samples were amplified with different tagged primer sets and performed using a PxE Thermal Cycler (Thermal Electron Corporation, USA) with the following cycle program: an initial step of 3 minutes at 94˚C; 30 seconds at 94˚C, 10 seconds at 57˚C, 30 seconds at 72˚C for a total of 5 amplification cycles, and followed by a final extension at 72˚C for 2 minutes. The 50-µl DT-PCR reaction mixture contained 1x *EX taq* buffer (Takara Bio Inc.), 0.2mM of dNTPs each, 0.4µM of each primer, 2.5U *TaKaRa Ex Taq*™ HS, and 20-50ng V6 amplicons. 3µl of the tagged PCR products were loaded into 15% non-denatured polyacrylamide gels ran on 100V for 6 hours. The tagged PCR products were visualized

and verified by the molecular weight from the gels after SYBR-Gold staining and purified by QIAEX II Gel Extraction Kit as described by manufacturer's protocol (QIAGEN, http://www.ebiotrade.com/buyf/productsf/qiagen/20021.htm).

### 2.1.5 Massively-parallel pyrosequencing

About 200ng of each tagged V6 amplicons were pooled together and sent to Mission Biotech Corporation (Taipei, Taiwan) for sequencing with the Genome Sequencer FLX System (454 Life Sciences, Roche Diagnostics Corporation).

We developed a program to remove below 74 bp reads or over 116 bp reads and to sort each V6 sequence back to its individual source through 5′ code sequence matching. This strategy was sufficient to remove the reads containing ambiguous bases (N) and the reads mismatched to the primer sequences. Eliminating low quality has ensured the accuracy and quality of reads for further analyses (Huse *et al*., 2007).

Table 1. A list of 5’ tagged PCR primers.

| Sample | Primers | Name | Sequence (5’-3’) |
|---|---|---|---|
| 10m bacteria | Forward | 967F-5 (5F) | **AATT**CAACGCGAAGAACCTTACC |
| 10m bacteria | Reversed | 1046R-5 (5R) | **AATT**CGACAGCCATGCANCACCT |
| 100m bacteria | Forward | 967F-6 (6F) | **AATC**CAACGCGAAGAACCTTACC |
| 100m bacteria | Reversed | 1046R-6 (6R) | **AATC**CGACAGCCATGCANCACCT |
| 1000m bacteria | Forward | 967F-7 (7F) | **AATG**CAACGCGAAGAACCTTACC |
| 1000m bacteria | Reversed | 1046R-7 (7R) | **AATG**CGACAGCCATGCANCACCT |
| 3000m bacteria | Forward | 967F-8 (8F) | **AACA**CAACGCGAAGAACCTTACC |
| 3000m bacteria | Reversed | 1046R-8 (8R) | **AACA**CGACAGCCATGCANCACCT |
| | | | |
| 10m archaea | Forward | 958arcF-10 (A_10F) | **AACC**AATTGGANTCAACGCCGG |
| 10m archaea | Reversed | 1048arcR-10 (A_10R) | **AACC**CGRCGGCCATGCACCWC |
| 100m archaea | Forward | 958arcF-11 (A_11F) | **AACG**AATTGGANTCAACGCCGG |
| 100m archaea | Reversed | 1048arcR-11 (A_11R) | **AACG**CGRCGGCCATGCACCWC |
| 1000m archaea | Forward | 958arcF-12 (A_12F) | **AAGA**AATTGGANTCAACGCCGG |
| 1000m archaea | Reversed | 1048arcR-12 (A_12R) | **AAGA**CGRCGGCCATGCACCWC |
| 3000m archaea | Forward | 958arcF-13 (A_13F) | **AAGT**AATTGGANTCAACGCCGG |
| 3000m archaea | Reversed | 1048arcR-13 (A_13R) | **AAGT**CGRCGGCCATGCACCWC |

## 2.2 Analysis of V6 sequence

### 2.2.1 Biodiversity analysis by using DOTUR

After removing low quality data, the sequences were aligned by the Clustal W package (version 2.10) (Thompson *et al*., 1994). The pairwise matrices with the Jukes-Cantor correction were calculated after alignment. These pairwise matrices served as input to the program distance-based OTUs and richness (DOTUR) for clustering sequences to operational taxonomic units (OTUs) with 3% dissimilarity as a cutoff value for defining the microbial populations (Stackebrandt and Goebel, 1994). Full bias-corrected Chao1 nonparametric richness estimator, abundance-based coverage estimator (ACE), the Shannon-Weaver and Simpson diversity indices, and rarefaction curves were also computed by using the DOTUR software package (Schloss and Handelsman, 2005). Rarefaction curves are performed to predict the probable total number of OTUs present in the source assemblage. As the curve goes like a linear behavior, it means the significantly increased estimators of total diversity will be obtained after the additional sampling. When the curve reveals by the asymptotic behavior, the library is considered to be ‘enough’ to have a probable estimate of richness.

Those identical (with 100% similarity) sequences were defined as the unique sequences. The abundance of OTUs and the unique sequences in each library were used to calculate the Good’s coverage values (Good, 1953). As defined by Good (1953), the coverage (C) is calculated by the equation, C = (1-n/N) x 100%, where n is the number of OTUs with single presence or the unique sequences occurring once in a library and N is the total number of sequences. The proportion of phylotypes in a sample of infinite

size represented in a subsample is estimated by the Good's coverage percentage.

### 2.2.2 The taxonomic identification of V6 sequences

After removing low quality data, a total of 38,704 bacterial V6 amplicons and 5,269 archaeal V6 amplicons were used for the taxonomic identification at the rank of Order if possible using a combinatory strategy. First, the initial taxonomic information of V6 tag sequences were determined by the Sequence Match tool of the Ribosomal Database Project II (RDP II) with default parameters except that the parameter of KNN matches was set as 5 (Cole *et al*., 2008). The format "Phylum_Class_Order" was used to describe the taxonomic information in this study. Taxonomic assignments were verified from the highest (Phylum) to the lowest rank (Order). Once a sequence at a rank couldn't be determined, the rest ranks of the sequence were named as "not assigned (N/A)". Second, the sequences containing N/A classification were re-verified by using the Blastn search for corresponding 16S rDNA sequences in the public database of the National Centre for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/). The results of two-step taxonomic assignments were presented by the 100% stacked bar charts. For convenient use in further analyses, the serial numbers with a capital letter "T" as the beginning were symbolized to denote the taxonomic assignments.

### 2.2.3 Profiles of the V6 unique sequences

Using a 100% sequence similarity cutoff to distinguish unique sequences in the whole dataset of the four depths libraries of V6 sequences, we obtained 8,111 and 1,475 unique sequences in the whole bacterial and archaeal dataset, respectively. The relative

abundances of each unique sequence in each library were calculated by the counts of each unique sequence divided by the total number of V6 sequences in each library.

### 2.2.4 Phylogenetic analysis of the V6 unique sequences

A total of 8,111 and 1,475 unique sequences in the bacterial and archaeal dataset were sorted by the taxonomic assignments, respectively. The serial numbers led by a capital letter “U” and combined with the taxonomic codenames were used to denote the unique sequences.

The unique sequences with the relative abundances higher than 0.45% in one of four libraries were selected to construct the phylogenetic trees. All sequences used for the phylogenetic analyses were aligned by the program MUSCLE (version 3.6; Edgar, 2004). Phylogenetic trees were accomplished with bootstrap analysis (1000 replicates) by the software MEGA 4 (Tamura *et al*., 2007) using the neighbor-joining and the maximum parsimony algorithms with evolutionary distances (Jukes-Cantor distances).

# Chapter 3 Results and Discussion

## 3.1 Sampling information

### 3.1.1 The study area

Sample was collected during a cruise of *R/V Ocean Research I* (Cruise 845) of the Oceanography Institute of National Taiwan University from 21st October to 30th October 2007. The methods of sample collection were described in Chapter 2. Sample depths were selected based on the previous study (Pei-Wen Chiang, in personal communication). According to the previous study, we found the surface water layer (10m~80m), and the mesopelagic layer (300m~2000m) had similar pattern in DGGE profiles respectively. The most different pattern in DGGE profiles was observed in the 100m-depth. Therefore, we selected 10m for the surface layer, 100m for its difference, 1000m for the mesopelagic layer, and 3000m for the bathypelagic layer. We obtained over 80 Liter seawater at each four depth (10m, 100m, 1000m, and 3000m at the SEATS Station; Table 2). Seawater samples were collected for preparing microbial V6 amplicon libraries.

### 3.1.2 Vertical profiles of CTD data, the T-S diagram, and oceanographic data

The hydrographic data were measured by the Sea-Bird CTD profilers and showed in Figure 6. We observed that the temperature profile is continuously stratified. The depth of 15˚C isotherm was defined as the thermocline depth by Zheng *et al*. (2007) in

the subtropic region, and the thermocline is between ~70m and ~200m in the northern SCS, separating the upper mixed layer from the calm deep water below. The temperature was nearly 28˚C at the upper mixed layer, and then declined with the increasing depth. A contrary pattern was found in the potential density profile. The salinity was about 33.6psu at the sea surface and increased with depth, peaked at the ~130m depth (~34.6psu), decreased slightly to a local minimum (~34.4psu) at depths between 350m and 430m, and then increased with depth again. The salinity was quite stable in the deep northern SCS (34.5psu and 34.6psu at the depths of 1000m and 3000m, respectively).

The water characteristics could be analyzed by the temperature-salinity (T-S) diagram, indicating that there are three water masses in the northern SCS (Figure 7). Our results in hydrographic data are similar to those reported previously in the northern SCS (Figure 6 and 7; Qu *et al*., 2000; Higginson *et al*., 2003; Wong *et al*., 2007), and the northern SCS surface water both influences by the freshwater input and Kuroshio intrusion so it has lower salinity than the Kuroshio water and the Pacific water. The intermediate water of SCS has similar characteristics with the North Pacific Intermediate Water (NPIW; Wong *et al*., 2007). The cold deep SCS water is a mixture of Circumpolar Deep Water (CDW: 0.1~2.0˚C in temperature and 34.62~34.73 in salinity; Emery, 2001) and Pacific Subarctic Intermediate Water (PSIW: 5.0~12.0˚C in temperature and 33.9~34.4psu in salinity; Emery, 2001; Chang *et al*., 2010).

Except the physical parameters, we measured other oceanographic data such as the concentrations of nitrite, nitrate, and dissolved oxygen (DO) as shown in Table 2. The concentration of DO was highest in the surface water (~200μM), decreased with increasing depth, and rose to about 110μM in the 3000m depth. The concentrations of nitrite, nitrate, dissolved inorganic phosphate (DIP), and silicate increased with depths,

and these vertical patterns are similar to the previous studies (Chen *et al*., 2001; Wong *et al*., 2007). In agreement with the earlier observations, the surface chlorophyll *a* (chl *a*) concentration was about 0.16μg/l in the northern SCS (Cai *et al*., 2004; Liu *et al*., 2007), and the concentration of dissolved inorganic carbon (DIC) was nearly to 2000μmol/kg and increased with depth (Cai *et al*., 2004). The vertical pattern of pH is also consistent and similar as observed in the previous report (Chen *et al*., 2006).

## 3.2 Direct cell counts

The cell density in 100m depth was the highest (8.9 x $10^4$ cell/ml on average). The cell densities in the surface layer and in 1000m depth were both 4.4 x $10^4$ cell/ml on average. The lowest cell density was obtained from the greatest depth (1.7 x $10^4$ cell/ml on average in the 3000m layer). The results are slightly lower than the cell density observed from the previous reports at the SEATS station (Liu *et al*., 2007; Zhang *et al*., 2009). This may be due to the different methods. Total cell counts declined with increasing depth between 100m and 3000m, and this trend is consistent with the earlier studies in the northern SCS (Zhang *et al*., 2009) and in the Pacific Ocean (Karner *et al*., 2001).

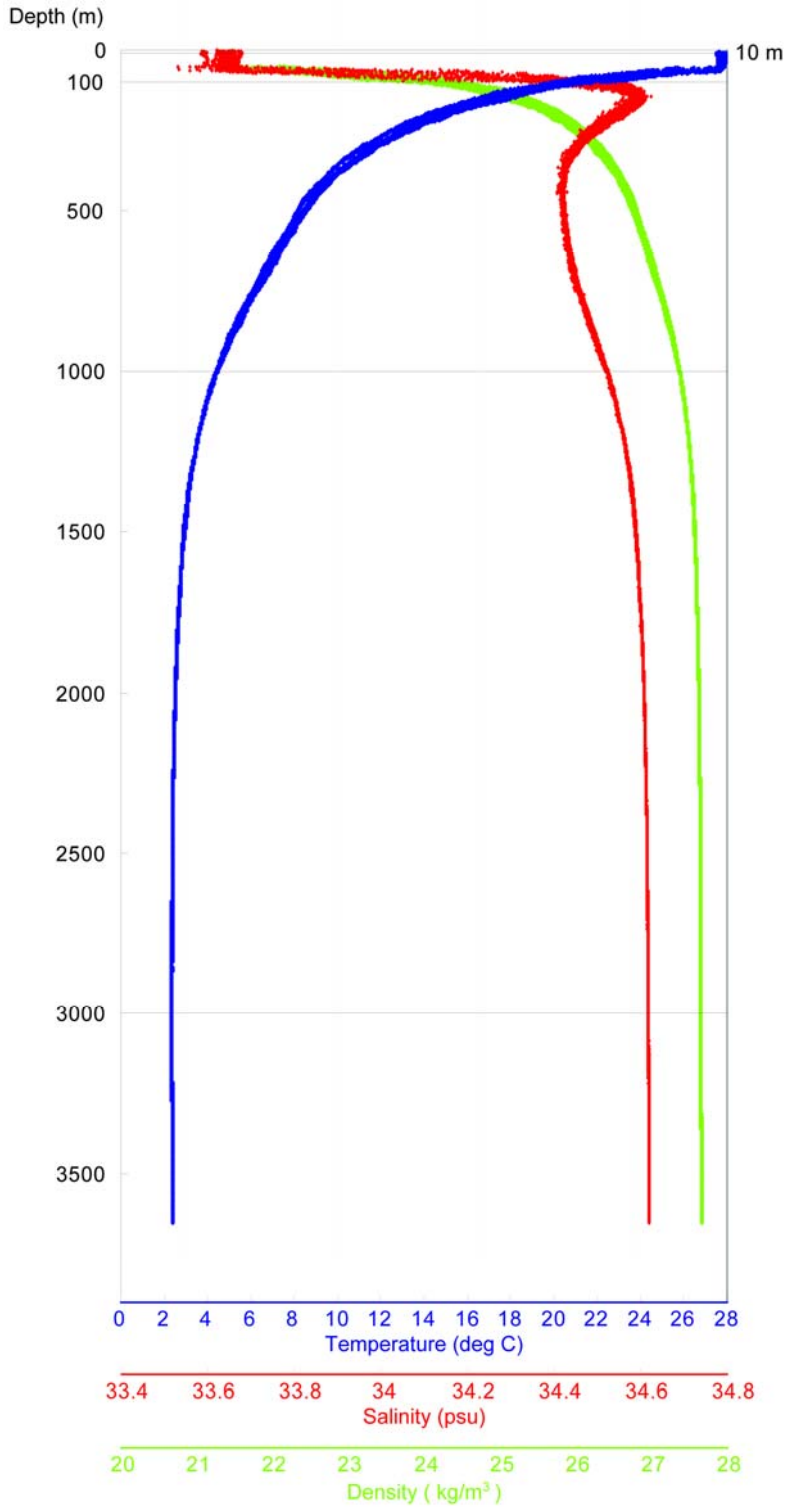


Figure 6. Vertical profiles of temperature (blue), salinity (red), and potential density (green) at the SEATS station (18°15' N, 115°30' E; datasets provided by Dr. Chun-Mao Tseng and his coworkers).

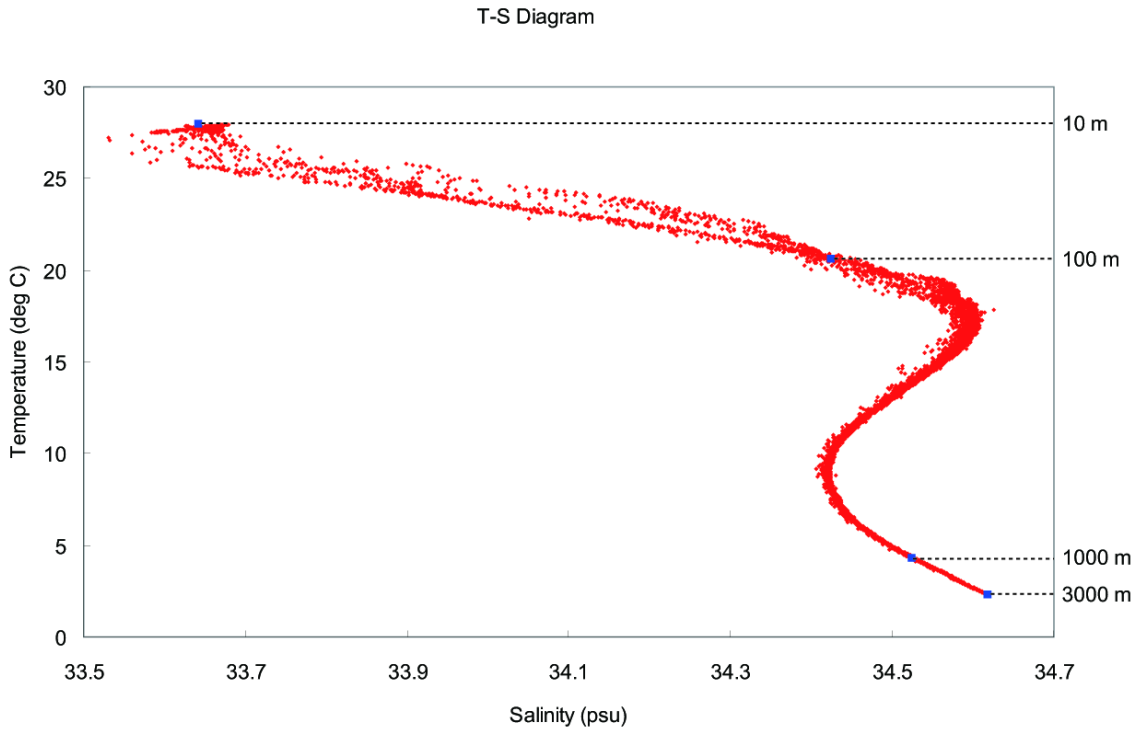


Figure 7. Temperature-salinity (T-S) diagram at the SEATS station (18°15' N, 115°30' E). The data envelope shows the temperature and salinity measured during the 845 cruise of R/V *Ocean Research I* from 21st October to 30th October 2007. The blue squares indicate the four depths of water samples collected in this study (Datasets provided by Dr. Chun-Mao Tseng and his coworkers).

Table 2. SEATS samples and oceanographic data during 2007/10/23 to 2007/10/26. Sample site, 18°15' N, 115°30' E (Datasets provided by Dr. Chun-Mao Tseng and his coworkers).

| | **10m** | **100m** | **1000m** | **3000m** |
|---|---|---|---|---|
| Volume filtered (Liters) | 140 | 140 | 80 | 140 |
| Temperature (˚C) | 27.79±0.09 | 20.56±0.59 | 4.42±0.02 | $2.36\pm2.00\times10^{-3}$ |
| Salinity (psu) | 33.65±0.02 | 34.42±0.05 | $34.52\pm2.00\times10^{-3}$ | $34.62\pm9.57\times10^{-5}$ |
| Dissolved Oxygen (μM) | 199.84±0.16 | 124.02±18.86 | 91.25±1.28 | $113.85\pm1.42\times10^{-4}$ |
| Nitrate, $NO_3^-$ (μM) | 0.02±0.03 | 12.88±1.62 | 36.29±0.15 | 37.99±0.10 |
| Nitrite, $NO_2^-$ (μM) | 0.00±0.00 | 0.04±0.01 | $1.00\times10^{-3}\pm1.00\times10^{-3}$ | $1.00\times10^{-3}\pm2.00\times10^{-3}$ |
| DIP (μM) | $0.03\pm4.00\times10^{-3}$ | 0.85±0.14 | 2.74±0.04 | $2.86\pm5.00\times10^{-4}$ |
| Silicate, $SiO_4^{2-}$ (μM) | 2.64±0.04 | 13.27±1.53 | 120.05±1.31 | 147.55±0.54 |
| Chlorophyll a (μg/l) | 0.16±0.02 | 0.11±ND | $4.00\times10^{-3}$±ND | ND |
| DIC (μmol/kg) | 1907.89±1.80 | 2073.21±1.13 | 2312.09±0.59 | 2344.50±1.30 |
| pH | $8.09\pm1.00\times10^{-3}$ | 7.86±0.02 | $7.53\pm3.00\times10^{-3}$ | $7.55\pm3.00\times10^{-3}$ |

Values in parentheses are the mean ± standard deviation of the CTD casts. Abbreviations are DIP, dissolved inorganic phosphate; DIC, dissolved inorganic carbon; and ND, not determined.

## 3.3 Analysis of V6 sequence

### 3.3.1 V6 sequencing information

We extracted DNA from four depth seawater samples and used DT-PCR technique with 5′-nucleotide tagged primers to generate microbial and archaeal V6 amplicons, followed by pyrosequencing (the FLX system, Roche/454 Life Sciences; as illustrated in the Chapter 2). The first 454 run generated 365,213 reads with an average length of 112.60 nucleotides and a total length of ~41 Mb, including SEATS bacterial V6 sequences and the samples of other projects. The second 454 run generated 47,410 reads with a length of 107.34 nucleotides on average and a total length of ~5 Mb, including SEATS archaeal V6 amplicons and the samples of other projects. We traced each V6 sequence back to its individual source through 5′ tag-analysis and trimmed either too short or too long reads. In all, 38,704 bacterial V6 amplicons and 5,269 archaeal V6 amplicons were used for further analyses in this study (Table 3 and 4).

Operational taxonomic units (OTUs) were determined by aligning tag sequences before calculating distances matrices and then using the program Distance Based OTUs and Richness Determination (DOTUR; Schloss and Handelsman, 2005) to create clusters at the unique and over 97% sequence identity level (Table 3 and 4).

In the 10m-depth data set, 2,861 unique V6 sequences of 11,230 bacterial V6 amplicons grouped into 680 phylotypes (OTUs); 2,037 unique V6 sequences from the 100m-depth data set of 6,747 bacterial V6 amplicons defined 581 phylotypes (OTUs). In the 1000m-depth data set, 669 OTUs were comprised from 2,755 unique V6 sequences, represented by 12,651 bacterial V6 amplicons. The 3000m-depth data set of 8,076 bacterial V6 amplicons wasn’t the smallest data set, but only 272 OTUs were

defined by 1,510 unique V6 sequences (Table 3).

In Table 4, the total number of archaeal V6 sequences at each depth varied between 691 and 1,720 (10m: 1,585; 100m: 1,720; 1000m: 691; and 3000m: 1,273). Based on a cut-off set on 100% similarity, the number of unique sequences ranged from 307 to 537 in the archaeal libraries (corresponding to 73-106 OTUs using 97% sequence similarity as a taxa proxy).

### 3.3.2 Bacterial and archaeal diversity and rarefaction curves

We also used the program DOTUR to calculate various diversity indices and richness within V6 amplicons (Table 3 and 4) and to generate rarefaction curves (Figure 8). The Good's coverage values were applied to evaluate whether screening these amount of sequences was sufficient to estimate diversity (Table 3 and 4).

**Bacterial diversity.** Nonparametric statistical richness estimates (Chao 1 and ACE) (Table 3) and rarefaction curves (Figure 8a) based on a 97% sequence identity showed that the bacterial richness values were higher in the 10m-, 100m-, and 1000m-depth data sets (800 phylotypes or more). These analyses predicted the lowest bacterial richness for the deepest seawater sample (365 and 424 phylotypes estimated by Chao 1 and ACE, respectively). The Shannon values decreased with depth increasing, indicating that the 3000m-depth layer is less diverse than the shallower layers (4.28 versus 4.71 or more). The evenness values were similar in four samples, with an average value of 0.75 (Table 3). The Good's coverage values of OTU at a 97% similarity level ($C_{OTU}$) were 98% on average in the bacterial libraries. With stricter criteria, the Good's coverage values of unique sequences (C) ranged from 79.5% to 86.9%. According to rarefaction curves, the other approach applied to evaluate sequencing effort, we obtained that the 3000m-depth curve more closely reached an asymptote than the curves of other depths (Figure 8a).

These results indicated that more sequencing effort would reveal further difference in diversity between the three shallower depths and the bathypelagic layer. Nevertheless, higher bacterial diversities are observed in upper layers in agreement with the previous study (Brown *et al*., 2009).

Except the ~~eutrophic~~ environments such as marine sediments, hydrothermal vents, and microbial mats of basalts, the values of observed OTUs, richness estimates, ACE, Chao1, and Shannon index obtained from the SEATS station in the northern SCS are higher than most previous studies in other oceans by methods such as 16S rRNA gene clone libraries and T-RFLP (Venter *et al*., 2004; Gallagher *et al*., 2004; Alonso-Sáez *et al*., 2007; Pommier *et al*., 2007; Santelli *et al*., 2008; Ma *et al*., 2009). This might indicate that the bacterial communities in the northern SCS are more diverse than that in other oceans. On the other hand, Kirchman *et al.* (2010) have studied the bacterial communities in the Arctic Ocean by pyrosequencing and analyzed 19,459 sequences on average for each library (we obtained an average of 9,676 sequences form the SEATS station). The values of observed OTUs, ACE, and Chao1 obtained from Kirchman *et al*. (2010) are 1.5 times higher than our results. However, these values are correlated extremely and tend to increase with the number of sequences analyzed (Schloss and Handelsman, 2005). Thus, the differences between our results and the results of other studies are much probably due to different sequencing efforts. In this case, in order to fairly compare microbial diversity in different environments, the sample sizes should be comparable.

Table 3. Sequencing data summary statistics and diversity indices for the **bacterial** communities as represented in the V6 region of 16S rRNA gene libraries[a].

| **Index** | **10m** | **100m** | **1000m** | **3000m** |
|---|---|---|---|---|
| Number of bacterial V6 tag sequences (N) | 11230 | 6747 | 12651 | 8076 |
| Unique bacterial V6 tag sequences | 2861 | 2037 | 2755 | 1510 |
| $n_1$[b] | 1930 | 1385 | 1847 | 1055 |
| Good's coverage (C) (%)[c] | 82.8 | 79.5 | 85.4 | 86.9 |
| S[d] | 680 | 581 | 669 | 272 |
| $S_1$[e] | 243 | 227 | 167 | 94 |
| Good's coverage ($C_{OTU}$) (%)[f] | 97.8 | 96.6 | 98.7 | 98.8 |
| Evenness[g] | 0.76 | 0.76 | 0.72 | 0.76 |
| Richness[h] | 167.64 | 151.47 | 162.84 | 69.36 |
| Shannon | 4.93 | 4.81 | 4.71 | 4.28 |
| Chao 1[i] | 990 (901, 1114) | 886 (796, 1015) | 878 (816, 966) | 365 (327, 430) |
| ACE[i] | 981 (946, 1020) | 863 (837, 891) | 910 (880, 944) | 424 (371, 504) |
| Simpson | 0.020 | 0.025 | 0.026 | 0.022 |

[a]Calculations are based on OTUs formed at an evolutionary distance of <0.03 (or greater than 97% similarity).

[b] $n_1$ is the number of the unique sequences appearing only once in a library.

[c] $C = (1- n_1/N) \times 100\%$

[d]S is defined as the number of OTUs.

[e] $S_1$ is the number of singleton OTUs.

[f] $C_{OTU} = (1- S_1/N) \times 100\%$

[g]Evenness is defined as Shannon / Ln(the number of OTUs)

[h]Richness = (number of singleton OTUs-1)/logN. The maximum value is (N-1)/logN

[i]Confidence intervals for the Chao 1 estimator and the ACE estimator are shown in parenthesis.

Table 4. Sequencing data summary statistics and diversity indices for the **archaeal** communities as represented in the V6 region of 16S rRNA gene libraries[a].

| **Index** | **10m** | **100m** | **1000m** | **3000m** |
|---|---|---|---|---|
| Number of archaeal V6 tag sequences (N) | 1585 | 1720 | 691 | 1273 |
| Unique archaeal V6 tag sequences | 423 | 537 | 307 | 395 |
| $n_1$[b] | 271 | 367 | 220 | 274 |
| Good's coverage (C) (%)[c] | 82.9 | 78.7 | 68.2 | 78.5 |
| S[d] | 106 | 99 | 73 | 80 |
| $S_1$[e] | 54 | 42 | 25 | 32 |
| Good's coverage ($C_{OTU}$) (%)[f] | 96.6 | 97.6 | 96.4 | 97.5 |
| Evenness[d] | 0.73 | 0.65 | 0.79 | 0.63 |
| Richness[e] | 32.81 | 30.29 | 25.36 | 25.44 |
| Shannon | 3.05 | 3.00 | 3.41 | 2.75 |
| Chao 1[i] | 236 (166, 386) | 165 (129, 247) | 111 (87, 176) | 108 (92, 146) |
| ACE[i] | 205 (159, 292) | 161 (132, 218) | 100 (85, 134) | 128 (103, 176) |
| Simpson | 0.083 | 0.108 | 0.057 | 0.120 |

[a]Calculations are based on OTUs formed at an evolutionary distance of <0.03 (or greater than 97% similarity).

[b] $n_1$ is the number of the unique sequences appearing only once in a library.

[c] $C = (1- n_1/N) \times 100\%$

[d]S is defined as the number of OTUs.

[e] $S_1$ is the number of singleton OTUs.

[f] $C_{OTU} = (1- S_1/N) \times 100\%$

[g]Evenness is defined as Shannon / Ln(the number of OTUs)

[h]Richness = (number of singleton OTUs-1)/logN. The maximum value is (N-1)/logN

[i]Confidence intervals for the Chao 1 estimator and the ACE estimator are shown in parenthesis.

Comparing to the rarefaction curves obtained from the Arctic Ocean by Kirchman *et al*. (2010) and Galand *et al*. (2010), at the number of sequences sampled ranging from 0 to 6500, the values of observed OTUs in the SEATS 3000m library is the lowest, indicating that the bacterial richness values are lower in the deep northern SCS than in the Arctic Ocean. The values of observed OTUs in the SEATS 10m, 100m, and 1000m libraries are similar to those obtained from the surface water in the Arctic Ocean (Kirchman *et al*., 2010) and lower than those obtained from the deep Arctic Ocean (Galand *et al*., 2010). These results show that the bacterial richness values in the northern SCS are similar to that in the shallow Arctic Ocean but lower than that in the deep Arctic Ocean (Kirchman *et al*., 2010; Galand *et al*., 2010). It's noticeable that the number of sequences sampled by Kirchman *et al*. (2010) and Galand *et al*. (2010) are 1- to 5-fold higher than the number of sequences analyzed in this study, and rarefaction curves may be influenced by the different sequencing efforts.

Our results in rarefaction curves were not saturated even if more than 6500 sequences were retrieved in bacterial communities, indicating that it needs analysis of an increasing number of sequences to properly evaluate the bacteria diversity in northern SCS (Table 3). We purpose that rarefaction curves are not saturated over 15000 sequences here, higher than Roesch *et al*. (2007) reported that in some soil. He pointed out that rarefaction curves in some soil didn't plateau up to 5000-10000 sequences, as similar the number of sequences analyzed in this study (Roesch *et al*., 2007). It's generally accepted that the microbial diversity in the soil is higher than that in the ocean (Quince *et al*., 2008). Therefore, we conclude that bacterial diversity in northern SCS (and likely elsewhere) is much greater than previous estimates in other oceans based on conventional methods such as 16S rRNA gene clone libraries and T-RFLP.

**Archaeal diversity.** The richness estimates of 1000m- and 3000m-depth archaeal

libraries obtained with Chao 1 and ACE were similar and lower than the 100m layer, while the surface layer contained the highest archaeal richness values (Table 4). The 3000m-depth sample had the lowest Shannon value (2.75 versus over 3.00), indicating that the archaeal communities in shallower layers are more diverse than that in deeper layers in agreement with earlier research (Galand *et al*., 2009a). The evenness values estimated from 100m- and 3000m-depth archaeal V6 libraries (0.64 on average) were slightly lower than that from the 10m- and the 1000m-depth libraries (0.73 and 0.79, respectively). The average value of $C_{OTU}$ from the archaeal libraries was 97%, but the other coverage values (C) ranged from 68.2% to 82.9% in the archaeal libraries (Good's coverage values calculating with stricter criteria), indicating that the actual diversity in archaeal libraries is at least covered two-thirds. Figure 8b show that rarefaction curves calculated from the archaeal communities didn't reveal an asymptotic behavior. We suggest that archaeal rarefaction curves wouldn't go to a flat until 5000 sequences in the northern SCS.

Comparing to the Shannon values obtained from the Tyrrhenian Sea (Cono *et al*., 2009) and the North Water (Galand *et al*., 2009b), we obtained the higher Shannon values. This may indicate the archaeal communities in the northern SCS are more diverse than that in other oceans. However, this may be also due to the different sequencing effort because the number of sequences in our libraries is at least 6-fold higher than the number of sequences analyzed by Cono *et al*. (2009-13) and Galand *et al*. (2009b). Galand *et al*. (2009a) have studied the archaeal community structures in the Arctic Ocean by pyrosequencing and recovered more than 16,000 sequences per sampling site. Comparing to the rarefaction curves obtained by Galand *et al*. (2009a) at the number of sequences sampled ranging from 0 to 1720, the values of observed OTUs in the four SEATS libraries are lower than that in the Arctic Ocean, indicating that the

archaeal richness values are lower throughout the water column in the northern SCS than in the Arctic Ocean. It's noticeable that the number of sequences sampled by Galand *et al*. (2009a) is higher than that in this study by an order of magnitude, and rarefaction curves may be influenced by the different sequencing efforts.

Our results show that the diversity of bacterial communities is higher than that of archaeal communities (Table 3 and 4; Figure 8), which is consistent with previous studies as shown through clone library (Aller and Kemp, 2008) and pyrosequencing of hydrothermal vents samples (Huber *et al*., 2007).

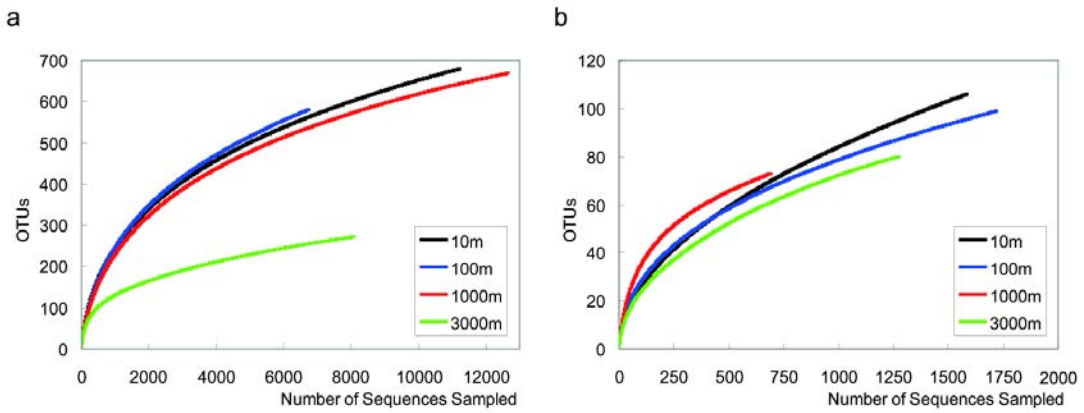


Figure 8. Analysis for species complexity. V6 sequences were aligned by the Clustal W package (version 2.10), and the pairwise matrices with the Jukes-Cantor correction were calculated after alignment. Rarefaction curves for bacterial and archaeal communities at four sampling depths were constructed by the DOTUR software package using the pairwise matrices. (a) Rarefaction curves for bacterial communities at 97% sequence identity level. (b) Rarefaction curves for archaeal communities at 97% sequence identity level.

### 3.3.3 Microbial community structures

**Bacterial communities.** The bacterial communities are dominated by the broad phylogenetic group, Phylum *Proteobacteria*, which accounted for about 68.24%, 91.46%, 95.57%, and 88.65% of all sequences in the 10m-, 100m-, 1000m-, and 3000m-depth datasets, respectively (Figure 9a). Phylum *Cyanobacteria* is relative abundant in the surface layer (about 20% of sequences in the 10m-depth library). Phyla *Bacteroides, Actinobacteria, Planctomyetes, Acidobacteria, Verrucomicrobia* are presented as minor groups. The patterns are similar in broad outline to those observed in other ocean such as the Sargasso Sea (Venter *et al*., 2004), the North Pacific Ocean (DeLong *et al*., 2006; Ma *et al*., 2009; Brown *et al*., 2009), the Baltic Sea (Andersson *et al*., 2010), the Mediterranean Sea (Feingersch *et al*., 2010), and the Arctic Ocean (Kirchman *et al*., 2010).

In Class-level groupings, the bacterial distributions are different between four depth samples. Classes *α*- and *γ-proteobacteria* are the top two abundant groups throughout the water column at the SEATS station in northern SCS (Figure 9a) except in the surface layer (10m), while *α-proteobacteria* is the most abundant group (~41.97%), followed by *Cyanobacteria* (~20.57%) and *γ-proteobacteria* (~19.88%). These results are in agreement with the previous study observed from the North Pacific Ocean by Brown *et al*. (2009).

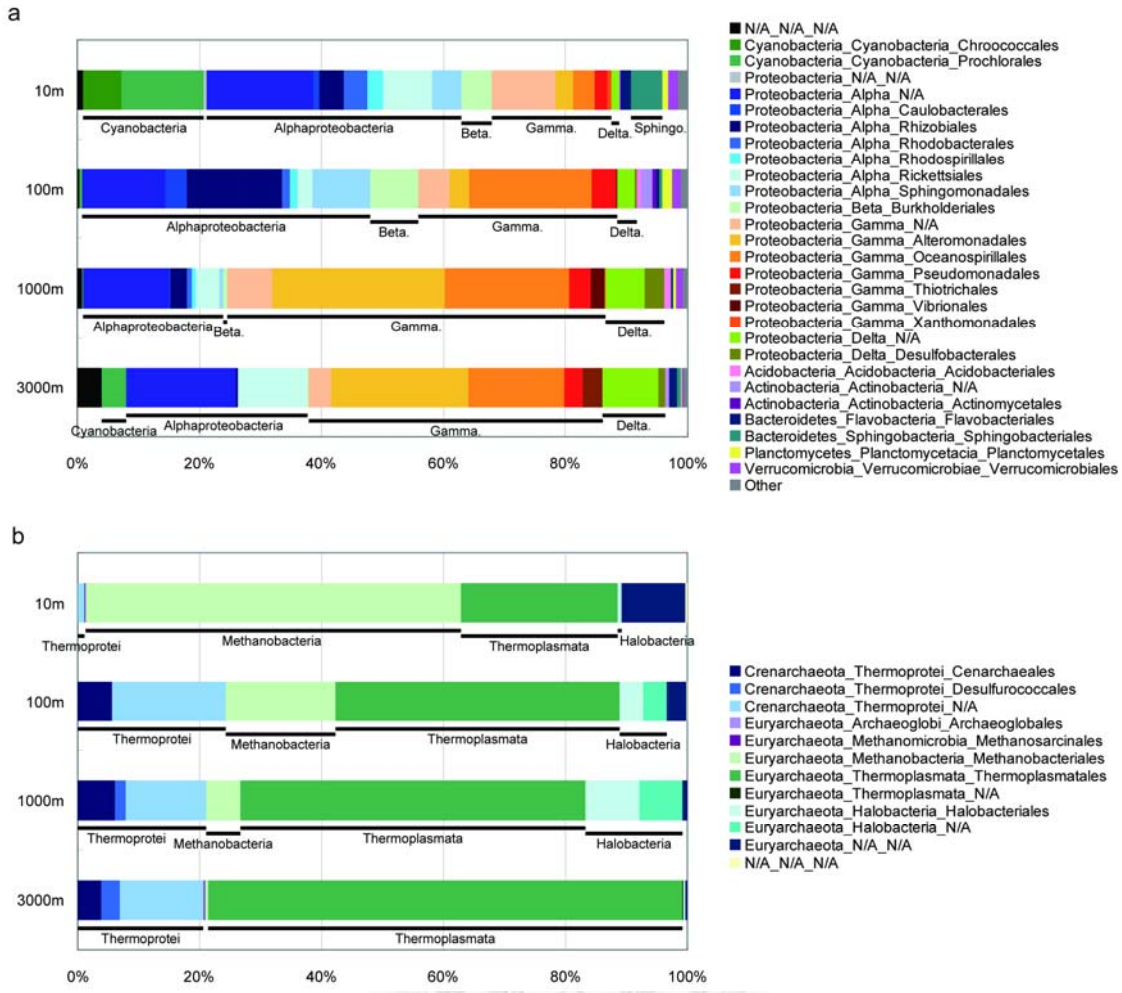


Figure 9. Taxonomic distributions of microbes identified in the V6 libraries. 100% stacked bar charts display the Phylum_Class_Order distribution for taxonomically assigned tags that presented over 0.5%; the remaining sequences are grouped into "Other". Abbreviations are N/A, not assigned; Alpha, *α-proteobacteria*; Beta, *β-proteobacteria*; Gamma, *γ-proteobacteria*; Delta, *δ-proteobacteria*; and Sphingo, *Sphingobacteria*. (a) The relative abundance of bacterial assemblages observed at 10m, 100m, 1000m, and 3000m depths. (b) The relative abundance of archaeal assemblages observed at 10m, 100m, 1000m, and 3000m depths.

Although the relative abundances of *α*- and *γ-proteobacteria* are high in all depths, at the finer taxonomic resolution, the compositions of Order-level groupings of *α*- and *γ-proteobacteria* vary in each depth. Taking *α-proteobacteria* as an example (Figure 9a), Order *Rickettsiales* are represented higher percents in 10m- and 3000m- depth layers than the middle two layers (more than 8% versus less than 4%). Orders *Sphingomonadales* and *Rhizobiales* show similar distribution patterns that they are more abundant at the 100m (*Sphingomonadales*: 9.57%; *Rhizobiales*: 15.65%) than that of 10m (*Sphingomonadales*: 4.84%; *Rhizobiales*: 3.98%), but occupy minor components at the deeper layers (below 3%). Besides, the unclassified alphaproteobacteria constitute ranging from 13.56% to 18.05% of the bacterial assemblages. As with *γ-proteobacteria* (Figure 9a), Order *Oceanospirillales* is abundant (15% or more) throughout the water column except for the sea surface (~3.47%). Order *Alteromonadales* is more abundant in the aphotic zone than in the euphotic zone (over 22% versus 3.5% or less).

More than 97% of the *β-proteobacteria* sequences are assigned to Order *Burkholderiales* (Figure 9a). Pommier *et al*. (2007) investigated the marine bacterial community structure globally from the Arctic Ocean, the North and South Atlantic Ocean, the North and South Pacific Ocean by 16S rRNA gene clone libraries, and Ma *et al*. (2009) revealed the vertical distribution and composition of bacteria in the Eastern Tropical North Pacific Ocean by 16S rRNA gene clone libraries. However, neither Pommier *et al*. (2007) nor Ma *et al*. (2009) observed *β-proteobacteria* which are inconsistent with our results. On the other hand, the studies on marine microbial community genomes from the Sargasso Sea, the North Pacific Ocean, and the Mediterranean Sea were obtained the reads that originated from *β-proteobacteria* (Venter *et al*., 2004; Martín-Cuadrado *et al*., 2007; Pham *et al*., 2008; Feingersch *et al*., 2010). The different results may be due to varied methods. The groups with low relative

abundance of total bacterial community such as betaproteobacteria are hard to be detected from fewer sequencing amount of 16S rRNA gene clone libraries compared to microbial community genome sequencing and pyrosequencing.

*β-proteobacteria* sequences were found in the recent studies from the North Pacific Ocean, the Arctic Ocean, and the Baltic Sea by pyrosequencing, the same method as this study (Brown *et al*., 2009; Kirchman *et al*., 2010; Galand *et al*., 2010; Andersson *et al*., 2010). Kirchman *et al*. (2010) observed that the relative abundance of total bacterial community of Order *Burkholderiales* was about 3.5% from the surface water in the Arctic Ocean, which is lower than our results in the upper two depths. However, Alonso-Sáez and his coworkers (2007) have found that *β-proteobacteria* in the Mediterranean Sea appeared at highly frequencies by the increase in the coastal and riverine discharges after the stormy period, and they suggested that this could be related to that *β-proteobacteria* is one of typical freshwater bacteria (Zwart *et al*., 2002). If that is the case, one of the possible reasons that caused the higher proportions of *β-proteobacteria* at the SEATS station is the offshore influence, because the Pearl River, the second largest river by volume in China, flows into the northern SCS in which the SEATS station located. Besides, most typhoons tend to form between May and November in the northwestern Pacific Ocean and have strong impact on the northern SCS, and our sampling in October is during the late typhoon seasons.

We also observed that Class *β-proteobacteria* is more abundant in the euphotic zone than in the aphotic zone in northern SCS, which is contrast to the results obtained by Brown *et al*. (2009) in North Pacific Ocean. The different patterns could be related to the different environmental factors, and it's reasonable that the relative abundances of the less abundant groups such as betaproteobacteria vary dramatically in different sampling sites (Galand *et al*., 2010).

In contrast to Class *β-proteobacteria*, the relative abundance of sequences from Class *δ-proteobacteria* increases with increasing depth, which is consistent with the observations from the North Pacific Ocean by Brown *et al*. (2009). The sequences affiliated to *δ-proteobacteria* are assigned to Order *Desulfobacterales* or the unclassified deltaproteobacteria grouping. The ratio of the unclassified deltaproteobacteria grouping increases with depth, while *Desulfobacterales* shows a peak (over 3%) at the 1000m-depth layer (Figure 9a).

Class *Cyanobacteria* accounted for about 20% of the total bacterial community in the surface layer, and the ratio sharply decreased to 0.74% in the 100m depth with less than 1% of the surface irradiance (Lee Chen *et al*., 2004; Chen *et al*., 2006). The relative abundant of *Cyanobacteria* is ~0.06% and ~4.04% of the total bacterial community in the 1000m layer and the 3000m depth, respectively. The subgroups of *Cyanobacteria* are Orders *Chroococcales* and *Prochlorales*. *Chroococcales* shows a peak (6.35%) at the sea surface, and decreases sharply in and below the 100m-depth layer. *Prochlorales* is also represented the highest ratio (13.42%) at the sea surface, and decreases to 0.43% and 0.09% at the 100m- and 1000m- depth datasets respectively, and the sequences of 3000m cyanobacteria all belonged to *Prochlorales* (Figure 9a).

It's known that cyanobacteria obtain their energy through photosynthesis, and we observed that Class *Cyanobacteria* is more abundant in the bathypelagic layer than in the middle two layers in contrast to our expectation. However, a recent survey of microbial population structures in the deep marine biosphere observed the appearance of *Cyanobacteria* (Huber *et al*., 2007), and in an earlier study, a sequence affiliated to *Cyanobacteria* was obtained from the 770m-depth library in the North Pacific Subtropical Gyre (Pham *et al*., 2008). Brown *et al*. (2009) also detected *Cyanobacteria* from the 800m-depth library and the 4400m-depth library in the North Pacific Ocean.

Therefore, the appearances of these photosynthetic organisms in the aphotic zone are not news and could be due to turbulent mixing, diffusion or other physical processes. We predicted that the relative abundance of *Cyanobacteria* should decrease with increasing depth through those physical processes; however, *Cyanobacteria* occupied nearly 4% of the 3000m-depth library and represented higher proportions than in the 1000m depth (less than 0.1%, Figure 9a). If the appearance of *Cyanobacteria* in the 3000m layer is not an episodic event, few questions emerge. Are they metabolically active organisms in the deep sea? If so, could they obtain their energy through the biofluorescent or live as heterotrophic bacteria? In order to answer these questions, more studies are needed.

We still found some minor groups from Figure 9a, such as Class *Sphingobacteria* accounting for about 5.17% of all sequences in the 10m-depth dataset, *Actinobacteria* contributing around 2.55% of all sequences in the 100m-depth dataset, and the unclassified bacteria reaching a relative abundance of *ca.* 3.96% of all sequences in the greatest depth. The remaining Classes are represented less than 1.7 percents.

**Archaeal communities.** As shown in Figure 9b, only two major archaeal lineages, Phyla *Euryarchaeota* and *Crenarchaeota*, were detected in this study. *Euryarchaeota* account for more than 75% of the archaeal V6 amplicons recovered from the entire 3000m span of the water column at the SEATS station in the northern SCS. In the surface layer, the relative abundance of archaeal sequences from *Crenarchaeota* is only around 1.14% (Figure 9b). *Crenarchaeota* suddenly increase in relative abundance at the 100-m depth layer (*ca.* 24.30%) and with a rather constant in relative abundance (*ca.* 20%) in the deeper layers (Figure 9b). Similar to earlier reports (Karner *et al*., 2001; Zhang *et al*., 2009), *Crenarchaeota* represent a higher ratio in the deep layers than in the upper layers. A surprising finding is that *Euryarchaeota* dominate throughout the water

column at SAETS station in the northern SCS, which is contrast to most of previous studies (Karner *et al*., 2001; Zhang *et al*., 2009). The dominance of *Euryarchaeota* has been revealed from 3000m depth at the Antarctic Polar Front (López-García *et al*., 2001a) and from the surface layer in the Arctic Ocean (Galand *et al*., 2009a). The reason for archaeal community variations is not fully understood, and more investigations are needed to resolve the variability in archaeal community structures in the open ocean.

In Class-level groupings, the compositions and the ratios of archaeal groups changed in different layers. *Methanobacteria* occupy a major component (~61.51%) at the sea surface, and decrease gradually with depth, and are only represented 0.39 percents in the 3000m-depth dataset (Figure 9b). On the other hand, *Thermoplasmata* shows the opposite trend (10m, ~25.80%; 100m, ~46.69%; 1000m, ~56.73%; and 3000m, ~78.00%). The relative abundance of Class *Thermoprotei* is 20% or more throughout the water column except for that of the sea surface (Figure 9b). *Halobacteria* have higher ratios at the middle two depths (~7.79% and ~15.77% in the 100m and 1000m libraries, respectively) (Figure 9b).

Comparing to Classes *α*- and *γ-proteobacteria*, the compositions of archaeal Order-level groupings of the same Class are similar even in different depths. For example, Order *Halobacteriales* and the unclassified halobacteria grouping are both belonged to *Halobacteria*, and they are represented about 3.90% at the 100m-depth layer; in addition, the relative abundance of *Halobacteriales* is almost equal to that of the unclassified halobacteria grouping at the 1000m depth layer (Figure 9b).

### 3.3.4 Profiles of unique sequences

**Bacterial V6 unique sequences.** Although the microbial compositions at the Order-level groupings even affiliated to the same class changed in different depths as

shown in Figure 9, the much more detailed results for the distribution of sequence-level groups were obtained from the analysis of unique sequences (Figure 10 and 11). For example, the relative abundance of Order *Rickettsiales* is more than 8% in the 10m and 3000m depth layers (Figure 9a), but unique sequences belonged to *Rickettsiales* peak differently between the surface layer and the deepest layer (Figure 10). Another example was observed among the unique sequences assingned to Order *Oceanospirillales* are different. In the 100m library, some groups affiliated to *Oceanospirillales* were detected and continued as major components in the deeper layers, but they were nearly undetectable in the sea surface. The bacterial groups contributed a peak in the pattern of unique sequences assingned to *Oceanospirillales* from the 10m library, but the groups almost disappeared from the 1000m and 3000m libraries while other groups were detected in deeper layers. Besides, some groups in Order *Burkholderiales* were only found in the 100m-depth layer. Similarly, some groups contributed peaks only in shallower depths, and some only highly presented in aphotic zone related to other Orders such as *Rhizobiales*, *Sphingomonadales*, *Desulfobacterales*, the unclassified deltaproteobacteria grouping, *Pseudomonadales*, and *Vibrionales* (Figure 10).

As showed in Figure 10, Order *Prochlorales* in the 10m-depth layer and the 3000m-depth layer are consisted by the same groups with identical sequences. The unclassified alpha- and gamma-proteobacteria groupings account for certain proportions of the bacterial assemblages (Figure 9a), and the pattern of unique sequences in each sample are different (Figure 10).

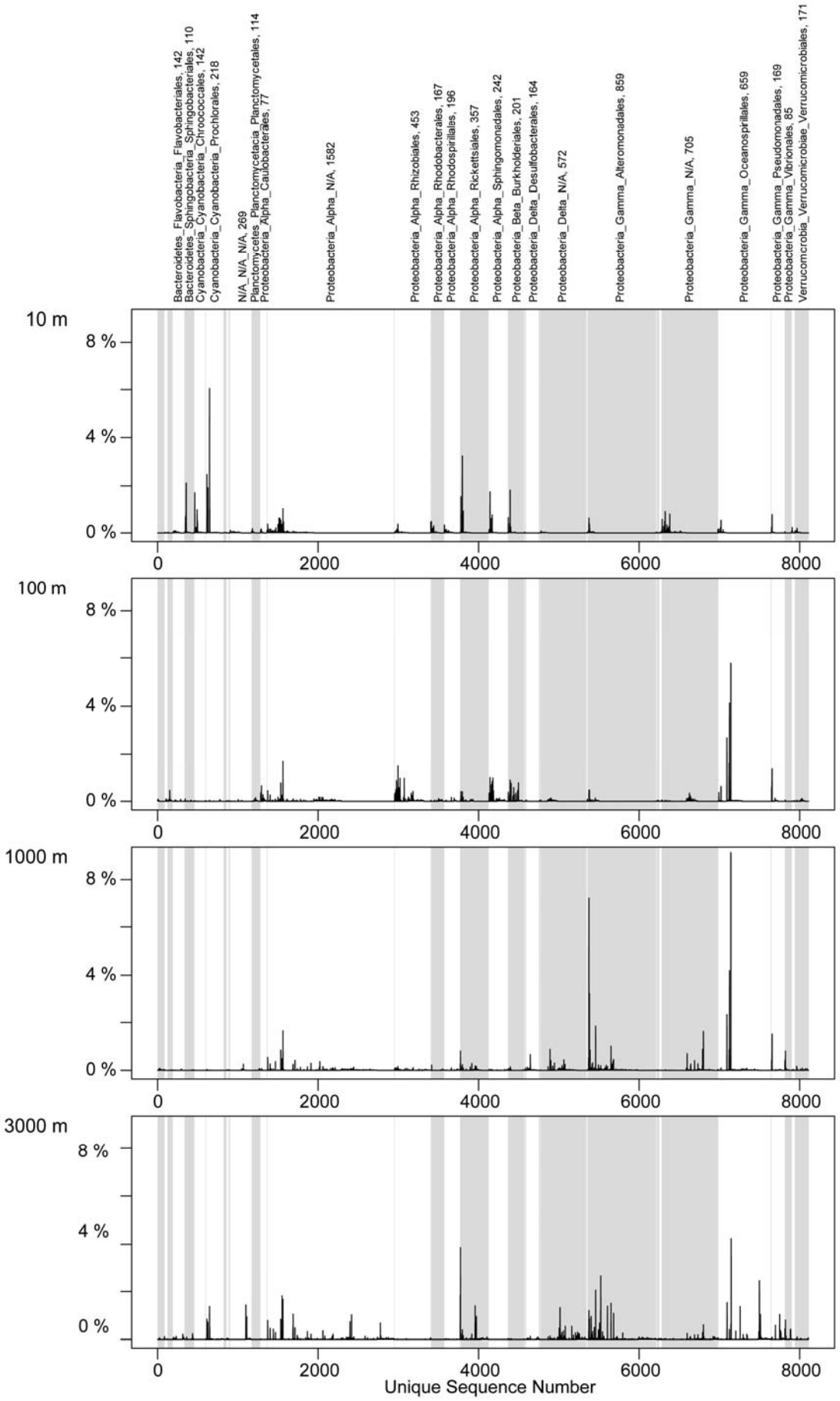
Bacteroidetes_Flavobacteriales, 142
Bacteroidetes_Sphingobacteriales, 110
Cyanobacteria_Chroococcales, 142
Cyanobacteria_Prochlorales, 218
N/A_N/A, 269
Planctomycetes_Planctomycetacia_Planctomycetales, 114
Proteobacteria_Alpha_Caulobacterales, 77
Proteobacteria_Alpha_N/A, 1582
Proteobacteria_Alpha_Rhizobiales, 453
Proteobacteria_Alpha_Rhodobacterales, 167
Proteobacteria_Alpha_Rhodospirillales, 196
Proteobacteria_Alpha_Rickettsiales, 357
Proteobacteria_Alpha_Sphingomonadales, 242
Proteobacteria_Beta_Burkholderiales, 201
Proteobacteria_Delta_Desulfobacterales, 164
Proteobacteria_Delta_N/A, 572
Proteobacteria_Gamma_Alteromonadales, 859
Proteobacteria_Gamma_N/A, 705
Proteobacteria_Gamma_Oceanospirillales, 659
Proteobacteria_Gamma_Pseudomonadales, 169
Proteobacteria_Gamma_Vibrionales, 85
Verrucomicrobia_Verrucomicrobiae_Verrucomicrobiales, 171
10 m
100 m
1000 m
3000 m
8 %
4 %
0 %
0
2000
4000
6000
8000
Unique Sequence Number

Figure 10. Profiles of the bacterial V6 unique sequences. The x-axis represents 8,111 unique sequences in the whole bacterial dataset in the arrangement according to their taxonomic assignments at the Order level. The y-axis represents the relative abundance of each unique sequence in each depth. The names of Orders contained more than 100 unique sequences are labeled on the top, following by the number of unique sequences assigned to the Order. To easily distinguish different Orders, we use the background color with gray and white. Abbreviations are N/A, not assigned; Alpha, *α-proteobacteria*; Beta, *β-proteobacteria*; Gamma, *γ-proteobacteria*; and Delta, *δ-proteobacteria*.

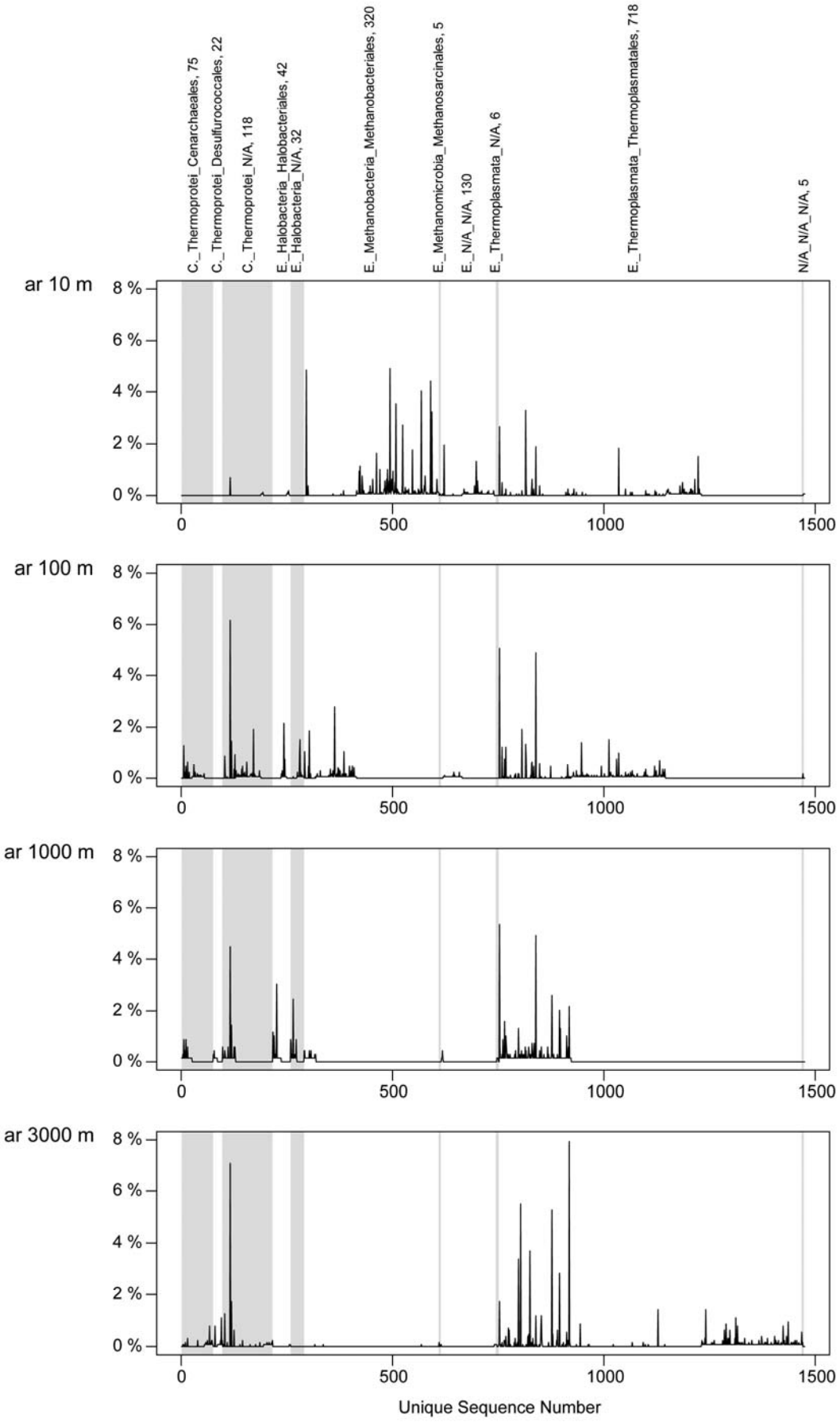
C._Thermoprotei_Cenarchaeales, 75
C._Thermoprotei_Desulfurococcales, 22
C._Thermoprotei_N/A, 118
E._Halobacteria_Halobacteriales, 42
E._Halobacteria_N/A, 32
E._Methanobacteria_Methanobacteriales, 320
E._Methanomicrobia_Methanosarcinales, 5
E._N/A_N/A, 130
E._Thermoplasmata_N/A, 6
E._Thermoplasmata_Thermoplasmatales, 718
N/A_N/A_N/A, 5
ar 10 m
ar 100 m
ar 1000 m
ar 3000 m
8 %
6 %
4 %
2 %
0 %
0
500
1000
1500
Unique Sequence Number

Figure 11. Profiles of the archaeal V6 unique sequences. The x-axis represents 1,475 unique sequences in the whole archaeal dataset in the arrangement according to their taxonomic assignments at the Order level. The y-axis represents the relative abundance of each unique sequence in each depth. The names of Orders are labeled on the top, following by the number of unique sequences assigned to the Order. To easily distinguish different Orders, we use the background color with gray and white. Abbreviations are N/A, not assigned; C., *Crenarchaeota*; and E., *Euryarchaeota*.

**Archaeal unique sequences.** Similar to the profiles of bacterial unique sequences, the results from the profiles of archaeal unique sequences show that some archaeal groups were found in the euphotic, aphotic zone, or only in a specific depth layer. From the different pattern in the profiles of Order *Methanobacteriales* at different depths, the following information was obtained. The relative abundance of *Methanobacteriales* decreases with depths in Figure 9b, and the decreaced ratio is not caused by the ralative abundance of the same groups decreasing with depth but caused by the combined outcomes from appearances and disappearances of varied groups belonged to this Order (Figure 11). In addition, some groups in *Thermoplasmatales* are depth-specific; for example, some of them were merely detected in the greastest depth layer, and other thermoplasmatales groups were found only in the 100m depth layer. Furthermore, the depth-specific patterns were found in many other Order-level groupings, such as *Cenarchaeales*, *Desulfurococcales*, *Halobacteriales*, the unclassified thermoprotei grouping, the unclassified halobacteria grouping, and the unclassified euryarchaeota grouping (Figure 11).

These results from the profiles of unique sequences illustrate that the microbial groupings may be different although the relative abundance of an Order or a higher taxonomic level is nearly constant throughout the water span of 3000m depth (Figure 10 and 11). The analyses of sequences provide much more detailed depth-related distributions than that of the 100% stacked bar charts at the Order-level.

### 3.3.5 Phylogenetic analysis of the V6 unique sequences

***α-proteobacteria.*** The 38 unique sequences selected divided into five named Orders and the unclassified supergroupings (Figure 12). The five named Orders are *Rhizobiales*, *Caulobacterales*, *Shingomonadales*, *Rhodobacterales*, and *Rickettiales*.

The unclassified supergroupings partitioned into four clusters as shown in Figure 8 (the SAR116 cluster and three unclassified alphaproteobacteria groupings). The sequence in the *Caulobacterales* cluster represented the highest proportion of the assemblages in the 100m depth layer and was related to members of Family *Caulobacteraceae*. All the sequences in the *Rhizobiales* cluster were affiliated with the genus *Methylobacterium* and also represented the highest proportion of the assemblages in the 100m depth layer. Members of the genus *Methylobacterium* are known as pink-pigmented facultative methylotrophic (PPFM) bacteria, and these methane-oxidizing bacteria can obtain the carbon for energy by oxidizing CH with O (Chistoserdova *et al.*, 2003; Kato *et al.*, 2008; Bosch *et al.*, 2009). Some of them can degrade important intermediates in the biogeochemical sulphur cycle such as methanethiol and dimethylsulphide (Anesti *et al.*, 2004) or formaldehyde, common found in the wastewater causing environmental pollution (Mirdamadi *et al.*, 2005). If our *Methylobacterium*-related sequences were originated from those bacteria with an capable of degraging hydrocarbon-containing matter, and if they were metabolic active at the SEATS station, the high relative abundances of these *Methylobacterium*-related sequences at the 100m depth could suggest that there probably exist hydrocarbon sources in this layer at the SEATS station and these bacteria play a role in the biogeochemical carbon and sulphur cycles in this layer.

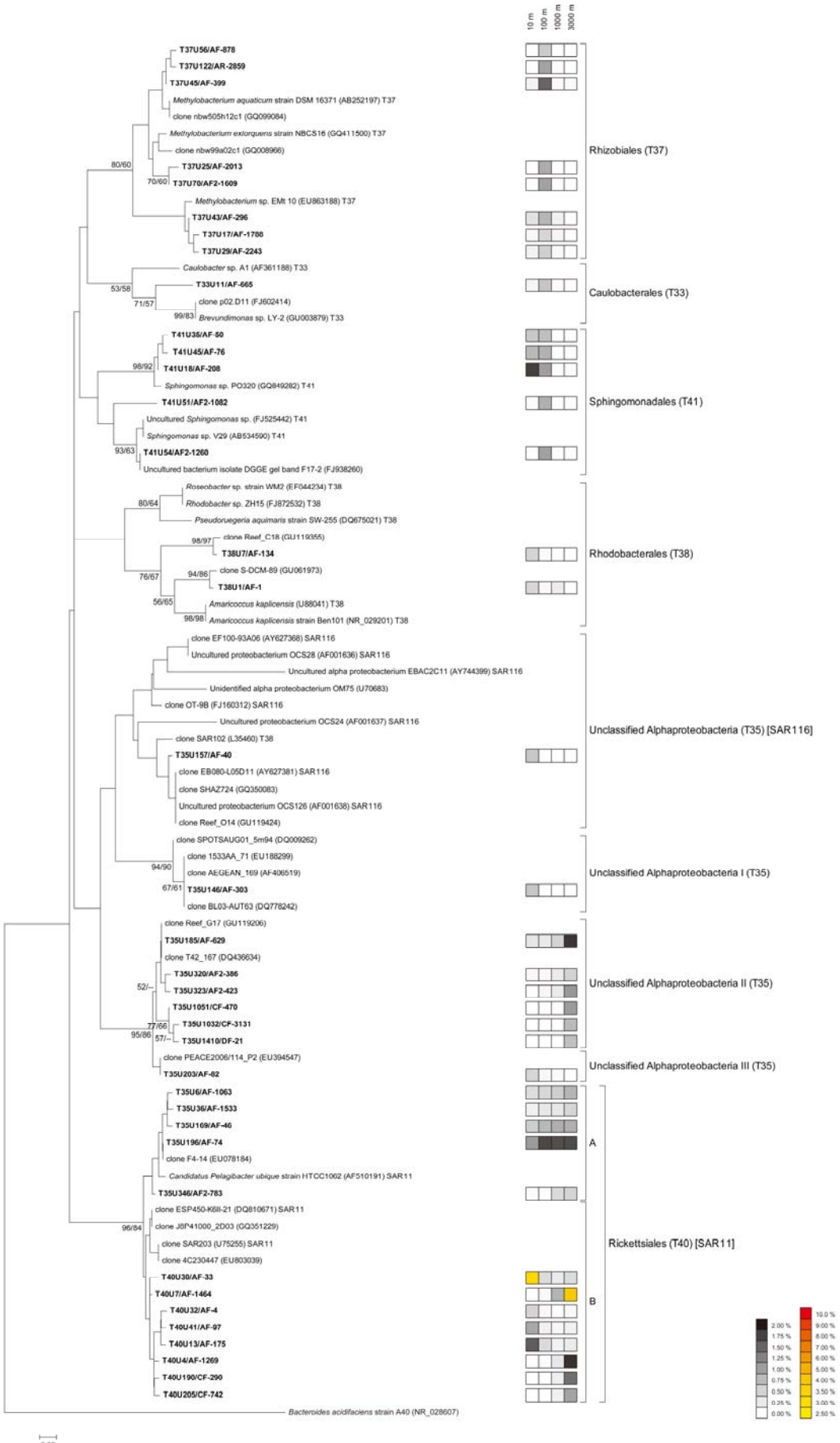
10 m
100 m
1000 m
3000 m
T37U56/AF-878
T37U122/AR-2859
T37U45/AF-399
Methylobacterium aquaticum strain DSM 16371 (AB252197) T37
clone nbw505h12c1 (GQ099084)
Methylobacterium extorquens strain NBCS16 (GQ411500) T37
clone nbw99a02c1 (GQ008966)
80/60
70/60
T37U25/AF-2013
T37U70/AF2-1609
Methylobacterium sp. EMt 10 (EU863188) T37
T37U43/AF-296
T37U17/AF-1786
T37U29/AF-2243
Rhizobiales (T37)
Caulobacter sp. A1 (AF361188) T33
53/58
T33U11/AF-665
71/57
clone p02.D11 (FJ602414)
99/83
Brevundimonas sp. LY-2 (GU003879) T33
Caulobacterales (T33)
T41U35/AF-50
T41U45/AF-76
98/92
T41U18/AF-208
Sphingomonas sp. PO320 (GQ849282) T41
T41U51/AF2-1082
Uncultured Sphingomonas sp. (FJ525442) T41
Sphingomonas sp. V29 (AB534590) T41
93/63
T41U54/AF2-1260
Uncultured bacterium isolate DGGE gel band F17-2 (FJ938260)
Sphingomonadales (T41)
Roseobacter sp. strain WM2 (EF044234) T38
80/64
Rhodobacter sp. ZH15 (FJ872532) T38
Pseudoruegeria aquimaris strain SW-255 (DQ675021) T38
98/97
clone Reef_C18 (GU119355)
T38U7/AF-134
76/67
94/86
clone S-DCM-89 (GU061973)
T38U1/AF-1
56/65
Amaricoccus kaplicensis (U88041) T38
98/98
Amaricoccus kaplicensis strain Ben101 (NR_029201) T38
Rhodobacterales (T38)
clone EF100-93A06 (AY627368) SAR116
Uncultured proteobacterium OCS28 (AF001636) SAR116
Uncultured alpha proteobacterium EBAC2C11 (AY744399) SAR116
Unidentified alpha proteobacterium OM75 (U70683)
clone OT-9B (FJ160312) SAR116
Uncultured proteobacterium OCS24 (AF001637) SAR116
clone SAR102 (L35460) T38
T35U157/AF-40
clone EB080-L05D11 (AY627381) SAR116
clone SHAZ724 (GQ350083)
Uncultured proteobacterium OCS126 (AF001638) SAR116
clone Reef_O14 (GU119424)
Unclassified Alphaproteobacteria (T35) [SAR116]
clone SPOTSAUG01_5m94 (DQ009262)
clone 1533AA_71 (EU188299)
94/90
clone AEGEAN_169 (AF406519)
67/61
T35U146/AF-303
clone BL03-AUT63 (DQ778242)
Unclassified Alphaproteobacteria I (T35)
clone Reef_G17 (GU119206)
T35U185/AF-629
clone T42_167 (DQ436634)
T35U320/AF2-386
52/--
T35U323/AF2-423
T35U1051/CF-470
77/66
T35U1032/CF-3131
95/86
57/--
T35U1410/DF-21
Unclassified Alphaproteobacteria II (T35)
clone PEACE2006/114_P2 (EU394547)
T35U203/AF-82
Unclassified Alphaproteobacteria III (T35)
T35U6/AF-1063
T35U36/AF-1533
T35U169/AF-46
T35U196/AF-74
A
clone F4-14 (EU078184)
Candidatus Pelagibacter ubique strain HTCC1062 (AF510191) SAR11
T35U346/AF2-783
clone ESP450-K6II-21 (DQ810671) SAR11
96/84
clone J8P41000_2D03 (GQ351229)
clone SAR203 (U75255) SAR11
clone 4C230447 (EU803039)
Rickettsiales (T40) [SAR11]
T40U30/AF-33
T40U7/AF-1464
B
T40U32/AF-4
T40U41/AF-97
T40U113/AF-175
T40U4/AF-1269
T40U190/CF-290
T40U205/CF-742
Bacteroides acidifaciens strain A40 (NR_028607)
0.02
2.00 %
1.75 %
1.50 %
1.25 %
1.00 %
0.75 %
0.50 %
0.25 %
0.00 %
10.0 %
9.00 %
8.00 %
7.00 %
6.00 %
5.00 %
4.00 %
3.50 %
3.00 %
2.50 %

Figure 12. Phylogenetic tree of *α-proteobacteria* V6 sequences. The tree is constructed by Jukes-Cantor (JC) distances with neighbor-joining (NJ) and maximum parsimony (MP) methods. Bootstrap values above 50% (NJ/MP method; 1000 replications) are shown at each node. Scale bar represents the nucleotide substitution percentage of NJ method. V6 sequences from this study are indicated in boldface and designated as the beginning of T*n*U*n*, in which T is an abbreviation of taxonomy, the first *n* represents the number of the taxonomy we assigned, U is an abbreviation of unique sequence, and the second *n* represents the number of the unique sequence we assigned. A representative V6 sequence of each unique sequence from this study is behind T*n*U*n* and slash. Accessions numbers of representative references retrieved from Genbank are shown in parentheses, and followed by the taxonomic codename (T*n*) according to the NCBI website description. Gray-scale and color-scale values in the squares indicate the occurrence frequency of the unique sequence in the corresponding libraries. *Bacteroides acidifaciens* strain A40 is used as the outgroup.

Two sequences in the *Rhodobacterales* cluster were related to the species *Amaricoccus kaplicensis*, isolated from activated sludge biomass (Maszenan *et al*., 1997). The *Shingomonadales* cluster could be further subdivided into two groups, and sequences in one group were abundant both in the 10m and 100m depths while sequences in the other group accounted for a higher ratio of the assemblages only in the 100m-depth layer. The sequences in the *Shingomonadales* cluster were affiliated with the genus *Sphingomonas*. Members of *Sphingomonas* are gram-negative aerobic bacteria and have the abilities to survive in oligotrophic niches. Some of them are potential microbial agents for bioremediation because of their abilities in metabolizing complex organic pollutants such as carbofuran, chlorinated biphenyls, and polychlorophenols (Yim *et al.*, 2010 and references therein). If our shingomonadales sequences were originated from the bacteria with an capable of metabolizing complex organic pollutants, and if they were metabolic active at the SEATS station, the high occurances of these shingomonadales sequences both at the depths of 10m and 100m suggest that there might be organic matter in the upper layers at the SEATS station probably due to the riverine and coastal runoff.

The ubiquitous SAR11 cluster consisted of two clades: a clade contained the sequences with “T35” as the beginning failed to assign to any named Order (clade A), and the other included those sequences assigned to *Rickettiales* (clade B). Sequences in the clade A were found throughout the water column with higher relative abundances in aphotic zones and affiliated with *Pleagibacter ubique* strain HTCC1062. However, the relative abundances of some sequences in the clade B were higher in the euphotic zones although they were detected in all libraries, and other sequences also in the clade B were only found in the deep two depths and accounted for higher proportions in the 3000m depth. These may support that the SAR11 is a board grouping likely constituted more

than one taxonomic species or even genus (Alonso-Sáez *et al*., 2007).

Through phylogenetic analyses, the results illustrate that the unclassified alphaproteobacteria occupied nearly 20% of the assemblages in the sea surface (Figure 9a) is actually made up of five groups, the SAR116 cluster, three unclassified alphaproteobacteria groupings, and some sequences in the clade A of the SAR11 cluster (Figure 12). The sequence T35U203/AF-82 related to the clone that obtained from the surface layer and formed the unclassified alphaproteobacteria III cluster (Oostende *et al*., 2008), which differed from the unclassified alphaproteobacteria II cluster of mainly bathypelagic layer sequences that represented a certain ratio in the unclassified alphaproteobacteria of the 3000m library in Figure 9a. Not only the sequences in the unclassified alphaproteobacteria II cluster but also that in the SAR11 cluster occupied largely in the unclassified alphaproteobacteria population of the meso- and bathy-pelagic layers (Figure 9a).

Our observations of the surface SAR116 sequences are similar to earlier results that the SAR116 clade was detected from the surface water of various oceans (Giovannoni *et al*., 1990; Britschgi and Giovannoni, 1991; Mullins *et al*., 1995; Giovannoni *et al*., 2000; Bano and Hollibaugh, 2002; Schauer *et al*., 2003; Pham *et al*., 2008). Besides, a representative of the SAR116 clade know as *Candidatus Puniceispirillum marinum* IMCC1322 was recently sequenced by Oh *et al*. (2010), and their results showed that its genome contained a gene for proteorhodopsin (Oh *et al*., 2010). Thus, this group is not only regarded as one kind of ubiquitous bacterioplanktons but also a proteorhodopsin-containing group.

***β-* and *δ- proteobacteria.*** As mentioned before, more than 97% of the *β-proteobacteria* sequences were assigned to Order *Burkholderiales*. In Figure 13, the *Burkholderiales* cluster mainly upper layers could partitioned into two groups: the

sequence affiliated with the genus *Herbaspirillum*, which differed from sequences clustered with sequences of the genera *Cupriavidus* and *Ralstonia* (a later synonym of the genus *Cupriavidus*) (La Duc *et al*., 2003; Vandamme and Coenye, 2004). Members of the genus *Cupriavidus* are well known as metal-resistant bacteria (Goris *et al*., 2001), and many of then can mineralize aromatic pollutants (Borde *et al*., 2003) and Trichloroethylene (TCE), a common soil and groundwater contaminant in industrialized countries (Nakamura *et al*., 2000 and references therein). *Cupriavidus basilensis* HMF14 can remove inhibitors from lignocellulosic hydrolysate under aerobic conditions (Wierckx *et al*., 2010). *Cupriavidus basilensis* JF1 is able to metabolize bisphenol A which is hazardous to humans (Fischer *et al*., 2010).

The sequences belonged to Class *δ-proteobacteria* of mainly aphotic zone sequences were grouped into at least two separate lineages, Order *Desulfobacterales* and the SAR324 cluster. The depth-stratified distribution was observed for the *Desulfobacterales* cluster; the sequence related with the clone obtained from 770m in the Pacific Ocean had the highest ratio of the assemblages in 1000m depth, and a clade of sequences found mainly in bathypelagic layer were clustered with the 4000m clone (HF4000_41A16). The 3000m sequences in the previous clade are phylogenetic closer to each other and failed to assign to any named order.

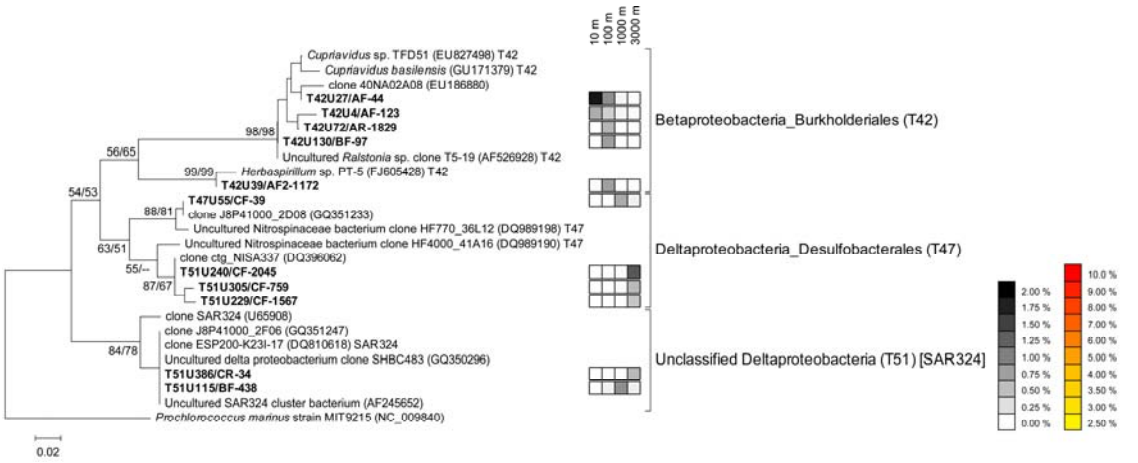


Figure 13. Phylogenetic tree of *β*- and *δ*- *proteobacteria* V6 sequences. The tree is constructed by Jukes-Cantor (JC) distances with neighbor-joining (NJ) and maximum parsimony (MP) methods. Bootstrap values above 50% (NJ/MP method; 1000 replications) are shown at each node. Scale bar represents the nucleotide substitution percentage of NJ method. V6 sequences from this study are indicated in boldface and designated as the beginning of T*n*U*n*, in which T is an abbreviation of taxonomy, the first *n* represents the number of the taxonomy we assigned, U is an abbreviation of unique sequence, and the second *n* represents the number of the unique sequence we assigned. A representative V6 sequence of each unique sequence from this study is behind T*n*U*n* and slash. Accessions numbers of representative references retrieved from Genbank are shown in parentheses, and followed by the taxonomic codename (T*n*) according to the NCBI website description. Gray-scale and color-scale values in the squares indicate the occurrence frequency of the unique sequence in the corresponding libraries. *Prochlorococcus marinus* str. MIT 9215 is used as the outgroup.

***γ-proteobacteria.*** Figure 14 shows the phylogenetic tree of sequences in Order *Alteromonadales* composed of five clades and the depth-related subpartitioning. Sequences in the Clade M were obtained from all depths but represented lower relative abundances in upper depths than that in the 3000m depth, while sequences in the clade N affiliated with the sequence that belonged to the genus *Moritella* were also highly represented in deeper depths but could not be detected in the sea surface. The relative abundances of those sequences in the clades O and P were all highest in the mesopelagic layer, and the relative abundances of all sequences in the clade Q were also highest in the bathypelagic layer. Besides, sequences in the clades O and P were found throughout the water column, while sequences in the clades Q could not be detected in the upper two depths.

The clade O sequence *Alteromonas* sp. WH189 has 99% sequence identity with *Alteromonas macleodii* AltDE (GenBank ID: CP001103), which is a deep ecotype of *Alteromonas macleodii* obtained from a depth of 1000m in the South Adriatic Sea basin (Ivars-Martinez *et al*., 2008). The clade P sequence *Alteromonas* sp. VS-63 has 98% sequence identity with *Alteromonas macleodii* AltDE, and three reference sequences in the clade M have 97% sequence identity with *Alteromonas macleodii* AltDE. The occurrence frequencies of our sequences of clades O and P are highest in the 1000m depth in agreement with the source of *Alteromonas macleodii* AltDE in the previous study (Ivars-Martinez *et al*., 2008). Sequences in the clade M have the lowest sequence identity with *Alteromonas macleodii* AltDE among these three clades, and the relative abundances of them are all highest near the sea bottom, suggesting that they may be a "deeper ecotype" of *Alteromonas macleodii* in the northern SCS.

Our sequences in the clade N are related to *Moritella* sp. JAM-GA22. The sequence *Moritella* sp. JAM-GA22 has 99% sequence identity with *Moritella viscosa*

strain NVI 88/478 (GenBank ID: NR_028880) and *Moritella japonica* strain DSK1 (GenBank ID: NR_025847). *Moritella viscosa*, well known as the pathogen of "winter ulcer" in farmed Atlantic salmon and cod, is a psychrophilic bacterium capable of surviving and reproducing in an oligotrophic and cold marine environment (Tunsjø *et al*., 2007). *Moritella japonica* strain DSK1 isolated from the Japan Trench at 6356m depth is a barophilic, halophilic, and psychrophilic bacterium (Nogi *et al*., 1998). Besides, a type strain of the genus *Moritella* is *Moritella marina*, one of the most common psychrophilic microbes from the ocean (Colwell and Morita, 1964). Our observations that the relative abundances of the clade N sequences are high in the deep-two layers are similar to the previous studies (Colwell and Morita, 1964; Nogi *et al*., 1998; Tunsjø *et al*., 2007).

The clade Q sequences are related to the uncultured bacterium clone F22 (GenBank ID: AY375116). According to the description on the NCBI website (http://www.ncbi.nlm.nih.gov/), the sequence of the uncultured bacterium clone F22 was obtained from a depth of 1901m sediments at the Western Pacific Ocean (Zeng *et al*., 2005). Through comparing to the public database, the clone F22 has 96% sequence similarity to cultivable bacterium belonged to the genus *Colwellia*. All characterized members of the genus *Colwellia* were obtained from stable cold marine environments, and these bacterial groups are strictly psychrophilic (Methé *et al*., 2005 and references therein). In consistent with the earlier knowledge, the relative abundances of the clade Q sequences were all highest in the bathypelagic layer, while they could not be detected in the upper two depths, suggesting that the clade Q sequences could be originated from novel psychrophilic bacteria in the deep northern SCS.

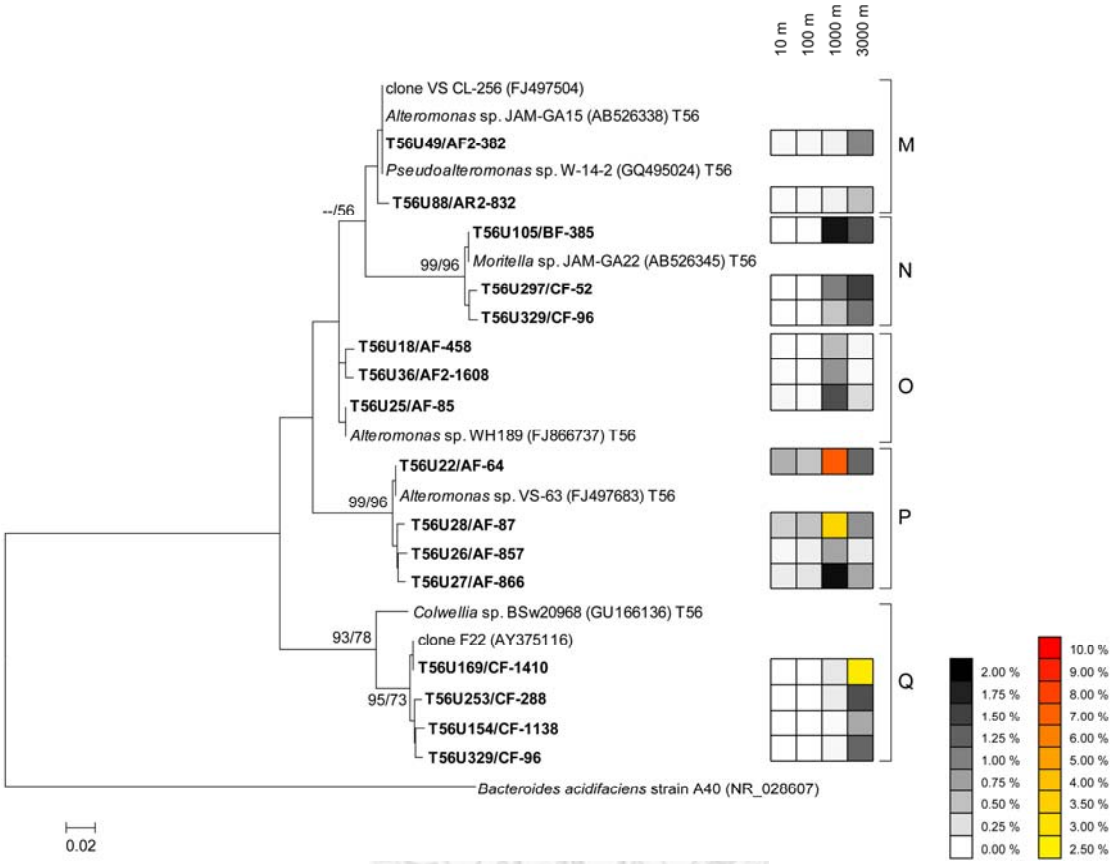


Figure 14. Phylogenetic tree of *Alteromonadales* (*γ-proteobacteria*) V6 sequences. The tree is constructed by Jukes-Cantor (JC) distances with neighbor-joining (NJ) and maximum parsimony (MP) methods. Bootstrap values above 50% (NJ/MP method; 1000 replications) are shown at each node. Scale bar represents the nucleotide substitution percentage of NJ method. V6 sequences from this study are indicated in boldface and designated as the beginning of T*n*U*n*, in which T is an abbreviation of taxonomy, the first *n* represents the number of the taxonomy we assigned, U is an abbreviation of unique sequence, and the second *n* represents the number of the unique sequence we assigned. A representative V6 sequence of each unique sequence from this study is behind T*n*U*n* and slash. Accessions numbers of representative references retrieved from Genbank are shown in parentheses, and followed by the taxonomic codename (T*n*) according to the NCBI website description. Gray-scale and color-scale values in the squares indicate the occurrence frequency of the unique sequence in the corresponding libraries. *Bacteroides acidifaciens* strain A40 is used as the outgroup.

The *Oceanospirillales* tree also had the depth-specific distribution and partitioned into two subclusters (Figure 15). The sequences related with the genera *Alcanivorax* and *Marinospirillum* formed a cluster (clades Y and Z), which separated from the mainly 3000m sequences hardly obtained in the euphoic zone and the 1000m-depth layer (clades W and X). The first cluster could be further divided in two clades, one contained sequences detected more frequencies in the 10m and 100m depths (clade Y), while the sequences in clade Z are found in the middle two layers. The two sequences in the clade Y are close related to *Alcanivorax borkumensis* strain Sk2, a hydrocarbon-degrading and surfactant-producing marine bacterium (Yakimov *et al*., 1998). If the clade Y sequences were originated from those functionally active bacteria previously mentioned, the high occurrence frequencies of these clade Y sequences in the euphotic zone could also suggest the existences of hydrocarbon pollutants in the northern SCS as a result of terrestrial and freshwater run-off, defusing matters from offshore oil refineries, coastal oil production, shipping activities and accidental spillage of fuels and other petroleum products, corresponding to our observations of the *Burkholderiales*, *Rhizobales*, *Caulobacterales*, and *Sphingomonadales* sequences.

The clade Z sequences are phylogenetically related to *Marinospirillum* sp. ANL-isoa. Although *Marinospirillum* sp. ANL-isoa couldn't utilize nitriles, the toxic organic compounds, they grew rapidly on the products of nitrile hydrolysis, amides, carboxylic acids and other simple compounds, thus acting as a scavenger of products released by the nitrile-degrading organisms (Sorokin *et al*., 2007). If the clade Z sequences were originated from the metabolic active bacteria that grow rapidly on those previously mentioned matters, the high relative abundances of the clade Z sequences in the middle two layers could be related to these characteristics of *Marinospirillum* sp. ANL-isoa, providing them advantages when competing with other bacteria at the

middle-two layers in the northern SCS.

Most representative relatives of the sequences in the clades W and X are belonged to the family *Oceanospirillaceae* such as *Marinomonas* sp. S3727, *Marinomonas pontica*, *Thalassolituus oleivarans*, and *Oleispira* sp. Gap-e-97. In consistent with that *Oceanospirillaceae* is one of the families of psychrophilic microorganisms (Schauer *et al*., 2010), our sequences of the clades W and X represented the highest frequencies in the relative abundance in the 3000 library.

Another depth-related subdivition observed in the tree of Order *Pseudomonadales* (Figure 16) comprised of two clades (clade I and J). The clade I sequences are more abundant in the middle two depths, while the sequences in the clade J could not be found in the surface layer and were over 0.5% of the total sequences in the 3000m library. The relative abundances of the clade J sequences (i.e., closely related to *Psychrobacter celer*) are highest in the greatest depth, and this observation is resemble other reports that members of the genus *Psychrobacter* could survive in extremely cold habitats, such as ice, soil, and sediments in the polar region, and in deep sea environments (Bozal *et al*., 2003; Yoon *et al*., 2005a, 2005b; Ayala-del-Río *et al*., 2010).

Corresponding to our other finding in many Orders such as *Burkholderiales* and *Sphingomonadales*, the clade I sequences which had the high occurrence frequencies in the middle two depths are related some functionally important bacteria such as *Pseudomonas otitidis* strain AOLR13 and *Pseudomonas* sp. P7(2009b). These two strains are both isolated from the aerobic granular sludges, compact and strong microbial aggregates, which are able to treat high-strength and toxic wastewater efficiently (Adav *et al*., 2010a, 2010b).

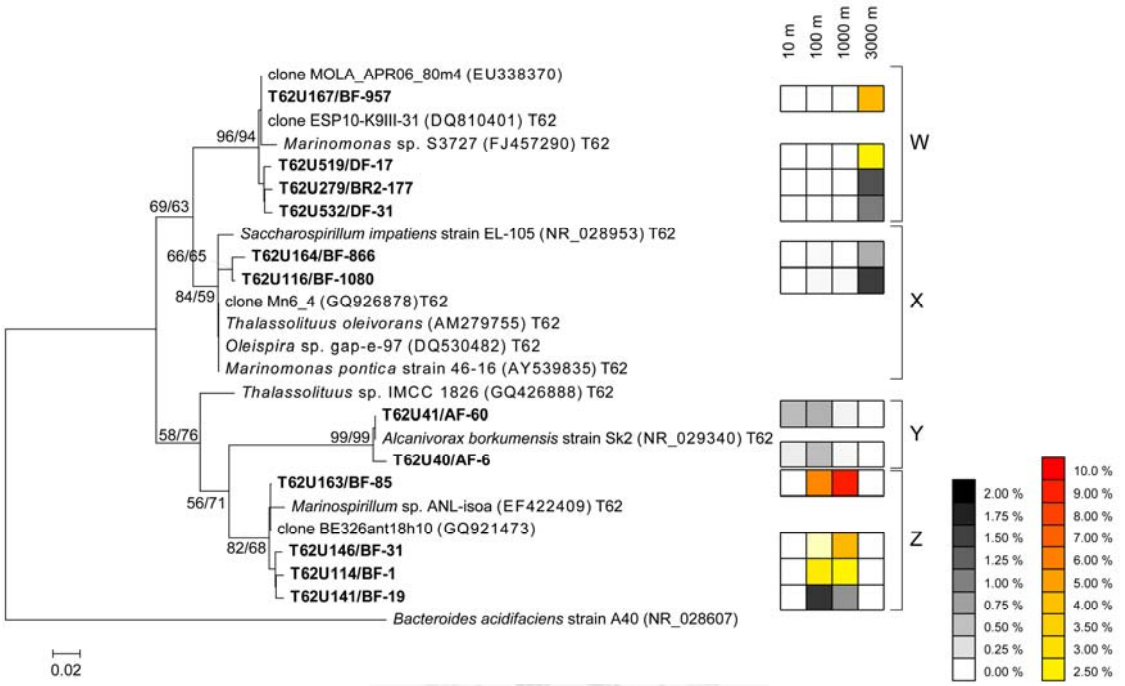


Figure 15. Phylogenetic tree of *Oceanospirillales* (*γ-proteobacteria*) V6 sequences. The tree is constructed by Jukes-Cantor (JC) distances with neighbor-joining (NJ) and maximum parsimony (MP) methods. Bootstrap values above 50% (NJ/MP method; 1000 replications) are shown at each node. Scale bar represents the nucleotide substitution percentage of NJ method. V6 sequences from this study are indicated in boldface and designated as the beginning of T*n*U*n*, in which T is an abbreviation of taxonomy, the first *n* represents the number of the taxonomy we assigned, U is an abbreviation of unique sequence, and the second *n* represents the number of the unique sequence we assigned. A representative V6 sequence of each unique sequence from this study is behind TnUn and slash. Accessions numbers of representative references retrieved from Genbank are shown in parentheses, and followed by the taxonomic codename (T*n*) according to the NCBI website description. Gray-scale and color-scale values in the squares indicate the occurrence frequency of the unique sequence in the corresponding libraries. *Bacteroides acidifaciens* strain A40 is used as the outgroup.

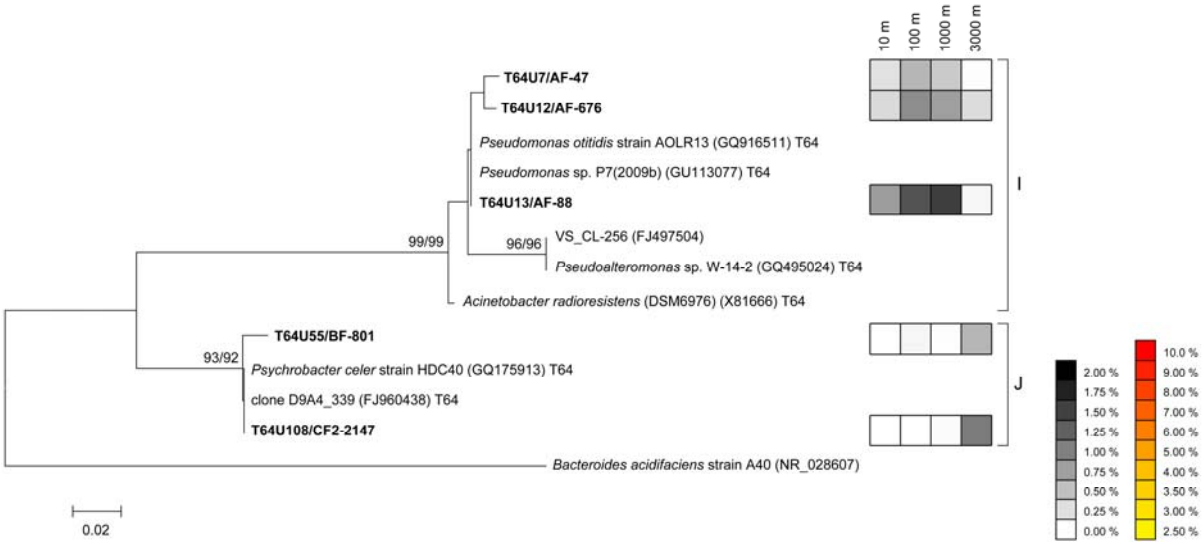


Figure 16. Phylogenetic tree of *Pseudomonadales* (*γ-proteobacteria*) V6 sequences. The tree is constructed by Jukes-Cantor (JC) distances with neighbor-joining (NJ) and maximum parsimony (MP) methods. Bootstrap values above 50% (NJ/MP method; 1000 replications) are shown at each node. Scale bar represents the nucleotide substitution percentage of NJ method. V6 sequences from this study are indicated in boldface and designated as the beginning of T*n*U*n*, in which T is an abbreviation of taxonomy, the first *n* represents the number of the taxonomy we assigned, U is an abbreviation of unique sequence, and the second *n* represents the number of the unique sequence we assigned. A representative V6 sequence of each unique sequence from this study is behind T*n*U*n* and slash. Accessions numbers of representative references retrieved from Genbank are shown in parentheses, and followed by the taxonomic codename (T*n*) according to the NCBI website description. Gray-scale and color-scale values in the squares indicate the occurrence frequency of the unique sequence in the corresponding libraries. *Bacteroides acidifaciens* strain A40 is used as the outgroup.

The sequences of Order *Vibrionales* and the unclassified gammaproteobacteria groupings were constructed in the same tree as shown in Figure 17. The *Vibrionales* cluster contained the sequence represented about 0.8% of the bacterial assemblages in the meso- and bathy- pelagic layers. In contrast with the close relative, *Vibrio* sp. S842, isolated from the sea surface (Gram *et al*., in press), the high ratios in the relative abundance of the *Vibrionales* sequences occurred at the deep two depths. However, some members of the genus *Vibrio* have been reported as psychrophilic bacteria such as *Vibrio diabolicus* (isolated from a deep-sea hydrothermal vent polychaete annelid), *Vibrio marinus*, *Vibrio wodanis*, *Vibrio pomeroyi*, and *Vibrio kanaloae* (the last two strains were isolated from sea water and marine animals; Colwell and Morita, 1964; Raguénès *et al*., 1997; DeLong *et al*., 1997; Lunder *et al*., 2000; Thompson *et al*., 2003; Bartlett, 2006). Thus, the *Vibrionales* sequence could be originated from one of the cold-adaptive bacteria or a novel strain that live in the deep northern SCS.

The sequences in the clades SAR86, PDB-OTU7, and KI89A were abundant in the surface layer but hardly detected in the three deeper depths. In contrast, the DHB-2 clade contained sequences accounted for higher proportions in and below the 1000m depth. The SAR86 clade well known as the proteorhodopsin-containing bacterium appeared mainly in the euphotic zone, and was observed in the upper layer from other oceans (Béjà *et al*., 2000; Man *et al*., 2003; Sabehi *et al*., 2003, 2004). Our observations in the distributions of these unclassified gammaproteobacteria groupings are consistent with the previous study (Pham *et al*., 2008).

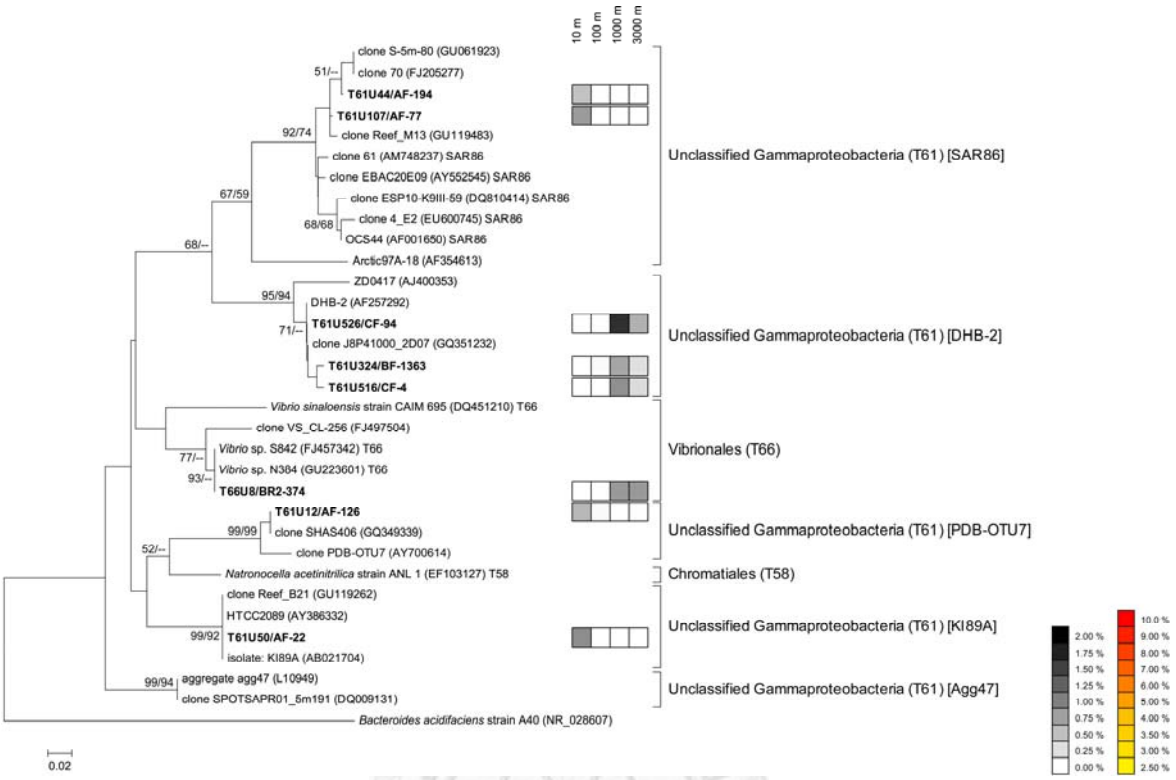


Figure 17. Phylogenetic tree of *Vibrionales* and the unclassified gammaproteobacteria V6 sequences. The tree is constructed by Jukes-Cantor (JC) distances with neighbor-joining (NJ) and maximum parsimony (MP) methods. Bootstrap values above 50% (NJ/MP method; 1000 replications) are shown at each node. Scale bar represents the nucleotide substitution percentage of NJ method. V6 sequences from this study are indicated in boldface and designated as the beginning of T*n*U*n*, in which T is an abbreviation of taxonomy, the first *n* represents the number of the taxonomy we assigned, U is an abbreviation of unique sequence, and the second *n* represents the number of the unique sequence we assigned. A representative V6 sequence of each unique sequence from this study is behind T*n*U*n* and slash. Accessions numbers of representative references retrieved from Genbank are shown in parentheses, and followed by the taxonomic codename (T*n*) according to the NCBI website description. Gray-scale and color-scale values in the squares indicate the occurrence frequency of the unique sequence in the corresponding libraries. *Bacteroides acidifaciens* strain A40 is used as the outgroup.

**Non-proteobacterial sequences** A total of 16 non-proteobacterial sequences were selected and grouped into three Orders (*Sphingobacteriales*, *Prochlorales*, and *Croococcales*), Class *Actinobacteria*, and one unclassified bacteria cluster (Figure 18). The *Sphingobacteriales* sequences and those cyanobacteria groups were mainly found in the surface layer. As illustrated before, the *Prochlorales* sequences in the 10m and 3000m depths are the same sequences. The cluster of unclassified bacteria comprised four sequences which were phylogenetically distinct from any known bacteria, showing that those 3000m sequences are likely originated from the unique and novel species unreported in current literature.

***Crenarchaeota*.** Among 16 SEATS archaeal sequences in the marine crenarchaeota group I cluster, only the sequence "T3U19ar1000m-39" was widely detected throughout the water column, and five wide-spread sequences were obtained from 100m, 1000m, and 3000m depths (T3U6ar1000m-173, T3U22ar1000m-53, T3U28ar1000m-67, T1U15ar1000m-510, and T1U6ar1000m-198; Figure 19). T3U74ar100m-4, T1U30ar100m-140, and T3U58ar100m-16 were the 100m sequences, and the sequences T3U14ar1000m-333 and T3U1ar1000m-101 were only detected at the 1000m depth, while the two sequences, T1U11ar1000m-434 and T3U30ar1000m-69, were found in both 100m and 1000m depths. Besides, the remaining SEATS sequences in this cluster were represented with the highest relative abundance of archaeal assemblages in the greatest depth layer, including two 3000m sequences (T2U20ar3000m-69 and T1U67ar3000m-689) and the T2U5ar1000m-309 sequence appearing in both 1000m and 3000m depths. However, there is no clearly depth-specific subpartitioning in the marine crenarchaeota group I cluster.

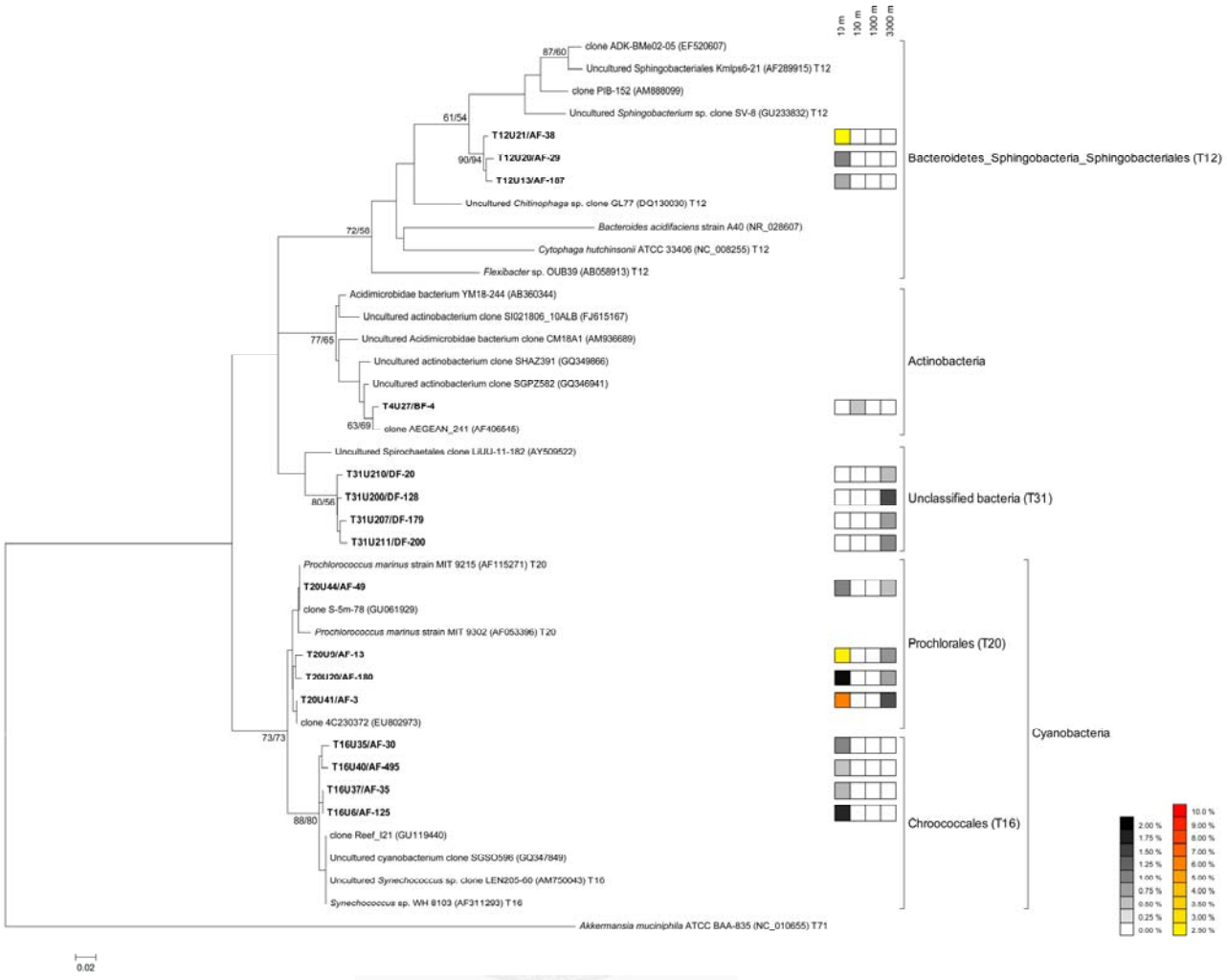


Figure 18. Phylogenetic tree of other non-proteobacteria V6 sequences. The tree is constructed by Jukes-Cantor (JC) distances with neighbor-joining (NJ) and maximum parsimony (MP) methods. Bootstrap values above 50% (NJ/MP method; 1000 replications) are shown at each node. Scale bar represents the nucleotide substitution percentage of NJ method. V6 sequences from this study are indicated in boldface and designated as the beginning of T*n*U*n*, in which T is an abbreviation of taxonomy, the first *n* represents the number of the taxonomy we assigned, U is an abbreviation of unique sequence, and the second *n* represents the number of the unique sequence we assigned. A representative V6 sequence of each unique sequence from this study is behind T*n*U*n* and slash. Accessions numbers of representative references retrieved from Genbank are shown in parentheses, and followed by the taxonomic codename (T*n*) according to the NCBI website description. Gray-scale and color-scale values in the squares indicate the occurrence frequency of the unique sequence in the corresponding libraries. *Akkermansia muciniphila* ATCC BAA-835 is used as the outgroup.

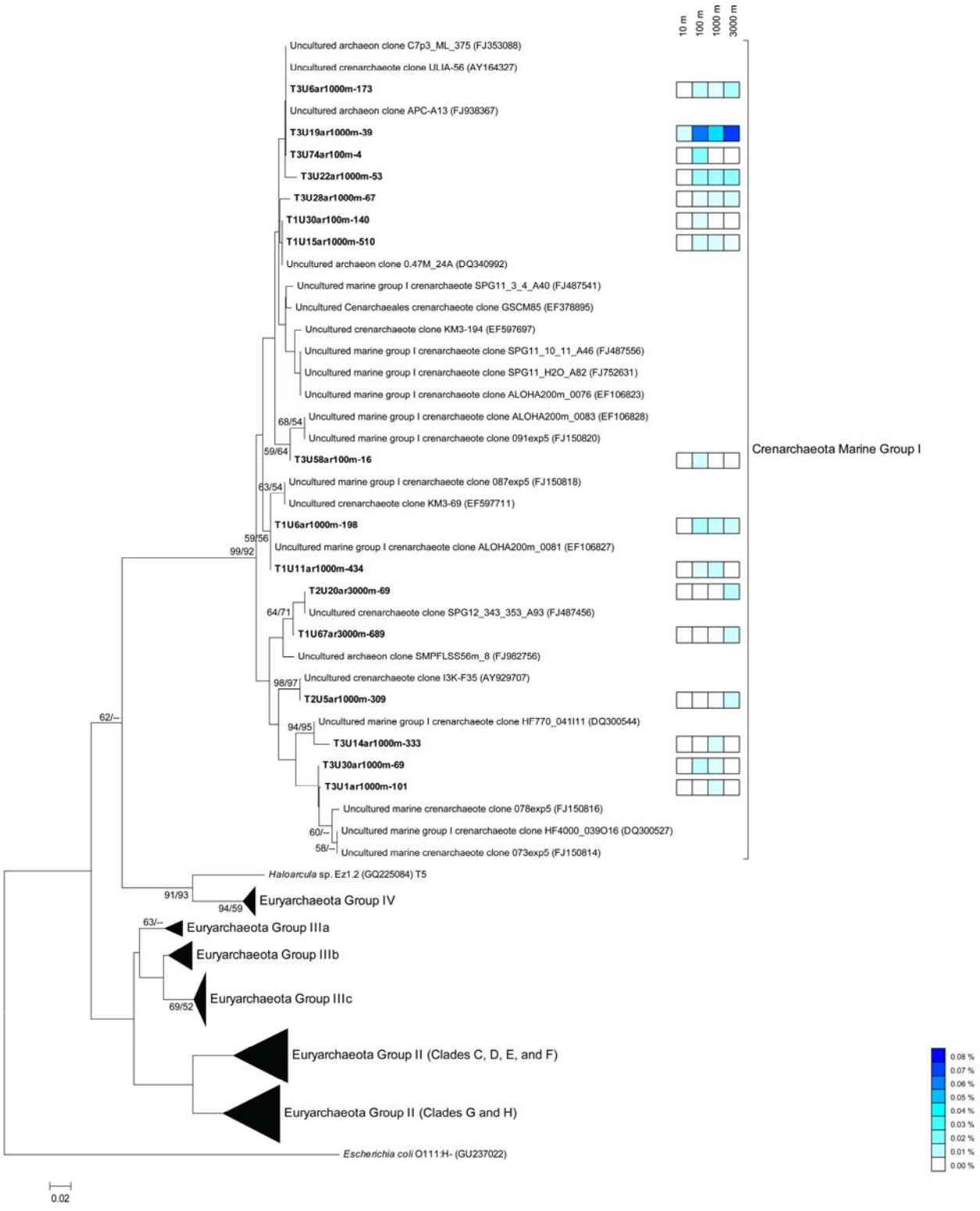
10 m
100 m
1000 m
3000 m
Uncultured archaeon clone C7p3_ML_375 (FJ353088)
Uncultured crenarchaeote clone ULIA-56 (AY164327)
T3U6ar1000m-173
Uncultured archaeon clone APC-A13 (FJ938367)
T3U19ar1000m-39
T3U74ar100m-4
T3U22ar1000m-53
T3U28ar1000m-67
T1U30ar100m-140
T1U15ar1000m-510
Uncultured archaeon clone 0.47M_24A (DQ340992)
Uncultured marine group I crenarchaeote SPG11_3_4_A40 (FJ487541)
Uncultured Cenarchaeales crenarchaeote clone GSCM85 (EF378895)
Uncultured crenarchaeote clone KM3-194 (EF597697)
Uncultured marine group I crenarchaeote clone SPG11_10_11_A46 (FJ487556)
Uncultured marine group I crenarchaeote clone SPG11_H2O_A82 (FJ752631)
Uncultured marine group I crenarchaeote clone ALOHA200m_0076 (EF106823)
68/54
Uncultured marine group I crenarchaeote clone ALOHA200m_0083 (EF106828)
Uncultured marine group I crenarchaeote clone 091exp5 (FJ150820)
59/64
T3U58ar100m-16
Crenarchaeota Marine Group I
63/54
Uncultured marine group I crenarchaeote clone 087exp5 (FJ150818)
Uncultured crenarchaeote clone KM3-69 (EF597711)
T1U6ar1000m-198
59/56
99/92
Uncultured marine group I crenarchaeote clone ALOHA200m_0081 (EF106827)
T1U11ar1000m-434
T2U20ar3000m-69
64/71
Uncultured crenarchaeote clone SPG12_343_353_A93 (FJ487456)
T1U67ar3000m-689
Uncultured archaeon clone SMPFLSS56m_8 (FJ982756)
98/97
Uncultured crenarchaeote clone I3K-F35 (AY929707)
T2U5ar1000m-309
62/--
94/95
Uncultured marine group I crenarchaeote clone HF770_041I11 (DQ300544)
T3U14ar1000m-333
T3U30ar1000m-69
T3U1ar1000m-101
Uncultured marine crenarchaeote clone 078exp5 (FJ150816)
60/--
Uncultured marine group I crenarchaeote clone HF4000_039O16 (DQ300527)
58/--
Uncultured marine crenarchaeote clone 073exp5 (FJ150814)
Haloarcula sp. Ez1.2 (GQ225084) T5
91/93
94/59
Euryarchaeota Group IV
63/--
Euryarchaeota Group IIIa
Euryarchaeota Group IIIb
69/52
Euryarchaeota Group IIIc
Euryarchaeota Group II (Clades C, D, E, and F)
Euryarchaeota Group II (Clades G and H)
Escherichia coli O111:H- (GU237022)
0.02
0.08 %
0.07 %
0.06 %
0.05 %
0.04 %
0.03 %
0.02 %
0.01 %
0.00 %

Figure 19. Phylogenetic tree of *Crenarchaeota* V6 sequences. The tree is constructed by Jukes-Cantor (JC) distances with neighbor-joining (NJ) and maximum parsimony (MP) methods. Bootstrap values above 50% (NJ/MP method; 1000 replications) are shown at each node. Scale bar represents the nucleotide substitution percentage of NJ method. V6 sequences from this study are indicated in boldface and designated as the beginning of T*n*U*n*, in which T is an abbreviation of taxonomy, the first *n* represents the number of the taxonomy we assigned, U is an abbreviation of unique sequence, and the second *n* represents the number of the unique sequence we assigned. A representative V6 sequence of each unique sequence from this study is behind T*n*U*n*. Accessions numbers of representative references retrieved from Genbank are shown in parentheses, and followed by the taxonomic codename (T*n*) according to the NCBI website description. The color-scale values in the squares indicate the occurrence frequency of the unique sequence in the corresponding libraries. *Escherichia coli* O111:H- is used as the outgroup.

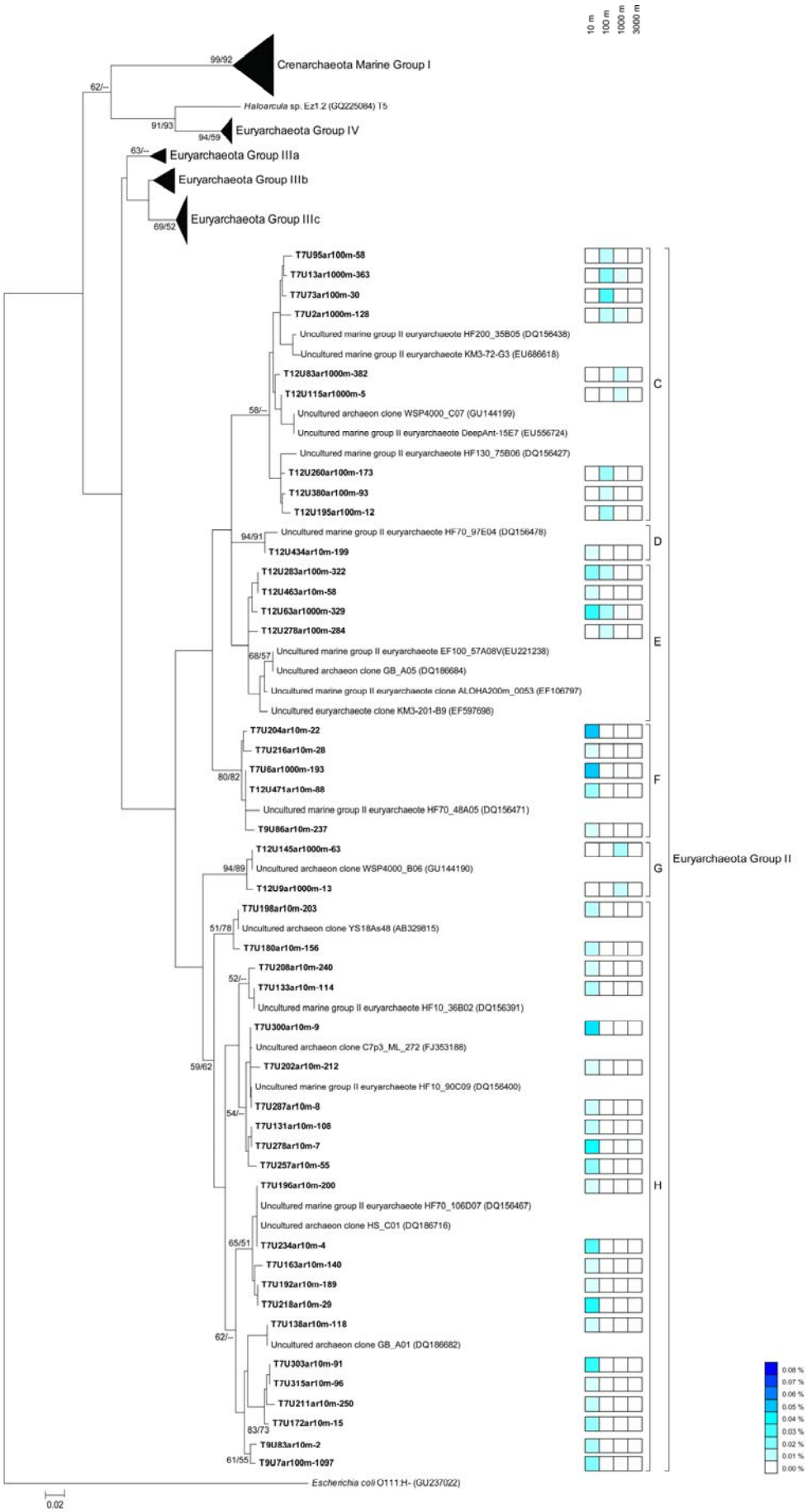
10 m
100 m
1000 m
3000 m
Crenarchaeota Marine Group I
Haloarcula sp. Ez1.2 (GQ225084) T5
Euryarchaeota Group IV
Euryarchaeota Group IIIa
Euryarchaeota Group IIIb
Euryarchaeota Group IIIc
T7U95ar100m-58
T7U13ar1000m-363
T7U73ar100m-30
T7U2ar1000m-128
Uncultured marine group II euryarchaeote HF200_35B05 (DQ156438)
Uncultured marine group II euryarchaeote KM3-72-G3 (EU686618)
T12U83ar1000m-382
T12U115ar1000m-5
Uncultured archaeon clone WSP4000_C07 (GU144199)
Uncultured marine group II euryarchaeote DeepAnt-15E7 (EU556724)
Uncultured marine group II euryarchaeote HF130_75B06 (DQ156427)
T12U260ar100m-173
T12U380ar100m-93
T12U195ar100m-12
C
Uncultured marine group II euryarchaeote HF70_97E04 (DQ156478)
T12U434ar10m-199
D
T12U283ar100m-322
T12U463ar10m-58
T12U63ar1000m-329
T12U278ar100m-284
Uncultured marine group II euryarchaeote EF100_57A08V(EU221238)
Uncultured archaeon clone GB_A05 (DQ186684)
Uncultured marine group II euryarchaeote clone ALOHA200m_0053 (EF106797)
Uncultured euryarchaeote clone KM3-201-B9 (EF597698)
E
T7U204ar10m-22
T7U216ar10m-28
T7U6ar1000m-193
T12U471ar10m-88
Uncultured marine group II euryarchaeote HF70_48A05 (DQ156471)
T9U86ar10m-237
F
T12U145ar1000m-63
Uncultured archaeon clone WSP4000_B06 (GU144190)
T12U9ar1000m-13
G
Euryarchaeota Group II
T7U198ar10m-203
Uncultured archaeon clone YS18As48 (AB329815)
T7U180ar10m-156
T7U208ar10m-240
T7U133ar10m-114
Uncultured marine group II euryarchaeote HF10_36B02 (DQ156391)
T7U300ar10m-9
Uncultured archaeon clone C7p3_ML_272 (FJ353188)
T7U202ar10m-212
Uncultured marine group II euryarchaeote HF10_90C09 (DQ156400)
T7U287ar10m-8
T7U131ar10m-108
T7U278ar10m-7
T7U257ar10m-55
T7U196ar10m-200
Uncultured marine group II euryarchaeote HF70_106D07 (DQ156467)
Uncultured archaeon clone HS_C01 (DQ186716)
T7U234ar10m-4
T7U163ar10m-140
T7U192ar10m-189
T7U218ar10m-29
T7U138ar10m-118
Uncultured archaeon clone GB_A01 (DQ186682)
T7U303ar10m-91
T7U315ar10m-96
T7U211ar10m-250
T7U172ar10m-15
T9U83ar10m-2
T9U7ar100m-1097
H
0.08 %
0.07 %
0.06 %
0.05 %
0.04 %
0.03 %
0.02 %
0.01 %
0.00 %
Escherichia coli O111:H- (GU237022)
0.02

Figure 20. Phylogenetic tree of the marine euryarchaeota group II V6 sequences. The tree is constructed by Jukes-Cantor (JC) distances with neighbor-joining (NJ) and maximum parsimony (MP) methods. Bootstrap values above 50% (NJ/MP method; 1000 replications) are shown at each node. Scale bar represents the nucleotide substitution percentage of NJ method. V6 sequences from this study are indicated in boldface and designated as the beginning of T*n*U*n*, in which T is an abbreviation of taxonomy, the first *n* represents the number of the taxonomy we assigned, U is an abbreviation of unique sequence, and the second *n* represents the number of the unique sequence we assigned. A representative V6 sequence of each unique sequence from this study is behind T*n*U*n*. Accessions numbers of representative references retrieved from Genbank are shown in parentheses, and followed by the taxonomic code (T*n*) according to the NCBI website description. The color-scale values in the squares indicate the occurrence frequency of the unique sequence in the corresponding libraries. *Escherichia coli* O111:H- is used as the outgroup.

***Euryarchaeota*.** In contrast to the marine crenarchaeota group I cluster, the depth-related distribution is observed for the marine euryarchaeota group II cluster. Fourty-three sequences were grouped into at least six subgroups (clades C-H, Figure 20). The sequences in the clade C were obtained from the 100m, 1000m, or both depths, which differed from the 10m sequence related to the HF70_97E04 sequence (clade D) and from the subgroup of mainly upper water column sequences (clade E). The clade C sequences were related with the relatives obtained from the depths between 130m and 3000m in other oceans. The HF70_97E04 sequence was retrieved from the large-insert DNA clone libraries in the North Pacific Subtropical Gyre and contained a bacterial-like proteorhodopsin (Frigaard *et al*., 2006). Through the phylogenetic analysis of SSU rRNA genes, the HF70_97E04 sequence was diverse from other sequences with no proteorhodopsin genes in the study of Frigaard *et al*. (2006). In agreement with the observations of Frigaard *et al*. (2006), the HF70_97E04 sequence and our sequence T12U434ar10m-199 form a clade (clade D) and are diverse from other sequences, indicating that our sequence T12U434ar10m-199 could be originated from a proteorhodopsin-containing euryarchaeota. In the clades F and H, the sequences were all represented with the highest ratio in the surface layer and were related with clones retrieved from the euphotic zone. The 1000m sequences, T12U145ar1000m-63 and T12U9ar1000m-13, were clustered into a seperate clade (affiliated with the clone WSP4000_B06; clade G) and phylogenetically distinct from any known arhcaea. On the other hand, these 43 sequences in the marine euryarchaeota group II cluster were assigned to *Thermoplasmatales* (T12), *Methanobacteriales* (T7) and to the unclassified euryarchaeota grouping (T9), illustrating that archaeal taxonomic assignments are still unsettled or that the resolution of V6 analysis is not enough to seperate the sequences in these two named orders.

The marine euryarchaeota group III cluster was comprised of 39 sequences assigned to *Thermoplasmatales* and could be further divided into three subdivisions, a subdivision of mainly 3000m sequences (clade IIIa), a subdivision of eurybathyal sequences (clade IIIb), and a subdivision of mainly meso- and bathy- pelagic layers sequences (clade IIIc, Figure 21). The sum of the relative abundances of these 39 sequences was ~7.51%, ~18.43%, ~29.38% and ~47.92% in the 10m, 100m, 1000m, and 3000m library, respectively.

A total of 13 sequences in the marine euryarchaeota group IV cluster were all assigned to *Halobacteria* after the comparisons to RDP and NCBI databases. Except for the T6U7ar1000m-24 sequence (appeared in the middle two layers), the remaining 12 sequences in the marine Euryarchaeota group IV cluster were only obtained either from 100m or 1000m depths and corresponding to the depth from which their relatives has been retrieved (Figure 22). For example, three sequences were related with DH148-Y16 that was isolated from deep-water masses of the South Atlantic (López-García *et al*., 2001b), and these three sequences were highly represented at the 1000m depth, as they were undetectable in other three depths. The sequence affiliated with the clone ALOHA200m_0026 (Mincer *et al*., 2007), for another example, were only observed at the 100m depth.

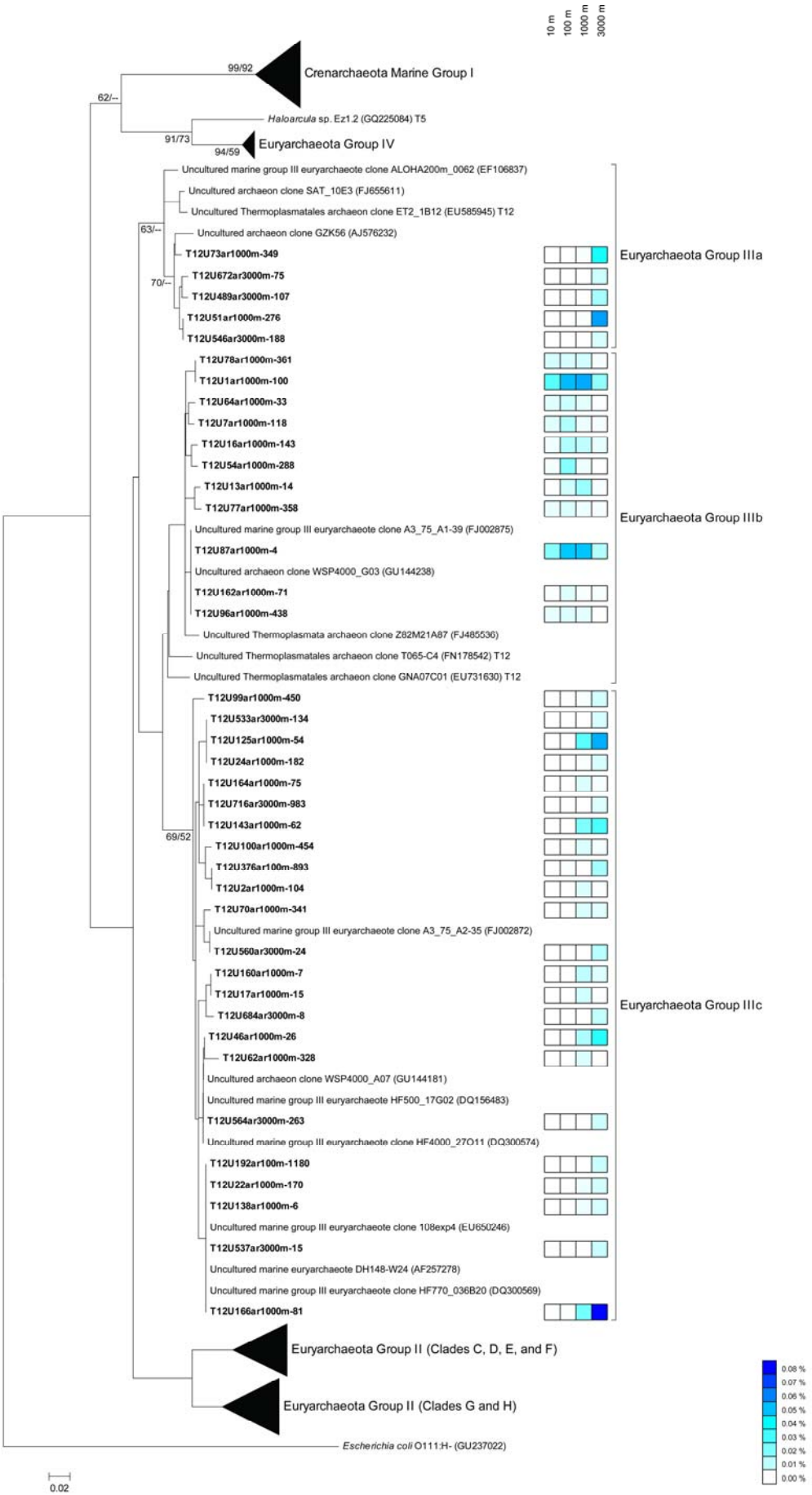
10 m
100 m
1000 m
3000 m
99/92
Crenarchaeota Marine Group I
62/--
Haloarcula sp. Ez1.2 (GQ225084) T5
91/73
94/59
Euryarchaeota Group IV
Uncultured marine group III euryarchaeote clone ALOHA200m_0062 (EF106837)
Uncultured archaeon clone SAT_10E3 (FJ655611)
Uncultured Thermoplasmatales archaeon clone ET2_1B12 (EU585945) T12
63/--
Uncultured archaeon clone GZK56 (AJ576232)
T12U73ar1000m-349
70/--
T12U672ar3000m-75
T12U489ar3000m-107
T12U51ar1000m-276
T12U546ar3000m-188
Euryarchaeota Group IIIa
T12U78ar1000m-361
T12U1ar1000m-100
T12U64ar1000m-33
T12U7ar1000m-118
T12U16ar1000m-143
T12U54ar1000m-288
T12U13ar1000m-14
T12U77ar1000m-358
Uncultured marine group III euryarchaeote clone A3_75_A1-39 (FJ002875)
T12U87ar1000m-4
Uncultured archaeon clone WSP4000_G03 (GU144238)
T12U162ar1000m-71
T12U96ar1000m-438
Uncultured Thermoplasmata archaeon clone Z82M21A87 (FJ485536)
Uncultured Thermoplasmatales archaeon clone T065-C4 (FN178542) T12
Uncultured Thermoplasmatales archaeon clone GNA07C01 (EU731630) T12
Euryarchaeota Group IIIb
T12U99ar1000m-450
T12U533ar3000m-134
T12U125ar1000m-54
T12U24ar1000m-182
T12U164ar1000m-75
T12U716ar3000m-983
T12U143ar1000m-62
69/52
T12U100ar1000m-454
T12U376ar100m-893
T12U2ar1000m-104
T12U70ar1000m-341
Uncultured marine group III euryarchaeote clone A3_75_A2-35 (FJ002872)
T12U560ar3000m-24
T12U160ar1000m-7
T12U17ar1000m-15
T12U684ar3000m-8
T12U46ar1000m-26
T12U62ar1000m-328
Uncultured archaeon clone WSP4000_A07 (GU144181)
Uncultured marine group III euryarchaeote HF500_17G02 (DQ156483)
T12U564ar3000m-263
Uncultured marine group III euryarchaeote clone HF4000_27O11 (DQ300574)
T12U192ar100m-1180
T12U22ar1000m-170
T12U138ar1000m-6
Uncultured marine group III euryarchaeote clone 108exp4 (EU650246)
T12U537ar3000m-15
Uncultured marine euryarchaeote DH148-W24 (AF257278)
Uncultured marine group III euryarchaeote clone HF770_036B20 (DQ300569)
T12U166ar1000m-81
Euryarchaeota Group IIIc
Euryarchaeota Group II (Clades C, D, E, and F)
Euryarchaeota Group II (Clades G and H)
Escherichia coli O111:H- (GU237022)
0.02
0.08 %
0.07 %
0.06 %
0.05 %
0.04 %
0.03 %
0.02 %
0.01 %
0.00 %

Figure 21. Phylogenetic tree of the marine euryarchaeota group III V6 sequences. The tree is constructed by JC distances with NJ and MP methods using MEGA4 program (Tamura *et al.*, 2007-7). Bootstrap values above 50% (NJ/MP method; 1000 replications) are shown at each node. Scale bar represents the nucleotide substitution percentage of NJ method. V6 sequences from this study are indicated in boldface and designated as the beginning of T*n*U*n*, in which T is an abbreviation of taxonomy, the first *n* represents the number of the taxonomy we assigned, U is an abbreviation of unique sequence, and the second *n* represents the number of the unique sequence we assigned. A representative V6 sequence of each unique sequence from this study is behind T*n*U*n*. Accessions numbers of representative references retrieved from Genbank are shown in parentheses, and followed by the taxonomic code (T*n*) according to the NCBI website description. The color-scale values in the squares indicate the occurrence frequency of the unique sequence in the corresponding libraries. *Escherichia coli* O111:H- is used as the outgroup.

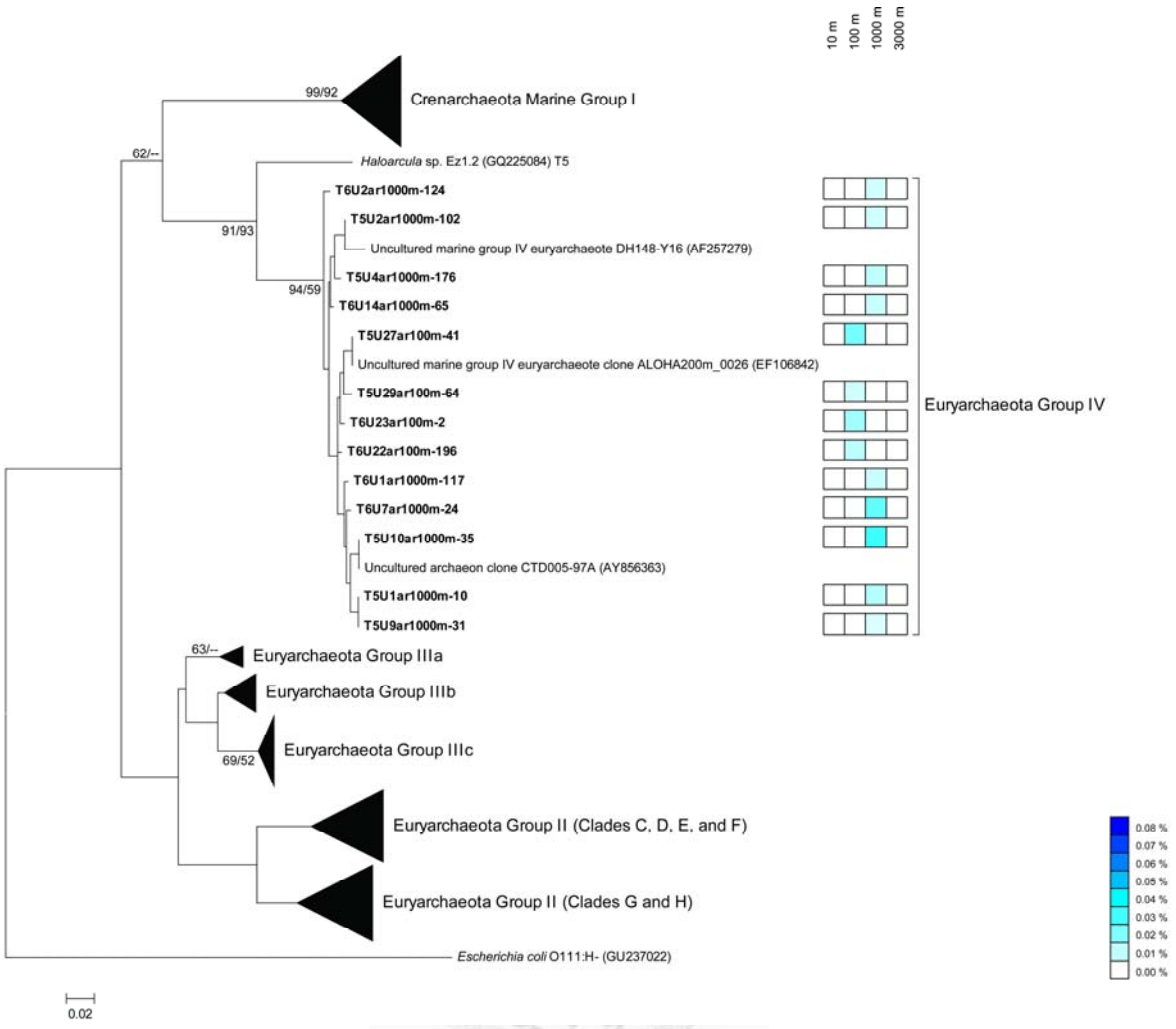


Figure 22. Phylogenetic tree of the marine euryarchaeota group IV V6 sequences. The tree is constructed by JC distances with NJ and MP methods using MEGA4 program (Tamura *et al.*, 2007-7). Bootstrap values above 50% (NJ/MP method; 1000 replications) are shown at each node. Scale bar represents the nucleotide substitution percentage of NJ method. V6 sequences from this study are indicated in boldface and designated as the beginning of T*n*U*n*, in which T is an abbreviation of taxonomy, the first *n* represents the number of the taxonomy we assigned, U is an abbreviation of unique sequence, and the second *n* represents the number of the unique sequence we assigned. A representative V6 sequence of each unique sequence from this study is behind T*n*U*n*. Accessions numbers of representative references retrieved from Genbank are shown in parentheses, and followed by the taxonomic code (T*n*) according to the NCBI website description. The color-scale values in the squares indicate the occurrence frequency of the unique sequence in the corresponding libraries. *Escherichia coli* O111:H- is used as the outgroup.

# Chapter 4 General discussion and conclusions

## 4.1 Diversity of the prokaryotes in the South China Sea

The diversity of the bacterial community in deep SCS water is lower than that in surface SCS water, implying that bacterial communities from the dark ocean are less diverse than surface communities (Table 3 and Figure 8a). Similar results have revealed from the North Pacific Ocean (ALOHA station, the site of HOT; Brown *et al*., 2009) and from the Arctic Ocean (Kirchman *et al*., 2010; Galand *et al*., 2010). As suggested by Brown *et al*. (2009), our consistent observation (i.e., bacterial diversity is higher in the sea surface) probably associates with an enhanced dissolved organic matter pool driven by the photosynthesis that could support many prokaryotic heterotrophs, and the availability of the energy from sunlight provides an opportunity to grow for mixotrophs in the euphotic zone, comparing to the deep dark ocean.

Moreover, archaeal communities in the northern SCS have a relatively low diversity comparing to that in the bacterial communities, in agreement with earlier reports that the bacterial community are generally more diverse archaeal community (Huber *et al*., 2007; Aller and Kemp, 2008; Table 3 and 4, and Figure 8).

We obtained the lower values of diversity indices, including the observed OTUs and richness estimates (ACE and Chao1), from the SEATS station (Table 3 and Figure 8a), in comparisons to earlier surveys by similar methods from other oceans, including the deep North Atlantic (Sogin *et al*., 2006), the North Pacific (Brown *et al*., 2009), and the Arctic (Kirchman *et al*., 2010; Galand *et al*., 2010). We suggest that the lower diversity of the northern SCS is caused by the oligotrophic character of the SCS (Wu *et al*., 2003; Lee Chen *et al*., 2004; Wen *et al*., 2006; Lee Chen *et al*., 2007; Wong *et al*.,

2007) and smaller sampling sizes comparing to those reports (Sogin *et al*., 2006; Brown *et al*., 2009; Kirchman *et al*., 2010; Galand *et al*., 2010). Noticeably, the values of putative OTUs, richness estimates, ACE, Chao1, and Shannon index obtained from the SEATS station by the pyrosequencing method are higher than that reported in many previous studies based on conventional analysis of 16S rRNA genes including clone libraries and molecular fingerprinting techniques (Gallagher *et al*., 2004; Alonso-Sáez *et al*., 2007; Pommier *et al*., 2007; Malmstrom *et al*., 2007; Santelli *et al*., 2008; Fuhrman *et al*., 2008; Cono *et al*., 2009; Galand *et al*., 2009b; Ma *et al*., 2009; Table 3 and 4), suggesting that the diversity of marine microbial communities are commonly underestimated. The main reason is insufficient sampling sizes in those studies that significantly affect the accuracy and sensitivity of surveying microbial communities in the ocean. Furthermore, the rarefaction curves of our study are still unsaturated (Figure 8), indicating that both the bacterial and archaeal diversity in SCS should be much greater than current estimates; more sequences survey is required.

## 4.2 Depth-related distribution

We provided the vertical profiles of marine microbial communities at Order level (Figure 9a), showing that the microbial community structures change in different depths. Although V6 sequences are short, they still offer a good taxonomical assignment for most of the bacterial assemblages as shown in Figure 9a. The short sequences are known may be assigned to a wrong taxon; however, we adopted a strategy with two-step taxonomic assignments to reduce the errors in taxonomic identification. The major compositions, such as *α*- and *γ-proteobacteria*, are generally similar to the present pyrosequencing data and earlier clone library descriptions of bacterioplankton

communities (Venter *et al*., 2004; DeLong *et al*., 2006; Ma *et al*., 2009; Brown *et al*., 2009; Andersson *et al*., 2010; Feingersch *et al*., 2010; Kirchma n *et al*., 2010). However, the compositions at Order level changes with depths; the relative abundance of *β-proteobacteria*, for instance, are higher in the upper layers, while *δ-proteobacteria* are represented higher in the deeper two layers. For another example, *Sphingobacterials* and *Cyanobacteria* mainly appear in the 10m-dpeth bacterial assemblages. Besides, *Methanobacterials* decreases with increasing depth, in contrast, *Thermoplasmatales* increases with depth.

The profiles of bacterial and archaeal V6 unique sequences also show that many sequences only appear at certain depths, indicating those sequences were originated from some microbial groups that are specifically inhabiting to the different depths. For example, unique sequences belonged to *Rickettiales* peak differently between the 10m and 3000m depths (Figure 10); some *Thermoplasmatales* sequences are merely detected in the bathypelagic layer, and several *Thermoplasmatales* sequences are found only the 100m depth (Figure 11). The profiles of bacterial and archaeal V6 unique sequences offer a finer resolution of the depth-specific distribution into Sequence level than that at Order level. In addition, through the phylogenetic analysis of bacterial and archaeal V6 sequences, we observe the depth-related subdivision in many microbial groups, including sequences belonged to *Sphingomonadales*, the unclassified alphaproteobacteria, *Alteromonadales*, *Oceanospirillales*, *Pseudomonadales*, and the marine euryarchaeota groups II and III (Figure 12, 14-16, 20, and 21).

According to our results described in previous paragraphs (Figure 9-12, 14-16, 20, and 21), we suggest that the varied physiological capacities and ecological roles among different pelagic bacteria groups lead to variances in spatial distribution and abundance of different bacteria groups.

**Information of microbial composition indicating the surface SCS water affected by offshore fluxes.** Despite several microbial groups occurring mainly in the upper layers, such as cyanobacteria, the SAR116 group, and the SAR86 group, probably involve in partially phototrophic lifestyles, an inference confirmed by genomic data (Béja *et al*., 2000; Cho and Giovannoni, 2004; Oh *et al*., 2010; Figure 12, 17, and 18); other sequences of *Rhizobiales*, *Rhodobacterales*, *Shingomonadales*, *Burkholderiales*, and the *Oceanospirillales* sequences related with the genera *Alcanivorax* and *Marinospirillum* (clades Y and Z) occur at highly frequencies in the 10m and 100m depths, and are closely related with some functionally important bacteria, such as the hydrocarbon-degrading bacteria and organic pollutant-metabolizing bacteria. If sequences of *Rhizobiales*, *Rhodobacterales*, *Shingomonadales*, *Burkholderiales*, and the *Oceanospirillales* sequences were originated from those functionally important bacteria, and if their metabalisms such as hydrocarbon-degrading and organic pollutant-metabolizing were active at the SEATS station, the high occurance frequencies of sequences of *Rhizobiales*, *Rhodobacterales*, *Shingomonadales*, *Burkholderiales*, and the *Oceanospirillales* sequences at the depths of 10m and 100m might suggest that there probably exists hydrocarbon sources in the upper layers at the SEATS station (Figure 12, 13, and 15). The existences of hydrocarbon sources in the northern SCS could be caused by offshore influences. The offshore influences are likely sourced from rivers and climate activities (tropical typhoons). The Zhujiang (the Pearl River), the second largest river by volume in China (the water discharge is ~316km$^3$ per year; Zhang, 1995; Wong *et al*., 2007), flows through densely populated areas (Dongguan, Foshan, Guangzhou, Hong Kong, Jiangmen, Macau, Shenzhen, Zhongshan, and Zhuhai) and drains into the northern SCS in which the SEATS station located. Moreover, most typhoons tend to form between May and November in the northwestern Pacific Ocean and frequently

have a great impact on the northern SCS that likely causes the coastal and riverine discharges increasing after typhoon events; due to our sampling processing in October, the time is during the late typhoon seasons, as the case reported in the Mediterranean Sea (Alonso-Sáez *et al*., 2007).

**Microbes survive in the cold, deep and dark sea.** On the other hand, sequences belonged to several clades, including the *Alteromonas*-related clades (clades M, O, and P), the *Moritella*-related clade (clade N), the clade Q, the *Oceanospirillaceae*-related clades (clades W and X), and the *Psychrobacter*-related clade (clade J), are affiliated to psychrophilic bacteria (Figure 14-16), suggesting that these groups have capacities for adapting to the cold niches and the high pressure environments and likely play a role in the deep oceanic ecosystem.

**Microbes mainly found in certain depths.** The *Marinospirillum*-related sequences (clade Z), some marine crenarchaeota group I sequences, the clades C and G sequences of marine euryarchaeota group II, and the marine euryarchaeota group IV sequences are represented at high frequencies neither in the sea surface nor in nearly sea bottom, but in the middle two layers, suggesting that microbial communities in the deeper surface layers of the northern SCS may be functionally specialized (Figure 15, 19, 20 and 22).

**The eurybathtal groups.** Furthermore, we observed many eurybathtal sequences, including some sequences of the unclassified alphaproteobacteria II and SAR11 groups, an *Alteromonas*-related clade (clade P), the clade I of *Pseudomonadales*, the marine crenarchaeota group I, and the marine euryarchaeota group IIIb, showing that these groups have unusual abilities to survive and adapt to a wide range of depths (Figure 12, 14, 16, 19, and 21).

## 4.3 Ambiguous taxonomically classified groups

**Novel microbes in the deep sea.** Several sequences mainly appear in the deep sea are not so closely related to the known bacteria, indicating that there could be novel bacterial species in the deep northern SCS. As shown in Figures 12 and 18, some sequences in the unclassified alphaproteobacteria II clade and the unclassified sequences with "T31" as the prefix are phylogenetically distinct from any known bacteria. Sequences in the clade M had the lowest sequence identity with *Alteromonas macleodii* AltDE among three clades (M, O, and P), and the relative abundances of the clade M sequences (they have the lowest sequence identity with *Alteromonas macleodii* AltDE isolated from a depth of 1000m) are all highest near the sea bottom, suggesting that they may be a "deeper ecotype" of *Alteromonas macleodii* in the northern SCS (Figure 14). These results indicate that a lot of novel bacteria not previously recognized exist in deep layers of the northern SCS and raise our interesting about their ecological roles in the deep sea.

**The SAR11 group.** According to our data, the subgroups of the SAR11 group are geographically-specific. Some SAR11 sequences are represented at high ratios in the upper layers (i.e., sequences T40U30/AF-33, T40U32/AF-4, T40U41/AF-97, and T40U13/AF-175 of the Clade B), some are abundant in and below the 1000m (i.e., the clade A sequence T35U346/AF2-783, and sequences T40U7/AF-1464, T40U4/AF-1269, T40U190/CF-290, and T40U205/CF-742 in the clade B), and some, however, are eurybathyal (i.e., the clade A sequences except the sequence T35U346/AF2-783, and sequences T40U30/AF-33, T40U41/AF-97, and T40U13/AF-175 of the Clade B; Figure 12).

*Pleagibacter ubique* strain HTCC1062 is one representative strain of the

ubiquitous SAR11 group, assigned to Order *Rickettsiales* (with a taxonomic code "T40"). However, not all sequences in the SAR11 cluster are grouped within the "T40" as prefix (clade A; Figure 12). Those SAR11 sequences clustered in the "T35" are failed to assign to any named Order but are closely related with *Pleagibacter ubique*. Based on these observations, we suggest that the SAR11 is a board group likely constituted more than one taxonomic even genus (Alonso-Sáez *et al*., 2007), and, obviously, their taxonomic assignments are still unsettled.

## 4.4 Unique archaeal assemblages

Our data reveals unique assemblages of Archaea in the northern SCS, with *Euryarchaeota* dominating throughout the entire 3000m water in the northern SCS (Figure 9b), which is different from most of previous studies (Karner *et al*., 2001; Zhang *et al*., 2009). *Crenarchaeota* has long been believed to be dominant in deep waters, whereas group II Euryarchaeota are often more abundant in surface waters (Karner *et al*., 2001; Herndl *et al*., 2005; DeLong *et al*., 2006). On the other hand, earlier reports from 3000m-depth layer at the Antarctic Polar Front (López-García *et al*., 2001a, 2001b) and from the surface layer in the Arctic Ocean (Galand *et al*., 2009a) have showed the dominance of *Euryarchaeota*. Few reports mentioned before (López-García *et al*., 2001a, 2001b; Galand *et al*., 2009a) and our date provides another view to euryarchaeota distribution in the ocean. The probable reasons for archaeal community variations could be associated to geographic specificity as well as the use of different probes, primers or methods (Casamayor *et al*., 2002; Galand *et al*., 2009a). It is deserved future investigations adopted a standardized approach combining quantitative and qualitative methods to have a comprehensive picture of the archaeal community

structures in the northern SCS.

The dominant archaeal unique sequences can be phylogenetically classified in four marine groups: the marine crenarchaeota group I, and the marine euryarchaeota groups II, III, and IV (Figure 19-22) as according the classification reported in previous studies (Martin-Cuadrado *et al*., 2008; Galand *et al*., 2008, 2009a). Galand and his coworkers (2009a) have found a mainly sea surface group of the marine euryarchaeota group II in the Arctic, and the group are belonged to the cluster that contained sequences obtained from the coastal seawater or sediments; thus, they suggested those sea surface group of the marine euryarchaeota group II could be involved in the anaerobic degradation of organic matter coming from river runoff. In our results, thirty-one of 43 marine euryarchaeota groups II sequences show the highest frequency of the appearance in the 10m depth, in agreement with the observations of Galand *et al.* (2009a). In addition, the high abundance of the thirty-one of 43 marine euryarchaeota groups II sequences in the surface layer is corresponded to the presence of the pollutant-associated bacterial groups (described in the Section 4.2 of this study), indicating that the surface SCS water is affected by offshore fluxes. On the other hand, The marine euryarchaeota group II sequences mainly found in the middle two layers (clades C and G) are distinct from those detected in surface waters (Figure 20). In addition, the sequences in the marine euryarchaeota groups IIIa and IIIc are represented highly in the meso- and bathy-pelagic layers and diverse from the eurybathyal sequences (marine euryarchaeota group IIIc, Figure 21). What factors drive the variances of distribution among these groups? Is there any association with the specific physiology of each group? These questions still remain unknown. More studies for different marine euryarchaeota groups II and III in different environments may provide some clues.

Notably, the marine euryarchaeota group III is usually scarce in marine ecosystems,

but they occupy largely in the deep northern SCS (over 47% of the 3000m archaeal library); besides, their occurrences can not be ignored in other three depths (at least 7%, 18%, and 29% of the 10m, 100m, and 1000m archaeal library respectively; Figure 21). The marine euryarchaeota group III was first identified from the depths of 500 and 300m in the Northeast Pacific (Fuhrman and Davis, 1997) and was later detected at over 400m depth in other oceans, including the Mediterranean Sea (Massana *et al*., 2000; Martín-Cuadrado *et al*., 2007, 2008), the south Atlantic (López-García *et al*., 2001a, 2001b; Martín-Cuadrado *et al*., 2007), and the Arctic (Galand *et al*., 2009a), suggesting that they were specific to the deep ocean. Two studies from the deep South Atlantic (Martín-Cuadrado *et al*., 2008) and the deep central Arctic (Galand *et al*., 2009a) have described the high abundances of the marine euryarchaeota group III, indicating the group successfully survives in deep high latitude water masses. One recent community genomics study in the North Pacific Ocean revealed the presence of the marine euryarchaeota group III throughout the entire 4000m-depth water span, although they were little represented (DeLong *et al*., 2006). Our observation on the abundant occupation of the marine euryarchaeota group III in the northern SCS suggests that the group successfully thrives not only in deep high latitude water masses but also in tropical region. Little known is about the metabolism and ecological role of the marine euryarchaeota group III, because members of the marine euryarchaeota group III are not-yet-cultivated to date. Nevertherless, Martín-Cuadrado and his coworkers (2008) suggested some members of the marine euryarchaeota group III may be oxidize-ammonia, because they detected the DNA fragments of genes originated from the marine euryarchaeota group III that resembled that found in ammonia oxidizing bacteria. If this is confirmed, the highly abundant occupations of the marine euryarchaeota group III imply that some members of the group may play an important

role in the nitrogen cycle. Because some members of the marine crenarchaeota group I have the ability to oxidize ammonia, our observation about the little represented *Crenarchaeota* and the high presence of the marine euryarchaeota group III raises a speculation that if some members of the marine euryarchaeota group III could use ammonia, the group may have success in the competition with ammonia-oxidizing crenarchaeota in the northern SCS.

The marine euryarchaeota group IV were mainly found in the middle two depths at SEATS station, but at low abundance (about 5.5% and 10.7% of the 100m and 1000m archaeal library, respectively; Figure 22). With the exception of one clone obtained at 55m depth from the upper Arctic (Bano *et al*., 2004), the marine euryarchaeota group IV sequences had been detected solely from samples collected at the meso- and bathy-pelagic depths from the Antarctic Polar Front, the North Atlantic, the Mediterranean Sea, and the central Arctic (López-García *et al*., 2001a, 2001b; Galand *et al*., 2009a). Our results confirm that the marine euryarchaeota group IV are specific to the deep oceans and represented as a minor group. The metabolism of this group at present reamins unknown.

## 4.5 Conclusions/significance

We present the diversity and distribution of marine prokaryotic communities along the vertical profiles by using the massively-parallel pyrosequencing technique at the site of SEATS station in the northern SCS (Figure 1). Several important contributions in this study are enumerated as following:

(1) For the first time, we successfully applied massively-parallel pyrosequencing to describe the diversity and taxonomic composition of both bacterial and archaeal

communities in the western tropical Pacific. Massively-parallel pyrosequencing give a high coverage of these assemblages, and this method offers a good taxonomical assignment for most of the bacterial groups as shown by the general agreement of bacterial assemblages between the present V6 data and earlier clone library descriptions.

(2) Noticeably, we obtained higher values of putative OTUs, richness estimates, ACE, Chao1, and Shannon index by the pyrosequencing method than that reported in many previous studies based on conventional analysis of 16S rRNA genes including clone libraries and molecular fingerprinting techniques (Gallagher *et al*., 2004; Alonso-Sáez *et al*., 2007; Pommier *et al*., 2007; Malmstrom *et al*., 2007; Santelli *et al*., 2008; Fuhrman *et al*., 2008; Ma *et al*., 2009; Cono *et al*., 2009; Galand *et al*., 2009b; Table 3 and 4), suggesting that marine microbial communities were more diverse than our earlier estimates. In addition, our data shows the diversity of the bacterial community in deep SCS water is lower than that in surface SCS water, suggesting that bacterial communities from the dark ocean are less diverse than surface communities (Table 3 and 4, and Figure 8). In agreement with earlier reports (Huber *et al*., 2007; Aller and Kemp, 2008), the bacterial community is more diverse than archaeal community in the SCS (Table 3 and 4, and Figure 8).

(3) Our results also reveal unique assemblages of Archaea in the northern SCS, with *Euryarchaeota* dominating throughout the entire 3000m-deep water column. In the deep northern SCS, the marine euryarchaeota group III has high appearances, while the marine euryarchaeota group II are dominate in the upper layers. The relative abundance of marine crenarchaeota group I is overwhelmed by that of the marine euryarchaeota group III from the four depths libraries. Those results support the new perspectives reported by Galand *et al*. (2009a) that there is a possible competition between the marine euryarchaeota group III and marine crenarchaeota group I for putative energy

sources such as ammonia oxidation.

(4) The repeated discovery of specifically inhabitating prokaryotic groups along vertical profile in the water column indicates the depth is one of the important variables affecting the microbial communities and is a good supplement to the current understanding of biogeographical traits for prokaryotic communities in the norther SCS. Our study provides an important foundational data for understanding marine microbial diversity and communities in the SCS and tropical seawaters.

# Chapter 5 Future work

New low-cost, high-throughput sequencing can greatly advance the analysis of marine microbial community structure, as presenting in this study. Including the focus about the taxonomic composition of marine microbial communities, the understanding about a possible change in the microbial community structure with respect to time is also a fundamental issue for subsequent studies in the northern SCS. Therefore, we suggest that extended seasonal or annual investigations could provide the relevant information about the variations of prokaryotic community structure during different time scales here.

In addition, a survey of a large spatial scale is important to enlarge our understanding about the marine microbial community. Thus, we are interested in the variations of microbial community among different sites in the SCS and between from the SCS and the Kuroshio. We purpose to make a line transect along the latitude of 18°15' N from SEATS station to the Kuroshio passing through the Luzon Strait, which is the deepest passage between the western Pacific and the SCS, and to study the large horizontal and vertical scales of microbial communities. These projects could offer us not only the information of marine microbial communities but also an examination of the interaction of water masses between the Kuroshio and the SCS through the analysis of microbial communities along the line transect. Moreover, to have a more comprehensive picture of the influence from the Pearl River, we could make another line transect form SEATS station to the mouth of Pearl River to study the differences of marine microbial communities.

Furthermore, linking the microbial taxonomic structures to their ecological roles is the fundamental to ecological studies in the marine ecosystems. Metagenomics, the

study of genetic and genomic information recovered directly from environmental samples, has brought some hope to get insights about the metabolic potential and evolutionary history of uncultured marine microbes. Extending metagenomic analyses into subsequent microbial studies in the SCS would help us link their taxonomic structures to their metabolisms, lifestyles, and ecological roles.

# REFERENCE


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

Adav, S. S., Lee, D. J., Lai, J. Y., Potential cause of aerobic granular sludge breakdown at high organic loading rates, *Applied Microbiology and Biotechnology* **85**: 753-762 (2010a)

Adav, S. S., Lee, D. J., Lai, J. Y., Microbial community of acetate utilizing denitrifiers in aerobic granules, *Applied Microbiology and Biotechnology* **85**: 1601-1610 (2010b)

Alford, M. H., Redistribution of energy available for ocean mixing by long-range propagation of internal waves, *Nature* **423**: 159-162 (2003)

Allen, A. P., Gillooly, J. F., Brown, J. H., Global biodiversity, biochemical kinetics, and the energetic-equivalence rule, *Science* **297**: 1545-1548 (2002)

Aller, J. Y., Kemp, P. F., Are archaea inherently less diverse than bacteria in the same environments? *FEMS Microbiol Ecol.* **65**: 74-87 (2008)

Alonso-Sáez, L., Balagué, V., Sà, E. L., Sánchez, O., González, J. M., Pinhassi, J., Massana, R., Pernthaler, J., Pedrós-Alió, C., Gasol, J. M., Seasonality in bacterial diversity in north-west Mediterranean coastal waters: assessment through clone libraries, fingerprinting and FISH, *FEMS Microbiol Ecol.* **60**: 98-112 (2007)

Alonso-Sáez, L., Balagué, V., Sà, E. L., Sánchez, O., Gonzalez, J. M., Pinhassi, J., Massana, R., Pernthaler, J., Pedrós-Alió, C., Gasol, J. M., Seasonality in

bacterial diversity in north-west Mediterranean coastal waters: assessment through clone libraries, fingerprinting, and FISH, *FEMS Microbiol Ecol.* **60**: 98-112 (2007)

Andersson, A. F., Riemann, L., Bertilsson, S., Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities, *ISME J* **4**: 171-181 (2010)

Anesti, V., Vohra, J., Goonetilleka, S., McDonald, I. R., Sträubler, B., Stackebrandt, E., Molecular detection and isolation of facultatively methylotrophic bacteria, including *Methylobacterium podarium* sp. nov., from the human foot microflora, *Environ. Microbiol.* **6**: 820-830 (2004)

Arrigo, K. R., Marine microorganisms and global nutrient cycles, *Nature* **437**, 349-355 (2005)

Ayala-del-Río, H. L., Chain, P. S., Grzymski, J. J., Ponder, M. A., Ivanova, N., Bergholz, P. W., Di Bartolo, G., et al., The Genome Sequence of *Psychrobacter arcticus* 273-4, a Psychroactive Siberian Permafrost Bacterium, Reveals Mechanisms for Adaptation to Low-Temperature Growth, *Appl. Environ. Microbiol.* **76**: 2304-2312 (2010)

Azam, A., Hodson, R. E., Size distribution and activity of marine microheterotrophs, *Limnol. Oceanogr.* **22**: 492-501 (1977)

Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyer-Reil, L. A., Thingstad, F., The ecological role of water-column microbes in the sea, *Mar. Ecol. Prog. Ser.* **10**: 257-263 (1983)

Baldwin, A. J., Moss, J. A., Pakulski, J. D., Catala, P., Joux, F., Jeffrey, W. H., Microbial diversity in a Pacific Ocean transect from the Arctic to Antarctic circles, *Aquat. Microb. Ecol.* **41**: 91-102 (2005)

Bano, N., Ruffin, S., Ransom, B., Hollibaugh, J. T., Phylogenetic Composition of Arctic Ocean Archaeal Assemblages and Comparison with Antarctic Assemblages, *Appl. Environ. Microbiol.* **70**: 781-789 (2004)

Bartlett, D. H., Extremophilic vibrionaceae, In *The biology of vibrios*, ed. by Thompson, F. L., Austin, B., Swings, J., American Society for Microbiology Press, Washington D.C., pp. 156-171 (2006)

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

Bentley, R., Chasteen, T. G., Microbial methylation of metalloids: arsenic, antimony and bismuth, *Microbiol. Mol. Biol. Rev.* **66**: 250-271 (2002)

Binladen, J., Gilbert, M. T., Bollback, J. P., Panitz, F., Bendixen, C., Nielsen, R., Willerslev, E., The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing, *PLoS ONE* **2**: e197 (2007)

Borde, X., Guieysse, B., Delgado, O., Muñoz, R., Hatti-Kaul, R., Nugier-Chauvin, C., Patin, H., Mattiasson, B., Synergistic relationships in algal-bacterial microcosms for the treatment of aromatic pollutants, *Bioresour Technol* **86**: 293-300 (2003)

Bosch, G., Skovran, E., Xia, Q., Wang, T., Taub, F., Miller, J. A., Lidstrom, M. E., Hackett, M., Comprehensive proteomics of *Methylobacterium extorquens* AM1 metabolism under single carbon and nonmethylotrophic conditions, *Proteomics* **8**: 3494-3505 (2009)

Bozal, N., Montes, M. J., Tudela, E., Guinea, J., Characterization of several

*Psychrobacter* strains isolated from Antarctic environments and description of *Psychrobacter luti* sp. nov. & *Psychrobacter fozii* sp. nov., *Int. J. Syst. Evol. Microbiol.* **53**: 1093-1100 (2003)

Britschgi, T. B., Giovannoni, S. J., Phylogenetic analysis of a natural marine bacterioplankton population by rRNA gene cloning and sequencing, *Appl. Environ. Microbiol.* **57**: 1313-1318 (1991)

Brodie, E. L., DeSantis, T. Z., Joyner, D. C., Baek, S. M., Larsen, J. T., Andersen, G. L., Hazen, T. C., Richardson, P. M., Herman, D. J., Tokunaga, T. K., Wan J. M., Firestone, M. K., Application of a High-Density Oligonucleotide Microarray Approach To Study Bacterial Population Dynamics during Uranium Reduction and Reoxidation, *Appl. Environ. Microbiol.* **72**: 6288-6298 (2006)

Brown, M. V., Philip, G. K., Bunge, J. A., Smith, M. C., Bissett, A., Lauro, F. M., Fuhrman J. A., Donachie, S. P., Microbial community structure in the North Pacific ocean, *ISME J* **3**: 1374-1386 (2009)

Cai, W. J., Dai, M., Wang, Y., Zhai, W., Huang, T., Chen, S., Zhang, F., Chen, Z., Wang, Z., The biogeochemistry of inorganic carbon and nutrients in the Pearl River estuary and the adjacent northern South China Sea, *Continental Shelf Research* **24**: 1301-1319 (2004)

Casamayor, E. O., Massana, R., Benlloch, S., Øvreås, L., Díez, B., Goddard, V. J., Gasol, J. M., Joint, I., Rodríguez-Valera, F., Pedrós-Alió, C., Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern, *Environ. Microbiol.* **4**: 338-348 (2002)

Castle, D., Kirchman, D. L., Composition of estuarine bacterial communities assessed by denaturing gradient gel electrophoresis and fluorescence *in situ*

hybridization, *Limnol. Oceanogr.: Methods* **2**: 303-314 (2004)

Chang, Y. T., Hsu, W. L., Tai, J. H., Tang T. Y., Chang, M. H., Chao, S. Y., Cold deep water in the South China Sea, *Journal of Oceanography* **66**: 183-190 (2010)

Chen, C. C., Shiah, F. K., Chung, S. W., Liu, K. K., Winter phytoplankton blooms in the shallow mixed layer of the South China Sea enhanced by upwelling, *Journal of Marine Systems* **59**: 97-110 (2006)

Chen, C. T. A., Wang, S. L., Chou, W. C., Sheu D. D., Carbonate chemistry and projected future changes in pH and $CaCO_3$ saturation state of the South China Sea, *Marine Chemistry* **101**: 277-305 (2006)

Chen, C. T. A., Wang, S. L., Wang, B. J., Nutrient budgets for the South China Sea basin, *Marine Chemistry* **75**: 281-300 (2001)

Cho, J. C., Giovannoni, S. J., Cultivation and Growth Characteristics of a Diverse Group of Oligotrophic Marine *Gammaproteobacteria*, *Appl. Environ. Microbiol.* **70**: 432-440 (2004)

Christoserdova, L., Chen, S. W., Lapidus, A., Lidstrom, M. L., Methylotrophy in *Methylobacterium extorquens* AM1 from a genomic point of view, *J. Bacteriol.* **185**: 2980-2987 (2003)

Clement, B. G., Kehl, L. E., DeBord, K. L., Kitts, C. L., Terminal restriction fragment patterns (TRFPs), a rapid, PCR-based method for the comparison of complex bacterial communities, *Journal of Microbiological Methods* **31**: 135-142 (1998)

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., Kulam-Syed-Mohideen, A. S., McGarrell, D. M., Marsh, T., Garrity, G. M., Tiedje, J. M., The ribosomal database project: improved alignments and new tools for rRNA analysis, *Nucleic Acids Res.* **37**: D141-145 (2008)

Colwell, R. R., Morita, R. Y., Reisolation and emendation of description of *Vibrio*

*marinus* (Russell) Ford, *J. Bacteriol.* **88**: 831-837 (1964)

Cono, V. L., Tamburini, C., Genovese, L., Spada, G. L., Denaro, R., Yakimov M. M., Cultivation-independent assessment of the bathypelagic archaeal diversity of Tyrrhenian Sea: Comparative study of rDNA and rRNA-derived libraries and influence of sample decompression, *Deep-Sea Res. II* **56**: 768-773 (2009)

DeLong, E. F., Franks, D. G., Yayanos, A. A., Evolutionary relationships of cultivated psychrophilic and barophilic deep-sea bacteria, *Appl. Environ. Microbiol.* **63**: 2105-2108 (1997)

DeLong, E. F., Preston, C. M., Mincer, T., Rich, V., Hallam, S. J., Frigaard, N. U., Martinez, A., Sullivan, M. B., Edwards, R., Rodriguez Brito, B., Chisholm, S. W., Karl, D. M., Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **27**: 496-503 (2006)

Duda, T. F., Lynch, J. F., *Senior Member, IEEE,* Irish, J. D., Beardsley, R. C., Ramp, S. R., Chiu, C. S., Tang, T. Y., Yang Y. J., Internal tide and nonlinear wave behavior in the continental slope in the northern South China Sea, *IEEE J. Ocean. Eng.* **29**: 1105-1131 (2004)

Edgar, R. C., MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.* **32**: 1792-1797 (2004)

Edwards R. A., Rodriguez-Brito, B., Wegley, L., Haynes, M., Breitbart, M., Peterson, D. M., Saar, M. O., Alexander, S., Alexander, E. C., Rohwer, F., Using pyrosequencing to shed light on deep mine microbial ecology, *BMC Genomics* **7**: 57-69 (2006)

Emery, W. J., Water types and water masses, In *Encyclopedia of Ocean Sciences*, ed. by J. H. Steele, S. A. Thorpe and K. K. Turekian, Academic Press, Inc., pp. 3179-3187 (2001)

Falkowski, P. G., Barber, R. T., Smetacek, V., Biogeochemical Controls and Feedbacks on Ocean Primary Production, *Science* **281**: 200-206 (1998)

Falkowski, P. G., Fenchel T., Delong E. F., The Microbial Engines That Drive Earth's Biogeochemical Cycles, *Science* **320**: 1034-1039 (2008)

Feingersch, R., Suzuki, M. T., Shmoish, M., Sharon, I., Sabehi, G., Partensky F., Béjà, O., Microbial community genomics in eastern Mediterranean Sea surface waters, *ISME J* **4**: 78-87 (2010)

Fenchel, T., Marine Plankton Food Chains, *Annual Review of Ecology and Systematics* **19**: 19-38 (1988)

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

Fischera, J., Kappelmeyera, U., Kastnera, M., Schauerb, F., Heipieper, H. J., The degradation of bisphenol A by the newly isolated bacterium *Cupriavidus basilensis* JF1 can be enhanced by biostimulation with phenol, *International Biodeterioration & Biodegradation* **64**: 324-330 (2010)

Frigaard, N. U., Martinez, A., Mincer, T. J., DeLong, E. F., Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea, *Nature* **439**: 847-850 (2006)

Fuhrman, J. A., Ammerman, J. W., Azam, A., Bacterioplankton in the coastal euphotic zone: distribution, activity and possible relationships with phytoplankton, *Marine Biology* **60**: 201-207 (1980)

Fuhrman, J. A., Azam, F., Bacterioplankton Secondary Production Estimates for Coastal Waters of British Columbia, Antarctica, and California, *Appl. Environ. Microbiol.* **39**: 1085-1095 (1980)

Fuhrman, J. A., Azam, F., Thymidine incorporation as a measure of heterotrophic bacterioplankton production in marine surface waters: evaluation and field results, *Marine Biology* **66**: 109-120 (1982)

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

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

Fuhrman, J. A., Steele, J. A., Hewson, I., Schwalbach, M. S., Brown, M. V., Green, J. L., Brown, J. H., A latitudinal diversity gradient in planktonic marine bacteria, *Proc Natl Acad Sci USA* **105**: 7774-7778 (2008)

Galand, P. E., Potvin, M., Casamayor, E. O., Lovejoy, C., Hydrography shapes bacterial biogeography of the deep Arctic Ocean, *ISME J* **4**: 564-576 (2010)

Galand, P. E., Casamayor, E. O., Kirchman, D. L., Potvin, M., Lovejoy, C., Unique archaeal assemblages in the Arctic Ocean unveiled by massively parallel tag sequencing, *ISME J* **3**: 860-869 (2009a)

Galand, P. E., Lovejoy, C., Hamilton, A. K., Ingram, R. G., Pedneault, E., Carmack, E. C., Archaeal diversity and a gene for ammonia oxidation are coupled to oceanic circulation, *Environ. Microbiol.* **11**: 971-980 (2009b)

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

Gallagher, J. M., Carton, M. W., Eardly, D. F., Patching, J. W., Spatio-temporal variability and diversity of water column prokaryotic communities in the eastern North Atlantic, *FEMS Microbiol Ecol.* **47**: 249-262 (2004)

García-Martínez, J., Rodríguez-Valera, F., Microdiversity of uncultured marine prokaryotes: the SAR11 cluster and the marine Archaea of Group I, *Mol Ecol* **9**: 935-948 (2000)

Giovannoni, S. J., Rappé, M. S., Evolution, diversity, and molecular ecology of marine prokaryotes, In *Microbial ecology of the oceans*, ed. by Kirchmann, D. L., Wiley Series in Ecological and Applied Microbiology, pp. 47-84 (2000)

Giovannoni, S. J., Rappe, M. S., Vergin, K. L., Adair, N. L., 16S rRNA genes reveal stratified open ocean bacterioplankton populations related to the green non-sulfur bacteria, *PNAS* **93**: 7979-7984 (1996)

Giovannoni, S. J., Stingl, U., Molecular diversity and ecology of microbial plankton, *Nature* **437**: 343-348 (2005)

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

Good, I. J., The population frequencies of species and the estimation of population parameters. *Biometrika* **40**: 237-264 (1953)

Goris, J., De Vos, P., Coenye, T., Hoste, B., Janssens, D., Brim, H., Diels, L., Mergeay, M., Kersters, K., Vandamme, P., Classification of metal-resistant bacteria from industrial biotopes as *Ralstonia campinensis* sp. nov., *Ralstonia metallidurans* sp. nov. and *Ralstonia basilensis* Steinle *et al*., 1998 emend, *Int. J. Syst. Evol. Microbiol.* **51**: 1773-1782 (2001)

Gram, L., Melchiorsen, J., Bruhn, J. B., Antibacterial activity of marine culturable bacteria collected from a global sampling of ocean surface waters and surface swabs of marine organisms, *Mar. Biotechnol*, in press. (DOI: 10.1007/s10126-009-9233-y)

Herndl, G. J., Reinthaler, T., Teira, E., Van Aken, H., Veth, C., Pernthaler, A., Pernthaler,

J., Contribution of *Archaea* to Total Prokaryotic Production in the Deep Atlantic Ocean, *Appl. Environ. Microbiol.* **71**: 2303-2309 (2005)

Higginson, M. J., Maxwell, J. R., Altabet, M. A., Nitrogen isotope and chlorine paleoproductivity records from the northern South China Sea: remote vs. local forcing of millennial- and orbital-scale variability, *Marine Geology* **201**: 223-250 (2003)

Hobbie, J. E., Holm-Hansen, O., Packard, T. T., Pomeroy, L. R., Sheldon, R. W., Thomas, J. P., Wiebe, W. J., A study of the distribution and activity of microorganisms in ocean water, *Limnol. Oceanogr.* **17**: 544-555 (1972)

Huber, J. A., Mark Welch, D. B., Morrison, H. G., Huse, S. M., Neal, P. R., Butterfield, D. A., Sogin, M. L., Microbial population structures in the deep marine biosphere, *Science* **318**: 97-100 (2007)

Hughes-Martiny, J. B., Bohannan, B. J. M., Brown, J., Colwell, R., Fuhrman, J., Green, J., Horner-Devine, M., Kane, M., Krumins, J., Kuske, C., Morin, P., Naeem, S., Ovreas, L., Reysenbach, A., Smith, V., Staley, J., Microbial biogeography: putting microorganisms on the map, *Nat Rev Microbiol* **4**: 102-112 (2006)

Hung, J. J., Wang S. M., Chen, Y. L., Biogeochemical controls on distributions and fluxes of dissolved and particulate organic carbon in the Northern South China Sea, *Deep-Sea Res. II* **54**: 1486-1503 (2007)

Huse, S. M., Dethlefsen, L., Huber, J. A., Welch, D. M., Relman, D. A., Sogin, M. L., Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing, *PLoS Genetics* **4**, e1000255 (2008)

Huse, S. M., Huber, J. A., Morrison, H. G., Sogin, M. L., Welch, D. M., Accuracy and quality of massively parallel DNA pyrosequencing, *Genome Biol.* **8**, R143 (2007)

Ivars-Martínez, E., Martín-Cuadrado, A. B., D'Auria, G., Mira, A., Ferriera, S., Johnson, J., Friedman, R., Rodriguez-Valera, F., Comparative genomics of two ecotypes of the marine planktonic copiotroph *Alteromonas macleodii* suggests alternative lifestyles associated with different kinds of particulate organic matter, *ISME J* **2**: 1194-1212 (2008)

Jarvie, T., Du, L., Knight, J., Shotgun sequencing and assembly of microbial genomes: comparing 454 and Sanger methods, *Biochemica* No. 4: 11-14 (2005)

Karner, M. B., DeLong, E. F., Karl, D. M., Archaeal dominance in the mesopelagic zone of the Pacific Ocean, *Nature* **409**: 507-510 (2001)

Kato, Y., Asahara, M., Goto, K., Kasai, H., Yokota A., *Methylobacterium persicinum* sp. nov., *Methylobacterium komagatae* sp. nov., *Methylobacterium brachiatum* sp. nov., *Methylobacterium tardum* sp. nov. and *Methylobacterium gregans* sp. nov., isolated from freshwater, *Int. J. Syst. Evol. Microbiol.* **58**, 1134-1141 (2008)

Kirchman, D. L., Cottrell, M. T., Lovejoy, C., The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes, *Environ. Microbiol.* **12**: 1132-1143 (2010)

La Duc, M. T., Nicholson, W., Kern, R. Venkateswaran, K., Microbial characterization of the Mars Odyssey spacecraft and its encapsulation facility, *Environ. Microbiol.* **5**: 977-985 (2003)

Lee Chen, Y. L., Chen, H. Y., Chung, C. W., Seasonal variability of coccolithophore abundance and assemblage in the northern South China Sea, *Deep-Sea Res. II* **54**: 1617-1633 (2007)

Lee Chen, Y. L., Chen, H. Y., Karl, D. M., Takahashi, M., Nitrogen modulates phytoplankton growth in spring in the South China Sea, *Continental Shelf*

*Research* **24**: 527-541 (2004)

Lee, S. H., Fuhrman, J. A., Species composition shift of confined bacterioplankton studied at the level of community DNA, *Mar. Ecol. Prog. Ser.* **79**: 195-201 (1991)

Liu, H., Chang, J., Tseng, C. M., Wen, L. S., Liu, K. K., Seasonal variability of picoplankton in the Northern South China Sea at the SEATS station, *Deep-Sea Res. II* **54**: 1602-1616 (2007)

López-García, P., López-López, A., Moreira, D., Rodríguez-Valera, F., Diversity of free-living prokaryotes from a deep-sea site at the Antarctic Polar Front, *FEMS Microbiol. Ecol.* **36**: 193-202 (2001a)

López-García, P., Moreira, D., López-López, A., Rodríguez-Valera, F., A novel haloarchaeal-related lineage is widely distributed in deep oceanic regions, *Environ. Microbiol.* **3**: 72-78 (2001)

Lunder, T., Sørum, H., Holstad, G., Phenotypic and genotypic characterization of *Vibrio viscosus* sp. nov. and *Vibrio wodanis* sp. nov. isolated from Atrantic salmon (Salmo salar) with 'winter ulcer', *Int. J. Syst. Bacteriol.* **50**: 427-450 (2000)

Ma, Y., Zeng, Y., Jiao, N., Shi, Y., Hong, N., Vertical distribution and phylogenetic composition of bacteria in the Eastern Tropical North Pacific Ocean, *Microbiological Research* **164**: 624-633 (2009)

Madigan, M. T., Martinko, J. M., Dunlap, P. V., Clark, D. P., Brock Biology of Microorganisms 12/e (Pearson Prentice Hall, Inc., 2009)

Malmstrom, R. R., Straza, T. R. A., Cottrell, M. T., Kirchman, D. L., Diversity, abundance, and biomass production of bacterial groups in the western Arctic Ocean, *Aquat. Microb. Ecol.* **47**: 45-55 (2007)

Man, D. L., Wang, W. W., Sabehi, G., Aravind, L., Post, A. F., Massana, R., Spudich, E.

N., Spudich, J. L., Béjà, O., Diversification and spectral tuning in marine proteorhodopsins, *EMBO J* **22**: 1725-1731 (2003)

Manter, D. K., Delgado, J. A., Holm, D. G., Stong, R. A., Pyrosequencing Reveals a Highly Diverse and Cultivar-Specific Bacterial Endophyte Community in Potato Roots, *Microbial Ecology* (the online version; doi: 10.1007/s00248-010-9658-x)

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., *et al*., Genome sequencing in microfabricated high-density picolitre reactors, *Nature* **437**: 376-380 (2005)

Martín-Cuadrado, A. B., López-Garcia, P., Alba, J. C., Moreira, D., Monticelli, L., Strittmatter, A., Gottschalk, G., Rodríguez-Valera, F., Metagenomics of the deep Mediterranean, a warm bathypelagic habitat, *PLoS ONE* **2**: e914 (2007)

Martín-Cuadrado, A. B., Rodriguez-Valera, F., Moreira, D., Alba, J. C., Ivars-Martínez, E., Henn, M. R., Talla, E., López-García, P., Hindsight in the relative abundance, metabolic potential and genome dynamics of uncultivated marine archaea from comparative metagenomic analyses of bathypelagic plankton of different oceanic regions, *ISME J* **2**: 865-886 (2008)

Massana, R., DeLong, E. F., Pedrós-Alió C., A Few Cosmopolitan Phylotypes Dominate Planktonic Archaeal Assemblages in Widely Different Oceanic Provinces, *Appl. Environ. Microbiol.* **66**: 1777-1787 (2000)

Maszenan, A. M., Seviour, R. J., Patel, B. K. C., Rees, G. N., McDougall, B. M., *Amaricoccus* gen. nov., a gram-negative coccus occurring in regular packages and tetrads, isolated from activated sludge biomass, and descriptions of *Amaricoccus veronensis* sp. nov., *Amaricoccus tamworthensis* sp. nov., *Amaricoccus macauensis* sp. nov. and *Amaricoccus kaplicensis* sp. nov., *Int J.*

*Syst. Bacteriol.* **47**: 727-734 (1997)

McInerney, J. O., Mullarkey, M. M., Werneke, M. E., Powell, R., Phylogenetic analysis of group I marine archaeal rRNA sequences emphasizes the hidden diversity within the primary group Archaea, *Proc. R. Soc. Lond. B* **264**: 1663-1669 (1997)

Methé, B. A., Nelson, K. E., Deming, J. W., Momen, B., Melamud, E., Zhang, X. et al., The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses, *Proc Natl Acad Sci USA* **102**: 10913-10918 (2005)

Mincer, T. J., Church, M. J., Taylor, L. T., Preston, C. Karl D. M., DeLong, E. F., Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre, *Environ. Microbiol.* **9**: 1162-1175 (2007)

Mirdamadi, S., Rajabi, A., Khalilzadeh, P., Norozian, D., Isolation of bacteria able to metabolize high concentrations of formaldehyde, *World J Microbiol Biotechnol* **21**: 1299-1301 (2005)

Moeseneder, M. M., Arrieta, J. M., Muyzer, G., Winter, C., Herndl, G. J., Optimization of Terminal-Restriction Fragment Length Polymorphism Analysis for Complex Marine Bacterioplankton Communities and Comparison with Denaturing Gradient Gel Electrophoresis, *Appl. Environ. Microbiol.* **65**: 3518-3525 (1999)

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

Muyzer, G., de Waal, E. C., Uitterlinden, A. G., Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase

chain reaction-amplified genes coding for 16S rRNA, *Appl. Environ. Microbiol.* **59**: 695-700 (1993)

Nakamura, K., Ishida, H., Iizumi, T., Constitutive trichloroethylene degradation led by tac promoter chromosomally integrated upstream of phenol hydroxylase genes of *Ralstonia* sp. KN1 and its nucleotide sequence analysis, *J. Biosci. Bioeng.* **89**: 47-54 (2000)

Nogi, Y., Kato, C., Horikoshi, K. *Moritella japonica* sp. nov., a novel barophilic bacterium isolated from a Japan Trench sediment, *J. Gen. Appl. Microbiol.* **44**: 289-295 (1998)

Oh, H. M., Kwon, K. K., Kang, I., Kang, S. G., Lee, J. H., Kim, S. J., Cho, J. C., Complete Genome Sequence of "*Candidatus Puniceispirillum marinum*" IMCC1322, a Representative of the SAR116 Clade in the *Alphaproteobacteria*, *Journal of Bacteriology* **192**: 3240-3241 (2010)

Oostende, N. V., Vyverman, W., Harlay, J., De Bodt, C., Chou, L., Suykens, K., Borges, A. V., Piontek, J., Engel, A., Sabbe, K., Coccolithophore bloom dynamics shape bacterioplankton communities in the northern Bay of Biscay (The 12th internatonal symposium on microbial ecology (ISME 12), CAIRNS, AUSTRALIA, 17th-22nd AUGUST 2008)

Pace, N. R., A Molecular View of Microbial Diversity and the Biosphere, *Science* **276**: 734-740 (1997)

Pham, V. D., Konstantinidis, K. T., Palden, T., DeLong, E. F., Phylogenetic analyses of ribosomal DNA-containing bacterioplankton genome fragments from a 4000 m vertical profile in the North Pacific Subtropical Gyre, *Environ. Microbiol.* **10**: 2313-2330 (2008)

Pommier, T., Canbäck, B., Riemann, L., Boström, K. H., Simu, K., Lundberg, P., Tunlid,

A., Hagström, Å., Global patterns of diversity and community structure in marine bacterioplankton, *Mol Ecol* **16**: 867-880 (2007)

Qu, T., Mitsudera, H., Yamagata, T., Intrusion of the North Pacific waters into the South China Sea, *Journal of Geophysical Research* **105**: 6415-6424 (2000)

Quince, C., Curtis, T. P., Sloan, W. T., The rational exploration of microbial diversity. *ISME J* **2**: 997-1006 (2008)

Raguénès, G., Christen, R., Guezennec, J., Pignet, P., Barbier, G., *Vibrio diabolicus* sp. nov., a new polysaccharide-secreting organism isolated from a deep-sea hydrothermal vent polychaete annelid, *Alvinella pompejana.*, *Int. J. Syst. Bacteriol.* **47**: 989-995 (1997)

Ramp, S. R., Tang, T. Y., Duda, T. F., Lynch, J. F., Liu, A. K., Chiu, C. S., Bahr, F. L., Kim, H. R., Yang, Y. J., Internal solitons in the northeastern South China Sea Part I: sources and deep water propagation, *IEEE J. Ocean. Eng.* **29**: 1157-1181 (2004)

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

Roesch, L. F. W., Fulthorpe, R. R., Riva, A., Casella, G., Hadwin, A. K. M., Kent A. D., Daroub, S. H., Camargo, F. A. O., Farmerie, W. G., Triplett, E. W., Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J* **1**: 283-290 (2007)

Roh, S. W., Kim, K. H., Nam, Y. D., Chang, H. W., Park E. J., Bae, J. W., Investigation of archaeal and bacterial diversity in fermented seafood using barcoded pyrosequencing, *ISME J* **4**: 1-16 (2010)

Sabehi, G., Béjà, O., Suzuki, M. T., Preston, C. M., DeLong, E. F., Different SAR86

subgroups harbour divergent proteorhodopsins, *Environ. Microbiol.* **6**: 903-910 (2004)

Sabehi, G., Massana, R., Bielawski, J. P., Rosenberg, M., Delong, E. F., Béjà, O. Novel proteorhodopsin variants from the Mediterranean and Red Seas, *Environ. Microbiol.* **5**: 842-849 (2003)

Santelli, C. M., Orcutt, B. N., Banning, E., Bach, W., Moyer, C. L., Sogin, M. L., Staudigel, H., Edwards, K. J., Abundance and diversity of microbial life in ocean crust, *Nature* **453**: 653-656 (2008)

Schauer, M., Balagué, V., Pedrós-Alió, C., Massana, R., Seasonal changes in the taxonomic composition of bacterioplankton in a coastal oligotrophic system, *Aquat. Microb. Ecol.* **31**: 163-174 (2003)

Schauer, M., Balagué, V., Pedrós-Alió, C., Massana, R., Seasonal changes in the taxonomic composition of bacterioplankton in a coastal oligotrophic system, *Aquat. Microb. Ecol.* **31**: 163-174 (2003)

Schauer, R., Bienhold, C., Ramette, A., Harder, J., Bacterial diversity and biogeography in deep-sea surface sediments of the South Atlantic Ocean, *ISME J* **4**: 159-170 (2010)

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

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

Schuster, S. C., Miller, W., Ratan, A., Tomsho, L. P., Giardine, B., Kasson, L. R., Harris, R. S., Petersen, D. C., Zhao, F., Qi, J., Complete Khoisan and Bantu genomes

from southern Africa, *Nature* **463**: 943-947 (2010)

Schwalbach, M. S., Fuhrman, J. A., Wide-Ranging Abundances of Aerobic Anoxygenic Phototrophic Bacteria in the World Ocean Revealed by Epifluorescence Microscopy and Quantitative PCR, *Limnol. Oceanogr.* **50**: 620-628 (2005)

Shiah, F. K., Liu, K. K., Tang, C. Y., The South East Asia Time-series Study (SEATS), *US-JGOFS Newsletter*: 8-9. (1999)

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

Sorokin, D. Y., van Pelt, S., Tourova, T. P., Muyzer, G., Microbial isobutyronitrile utilization at haloalkaline conditions, *Appl. Environ. Microbiol.* **73**: 5574-5579 (2007)

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

Strom, S. L., Microbial Ecology of Ocean Biogeochemistry: A Community Perspective, *Science* **320**: 1043-1045 (2008)

Sverdrup, H. U., Johnson, M. W., and Fleming, R. H., The Oceans: their physics, chemistry and general biology (Prentice-Hall, Englewood Cliffs, New Jersey, 1942)

Tamura, K., Dudley, J., Nei, M., Kumar, S., MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0, *Mol. Biol. Evol.* **24**: 1596-1599 (2007)

Thompson, F. L., Thompson, C. C., Li, Y., Gomez-Gil, B., Vandenberghe, J., Hoste, I. B., Swings, J., *Vibrio kanaloae* sp. nov., *Vibrio pomeroyi* sp. nov. and *Vibrio*

*chagasii* sp. nov., from sea water and marine animals, *Int. J. Syst. Evol. Microbiol.* **53**: 753-759 (2003)

Thompson, J. D., Higgins, D. G., Gibson, T. J., CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucl. Acids Res.* **22**: 4673-4680 (1994)

Tunsjø, H. S., Paulsen, S. M., Mikkelsen, H., L'Abée-Lund, T. M., Skjerve, E., Sørum, H., Adaptive response to environmental changes in the fish pathogen *Moritella viscosa*, *Res Microbiol* **158**: 244-250 (2007)

Vandamme, P., and Coenye, T., Taxonomy of the genus Cupriavidus: a tale of lost and found, *Int. J. Syst. Evol. Microbiol.* **54**: 2285-2289 (2004)

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

Wen, L. S., Jiann, K. T., Santschi, P. H., Physicochemical speciation of bioactive trace metals (Cd, Cu, Fe, Ni) in the oligotrophic South China Sea, *Marine Chemistry* **101**: 104-129 (2006)

Whittaker, R. H., New concepts of kingdoms or organisms, *Science* **163**: pp.150-160 (1969)

Wierckx, N., Koopman, F., Bandounas, L., de Winde, J. H., Ruijssenaars, H. J., Isolation and characterization of *Cupriavidus basilensis* HMF14 for biological removal of inhibitors from lignocellulosic hydrolysate, *Microb Biotechnol* **3**: 336-343

(2010)

Woese, C. R., Bacterial Evolution, *Microbiol. Rev.* **51**: 221-271 (1987)

Woese, C. R., Fox, G. E., Phylogenetic structure of the prokaryotic domain: the primary kingdoms, *PNAS* **74**: 5088-5090 (1978)

Wong, G. T. F., Ku, T. L., Mulholland, M., Tseng, C. M., Wang, D. P., The SouthEast Asian time-series study (SEATS) and the biogeochemistry of the South China Sea—an overview, *Deep-Sea Res. II* **54**: 1434-1447 (2007)

Wong, G. T. F., Tseng, C. M., Wen, L. S., Chung, S. W., Nutrient dynamics and N-anomaly at the SEATS station, *Deep-Sea Res. II* **54**: 1528-1545 (2007)

Wu, J., Chung, S. W., Wen, L. S., Liu, K. K., Lee Chen, Y. L., Chen, H. Y., Karl, D. M., Dissolved inorganic phosphorus, dissolved iron, and *Trichodesmium* in the oligotrophic South China Sea, *Global Biogeochem. Cycles* **17**: 1008 (2003)

Yakimov, M. M., Golyshin, P. N., Lang, S., Moore, E. R., Abraham, W. R., Lunsdorf, H., Timmis, K. N., *Alcanivorax borkumensis* gen. nov., sp. nov., a new, hydrocarbon-degrading and surfactant-producing marine bacterium, *Int. J. Syst. Bacteriol.* **48**: 339-348 (1998)

Yim, M. S., Yau, Y. C. W., Matlow, A., So, J. S., Zou, J., Flemming, C. A., Schraft, H., Leung, K. T., A novel selective growth medium-PCR assay to isolate and detect Sphingomonas in environmental samples, *Journal of Microbiological Methods* **82**: 19-27 (2010)

Yoon, J. H., Lee, C. H., Yeo, S. H., Oh, T. K., *Psychrobacter aquimaris* sp. nov. and *Psychrobacter namhaensis* sp. nov., isolated from sea water of the South Sea in Korea. *Int. J. Syst. Evol. Microbiol.* **55**: 1007-1013 (2005b)

Yoon, J. H., Lee, C. H., Yeo, S. H., Oh, T. K., *Psychrobacter celer* sp. nov., isolated from sea water of the South Sea in Korea. *Int. J. Syst. Evol. Microbiol.* **55**:

1885-1890 (2005a)

Zeng, R., Zhao, J., Zhang, R., Lin, N., Bacterial community in sediment from the Western Pacific "Warm Pool" and its relationship to environment, *Science in China Series D: Earth Sciences* **48**: 282-290 (2005)

Zhang, J., Geochemistry of Trace Metals from Chinese River/Estuary Systems: An Overview, *Estuarine, Coastal and Shelf Science* **41**: 631-658 (1995)

Zhang, Y., Sintes, E., Chen, J., Dai, M., Jiao, N., Herndl, G. J., Role of mesoscale cyclonic eddies in the distribution and activity of Archaea and Bacteria in the South China Sea, *Aquat. Microb. Ecol.* **56**: 65-79 (2009)

Zheng, Q., Susanto, R. D., Ho, C. R., Song, Y. T., Xu, Q., Statistical and dynamical analyses of generation mechanisms of solitary internal waves in the northern South China Sea, *Journal of Geophysical Research* **112**: C03021 (2007)

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