2014

IDENTIFICATION OF SECONDARY METABOLITE GENE CLUSTERS OF BACTERIA FROM SOUTH PACIFIC GYRE SUBSEAFLOOR SEDIMENT

Liwei Zhu
University of Rhode Island, liwei_zhu@my.uri.edu

IDENTIFICATION OF SECONDARY METABOLITE GENE CLUSTERS OF BACTERIA FROM SOUTH PACIFIC GYRE SUBSEAFLOOR SEDIMENT

BY

LIWEI ZHU

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

OCEANOGRAPHY

UNIVERSITY OF RHODE ISLAND

2014

MASTER OF SCIENCE IN OCEANOGRAPHY THESIS

OF

LIWEI ZHU

APPROVED:

Thesis Committee:

Major Professor David C. Smith

David C. Rowley

Steven L. D'Hondt

Nasser H. Zawia

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2014

ABSTRACT

Secondary metabolites are organic compounds that are not directly involved in the key processes (growth, reproduction and development) of an organism. They are commonly targeted in pharmaceutical science for drug discovery. Secondary metabolites that have been used in drug discovery have been derived from plants, invertebrates and microbes. Microbes, bacteria in particular, have contributed greatly and will continue to play an important role in new drug discovery. Among the bacteria from all environments, marine bacteria are a vast reservoir for many potential useful bioactive compounds. Recent studies using marine bacteria for pharmaceutical use mainly focused on the bacteria collected from near-shore sediments. However, bacteria from deep-sea sediments remain unexplored.

The South Pacific Gyre (SPG) is the most oligotrophic region of the world ocean. Due to the low surface productivity and distance from land, sediments below the gyre accumulate very slowly and are characterized by very low organic carbon content and relatively high dissolved oxygen concentrations. Sediments from South Pacific Gyre were found to host a living microbial community that, compared to other marine sediments, contains very low microbial biomass and very low metabolic activity.

Thus, the goal of this study is to: (1) document the diversity of bacteria isolated in pure culture; and (2) explore the pharmaceutical potential of deep-sea

bacteria from South Pacific Gyre sediment. To address this, bacteria were isolated in pure culture from sediments from seven sites of the Integrated Ocean Drilling Program (IODP) Expedition 329 in the South Pacific Gyre. 16S rRNA genes from 81 bacterial isolates throughout six SPG sites (U1366, U1367, U1368, U1369, U1370 and U1371) were sequenced for phylogenetic analysis using the RDP (Ribosomal Database Project). 16S rRNA genes were amplified with bacterial primers that have been proven to amplify bacterial sequences well (27F, 1392R). Whole genomes from nine *Rhodococcus* isolates (with two isolates sequenced in duplicate) from four SPG sites (U1366, U1367, U1370 and U1371) were sequenced for secondary metabolites gene clusters discovery. By using antiSMASH (antibiotics & Secondary Metabolite Analysis SHell), secondary metabolite biosynthesis gene clusters in the bacterial genome were identified, annotated and analyzed.

Of the 81 16S rRNA gene clone libraries constructed, most of the clones (63%) affiliated with the genus *Bacillus*, 35.8% were affiliated with the genus *Rhodococcus* and one clone was identified as a *Corynebacterium*. The phylogenetic tree indicated that all the *Rhodococci* were identified as *Rhodococcus erythropolis*. By using antiSMASH to look for the secondary metabolites gene clusters from the *Rhodococcus* genomes, many gene clusters, most of which were non ribosomal peptides (NPRS) and polyketide synthases (PKS), were found in the genomes. This study suggests that deep-sea sediments harbor bacteria with the potential to produce pharmaceutically important secondary metabolites.

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my advisor Prof. David Smith for the continuous support of my study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. Without his supervision and constant help this thesis would not have been possible. I could not have imagined having a better advisor and mentor for my study.

Besides my advisor, I would like to thank the rest of my thesis committee, Prof. David Rowley, Prof. Steven D'Hondt, for their encouragement, insightful comments, and hard questions.

My sincere thanks go to Dr. Daniel Udwary, for his guidance and helpful suggestions on my study in bioinformatics. His guidance has served me well and I owe him my heartfelt appreciation. Special thank you go to Dr. Stephanie Forschner-Dancause, who prepared all the bacterial isolates for my research and friendly assistance at the very beginning of study in the project.

In addition, I would like to thank my fellow labmates in David Rowley's lab: Chrissy Dao, Rob Deering, Jiadong Sun, for the good companion and collaboration in the last two years. Also I thank my friends, Qianqian Liu, Wenjing Zhao, Xiangyu Mou, Janalyn Taylor, Amir Shirazi. Their friendship and assistance has meant me more to me than I could ever express. In particular, I am grateful to Dr. Gang Chen

and Dr. John Kirkpatrick for enlightening me of my research. I also want to thank the NIH for their financial support for my study (project number 1R15AI093158-01).

Furthermore, I should thank Genomics and Sequencing Center (GSC) and IDeA Network of Biomedical Research Excellence (INBRE) of URI for providing me with facilities and equipment for this study. Special thanks go to Janet Atoyan for sequencing my samples.

Last but no least, I wish to thank my parents, Qunying Zhang and Zhiyong Zhu, for giving birth to me at the first place and supporting me spiritually throughout my life. Their love provided my inspiration and was my driving force. I owe them everything and wish I could show them just how much I love and appreciate them. I also would like to thank my younger brother Jipeng Zhu, for bringing me a lot of happiness and cheering me up when I felt down.

PREFACE

This thesis is made as a completion of the Master of Science in Oceanography. It is one of the fruit of my labor in David Rowley's lab in College of Pharmacy and David Smith's lab in Graduate School of Oceanography over the past two years and half. During my research master studies, I gained experience in DNA and genome extraction, sequencing and bioinformatics methods. Although I went through some difficulties in the genome extraction and genome data analysis, it was delight that I had so many people to discuss with and finally got the problems solved.

I am grateful to all those people from Graduate School of Oceanography, College of Pharmacy and Center for Biotechnology & Life Sciences that were involved in this study. In particular, I would like to thank Prof. David Smith and Prof. David Rowley for giving me the opportunity to study here in United States and take part in the natural products project.

I hope you will enjoy reading this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

MANUSCRIPT-1

This thesis is prepared for submission to FEMS Microbiology Ecology

Identification of Secondary Metabolite Gene Clusters of Bacteria from South Pacific Gyre Subseafloor Sediment

Liwei Zhu[1], David C. Rowley[2], and David C. Smith[1]

[1]Graduate School of Oceanography, University of Rhode Island, Narragansett, RI, USA 02882

2 Biomedical and Pharmaceutical Sciences, University of Rhode Island, Kingston, RI, USA 02881

CHAPTER 1 INTRODUCTION

1.1 Microbial abundance in marine subsurface sediment

Beginning in 1986, on the Ocean Drilling Program (ODP) Leg 112, and on several subsequent Legs, sediment samples were analyzed with direct microscopic observation to estimate bacterial abundance with depth. This work established that bacteria were found in sediment to a depth of 518 mbsf. Whitman et al. (1998) used ODP cell abundance data to extrapolate the global abundance of bacterial biomass in the marine sediment and argued that the biomass of bacteria in marine sediments is equivalent to ~30 of the total biomass on Earth. Parkes et al. (2000) suggested that the biomass of subsurface marine bacteria is equivalent to ~10% of the total surface biosphere. Using a much richer dataset that considered factors of sedimentation rate and distance from shore, Kallmeyer et al. (2012) concluded that the microbial biomass in marine sediments is equivalent to ~0.6% of the total biomass on Earth. Despite this reduction in the revised estimate, marine sediment hosts an enormous amount of microbial biomass and therefore, represents a vast reservoir of genomic potential.

1.2 Microbial biosynthetic secondary metabolites

Antibiotic compounds found in drug discovery efforts have been derived from plants, invertebrates, vertebrates and microbes. Microbes have made enormous contributions to the health and well-being of people throughout the

world. More than two-thirds of the antibiotics used to treat humans are microbial natural products or semi-synthetic derivatives of these (Fischbach & Walsh 2009). Though secondary metabolic products can be utilized as pharmaceuticals (antimicrobials and anticancer agents etc.), secondary metabolism genes are often overlooked because of their evolutionary relationships to primary metabolic genes. Advances in genetics, biochemistry, and bioinformatics have contributed to the study of antibiotics and other natural products, not only by revealing how they are synthesized but also by casting them as phenotypes encoded by gene collectives that be studied through an evolutionary lens (Fischbach & Walsh 2009). Antibiotic-encoding gene collectives can converge evolutionarily on similar phenotypes just like other sets of genes that encode adaptive traits. Some distinct gene clusters on similar phenotypes have been converged into a single functional unit, called gene clusters. Examining antibiotics have provided an entry point for studying the natural roles of these natural products. Because the useful lifetime for an antibiotic is relatively short compared to the time that clinically significant resistance emerges (Walsh, 2003), the continuing and cyclical need for new antibiotics to combat the current generation of resistant pathogens compels the scientists and clinicians to search for new sources of antibiotics (Clardy et al. 2006).

Rhodococcus species have been shown to have potential as a commercial product. The wide range of chemicals transformed or degraded by *Rhodococcus* makes them good candidates for use in both environmental and industrial biotechnology. They play an important role in bioremediation and biodegradation

pollutants (Warhurst et al. 1994), biosurfactants and bioflocculants (Finnerty 1992), desulphurization of fossil fuels (Gray et al. 1996) and oil prospecting (Ashraf et al. 1994). Besides, a range of other transformations using *Rhodococcus* cells or enzymes for pharmaceutical use has been described. A novel and efficient biotransformation producing sec-cedrenol, a compound with potential medical value, has been described (Takigawa et al. 1993). Peters et al. (1993)report possible synthetic uses of carbonyl reductases from *Rhodococcus* to give a range of compounds that can be used for synthesis of pharmaceuticals and agrochemicals. Cholesterol oxidases of *Rhococcus rhodochrous* (Warhurst et al. 1994) have been studied and could have applications in the food industry or in steroid drugs' production (Finnerty 1992, Christodoulou et al. 1994, Kreit et al. 1994). Due to their remarkable metabolic versatility, *Rhodococcus* isolates were chosen for whole genome sequencing, in order to explore the pharmaceutical potential of *Rhodococcus* isolates from South Pacific Gyre deep sea subsurface sediment.

1.3 Integrated Ocean Drilling Program Expedition 329

In 2011, Integrated Ocean Drilling program (IODP) Expedition 329 cored and logged deep-sea sediment and basement at a series of seven sites in the South Pacific Gyre. Gyres are semi-still areas in the middle of the oceans where there is little wind, little current, and very little upwelling of deep nutrient rich water, thus the water is clear and nutrient-poor. The South Pacific Gyre (SPG) is the largest subtropical gyre on Earth. The center of South Pacific Gyre is considered the biggest

oceanic desert because of its low rate of primary production. In addition, the middle of the South Pacific Gyre is far from continents so there is no riverine input and little aeolian input. Seven SPG sites (Fig. 1) were explored during Expedition 329.

The seven sites are characterized by varying levels of productivity in the overlying water column (Fig. 2) as well as basement ages (6 Ma to 84-124.6 Ma). Of the seven drilling sites explored in the expedition (Fig. 1), six drilling sites (Site U1365-U1370) are in the South Pacific Gyre and the remaining site (Site U1371) is in the upwelling zone just south of the low-chlorophyll gyre (Expedition 329 Scientist, 2011). Mean sedimentation rates are extremely low in the low-chlorophyll region of the South Pacific Gyre. Mean sedimentation rate is higher at eastern-most sites where water depth is shallowest (allowing carbonate sediment to accumulate). The rate is highest at U1371, which is just outside of the central gyre because of sedimentation of siliceous debris. Mean sedimentation rates within the South Pacific Gyre are among the lowest that occur at the Earth's surface (D'Hondt et al. 2009).

The primary purpose of the expedition was to document the microbial communities and test and test the energetic limit to life in nutrient-poor deep-sea sediment. One of the fundamental questions about the subseafloor biosphere are addressed in this expedition is whether these sites contain previously unknown kinds of microorganisms. During the expedition, sediment samples were collected for subsequent laboratory-based attempts to isolate bacteria in pure culture (Forschner-Dancause 2012).

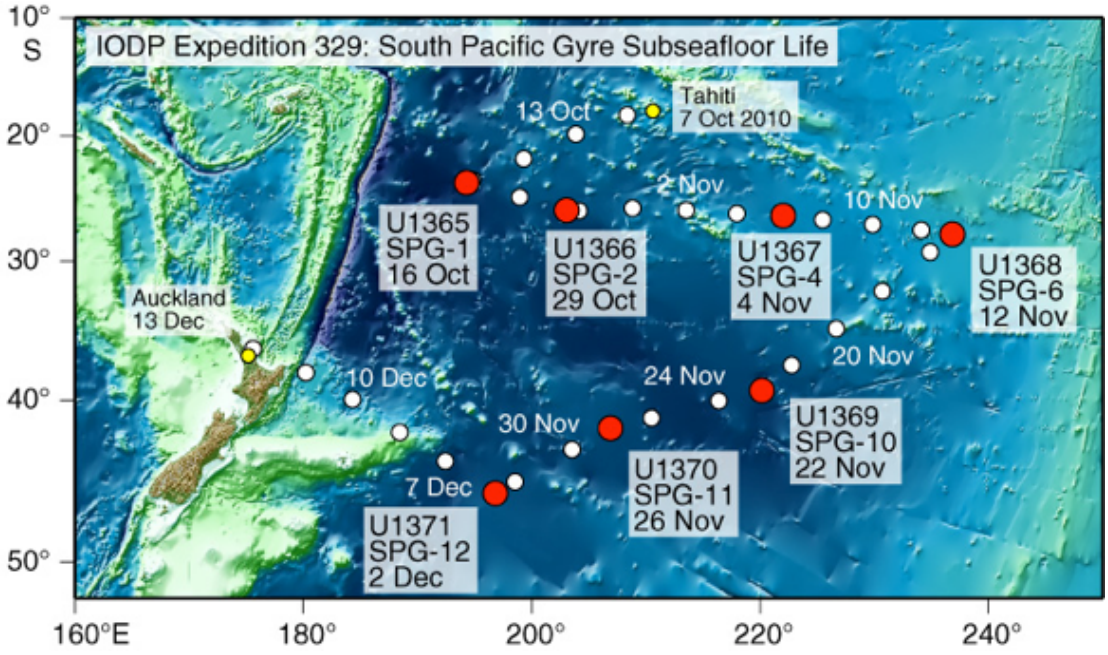


Figure 1 Seven sites explored in IODP EXP. 329

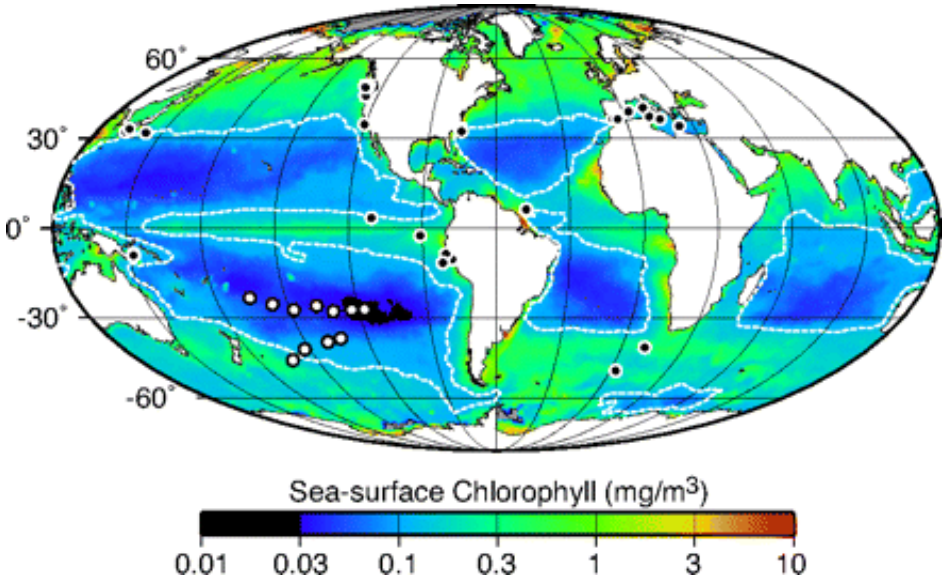


Figure 2 Site location on a map of time-averaged sea-surface chl-a concentrations (from (D'Hondt et al. 2009))

1.4 This study

This study examines the prospects for the discovery of novel microbial biosynthetic secondary metabolites from bacteria isolated from the SPG subseafloor sediment. The phylogentic affiliations of 81 bacteria previously isolated from SPG sediments (Forschner-Dancause 2012) were determined using 16S rDNA sequence analysis. Of the 81 isolates, nine were chosen for whole genome sequencing. The genomes were then analyzed for secondary metabolite gene clusters.

CHAPTER 2 MATERIALS AND METHODS

2.1 Core Handling and Sample Collection

The Integrated Ocean Drilling Program (IODP) Expedition 329, "South Pacific Gyre Subseafloor Life" was conducted in October – December 2010 onboard the drillship "JOIDES Resolution". Sediment core samples were collected at seven sites along two transects: from the western side of the gyre to the center of the gyre (Site U1368) and from the center southerly, ending at a control site in the upwelling region located southwest of the gyre (Fig. 1). Once a core was retrieved, it was immediately transferred to the catwalk for labeling and cutting of sections before the next core barrel was deployed (Expedition 329 Scientists 2011). This reduced the amount of time the core remained on deck and therefore minimized warming of the samples. All core sections to be sampled for microbiological studies were transferred from the drilling platform to the Hold Deck refrigerator (~7 °C – 10 °C) as quickly as possible and kept as whole-core sections until processed. The core liner was cut by the standard IODP core cutter and cut with an ethanol-wiped spatula. Since the core liner is not sterile and the outer surface of the core is contaminated during drilling, subsampling of whole-round cores excluded the sediment next to the core liner. Subcores were taken with sterile 5 cm^3 syringes that had the luer-lock end removed. The subcores were remained in the syringe barrel inside sterile bags and stored and shipped at 4 °C for future onshore cultivation efforts (Expedition 329 Scientists 2011).

2.2 Bacterial Cultivation and Cultured Strains Purification

Bacteria were isolated in pure culture by Dr. Forschner-Dancause (Forschner-Dancause 2012). Sediment slurries were made by aseptically removing the exposed end of the subcore and then sampling approximately 0.5 cm^{-3} of sediment from the center of the core. The samples were added to a 4 mL vial containing 1 mL sterile artificial seawater (30 g Instant Ocean per 1 L diH_2O) and mixed using a vortex mixer. The resulting slurry was allowed to settle. 100 µL of overlaying water was drawn off the slurry and spread on plates made of various marine media. The sealed plates were incubated at room temperature and observed daily for bacterial growth.

Individual bacterial colonies were sampled with a sterile loop and streaked out onto a fresh isolation plate of the same media from which it was isolated. To purify the colony, a well isolated colony was streaked from the first isolation plate onto a second one of the same media and then onto a third one containing Yeast Peptone media (1 g yeast extract, 5 g peptone, 22.5 g Instant Ocean and 15 g agar per 1 L diH_2O). For cryopreservation into the culture collection, the strain was grown in YP broth (1 g yeast extract, 5 g peptone, 22.5 g Instant Ocean and 1 L diH_2O) for 48 hrs at 24 °C, shaking at 175 rpm. For preservation of the cells, glycerol was added to the culture for a final concentration of 20% and the strain was frozen at -80 °C (Forschner-Dancause 2012).

2.3 16S rDNA sequencing

DNA Extraction

For this project, frozen bacterial isolates were revived by incubating them on YP media plates for 3 to 4 days at room temperature. Well isolated colonies were streaked onto fresh media and allowed to grow. Colonies from the new plates were picked and the DNA was extracted using the Lyse and Go PCR Reagent (Thermo Scientific, Cat. No.: PI78882) according to the manufacturer's instructions. Each colony was suspended in a 10 µL Lyse and Go Reagent.

16S rDNA PCR

The primers used in the DNA amplification were: 10 µM B27F and 10 µM B1392R (Table 1). The 5 µL Bacterial DNA extract and controls were amplified with 25 µL Taq master mix, 1 µL B27F, 1 µL B1392R and 18 µL diH_2O. Amplification conditions for PCR were as follows: 2 min at 94 °C to denature the DNA, followed by 30 cycles of: denaturation at 94 °C for 30 s, primer annealing at 60 °C for 1 min and strand denaturation at 72 °C for 2 min in a thermal cycler (Rotorgene). PCR products were separated on a 1% agarose gel and DNA bands were visualized with ethidium bromide. The recipe of 1% agarose gel was: 35 mL 1xTAE and 0.35 g agar. 3 µL sample were well mixed with 2 µL loading buffer and 3 µL PCR product and then added into the wells of the agarose gel.

Table 1 Bacterial 16S rDNA targeted PCR primers

PCR primer	Sequence (5'-3')	Melting Temp.(℃)
27F	AGA GTT TGA TCC TGG CTC AG	54.3
1392R	ACG GGC GGT GTG TRC	57.4

Primers and excess nucleotides were removed from the amplified DNA using QIAquick PCR Purification Kit (QIAGEN, Cat. No.: 28104). The amount of DNA in the cleaned-up product was quantified spectrophotometrically (NanoDrop 2000) in Rhode Island INBRE (IDeA Network of Biomedical Research Excellence) core lab. The DNA templates for sequencing were prepared in 12 μL reaction volumes containing primers and molecular biology grade water according to the following standards: (1) PCR products: 2.5 ng DNA/100 bases per reaction; (2) Plasmid templates: 300-500 ng per reactions; (3) Primers: 5 pmol per reaction. In order to facilitate pipetting to mix the solution well, the reactions were submitted to Genomics and Sequencing Center (GSC) in URI in duplicate (24 μL) and more than 16 reactions were prepared in strip-tubes.

Sequence Analysis

Purified PCR products were sequenced in the forward and reverse direction in separate reactions. The forward and reverse sequences for each sample were aligned using Geneious software 7.0.6 to obtain a composite sequence. The quality of each sequence trace was visually assessed and poor quality sequence was edited and removed manually. Samples were identified for each assay by running the

composite sequence though the Ribosomal Database Project (RDP) Classifier. The aligned 16S rRNA gene sequences were assigned a taxonomical hierarchy by using the Classifier analysis tools in RDP. The RDP obtains bacterial rRNA sequences from the International Nucleotide Sequence Databases (GenBank/EMBL/DDBJ) on a monthly basis. RDP Classifier placed sequences into the RDP hierarchy. It could be used to give an initial taxonomic placement for sequences (Cole et al. 2005).

Phylogenetic analysis was performed after the RDP analysis in order to determine the evolutionary relationships among various organisms based on similarities and differences in their genetic sequence. The phylogenetic tree was reconstructed by neighbor-joining method using MUSCLE alignment and calculating evolutionary distances by the Jukes and Cantor method that assumes an equal substitution rate among the four nucleotides. Bootstrap values were calculated from 100 replicate trees and sequence identity determined by database comparisons refined with phylogenetic analysis. The 16S rRNA sequences were transformed to the FASTA format and then analyzed using the Geneious alignment and assembly software. All the 16S rRNA sequences were aligned using E. coli as the outgroup. This was achieved by following the above steps: (1) Find and download the reference sequences by clicking “Sequence Search” in Geneious (Database: nr; Program: Megablast; Max Hits: 100); (2) Choose all the consensus sequences and reference sequences; (3) Run the “Multiple Alignment” under the tab “Align/Assemble” (93% similarity); (3) A new alignment sequence will be built, trim

all the aligned consensus sequences to equal lengths and save the files; (4) Choose *E. coli* AY776275.1 as an outgroup.

2.4 Whole Genome Sequencing

Whole Genome Extraction

Genomic DNA of nine *Rhodococcus* isolates (with two individual isolates sequenced in duplicate) was extracted by using Wizard® Genomic DNA Purification Kit (Promega, Cat. No.: A1120). Considering that *Rhodococcus* is a Gram-positive bacterium that has cell envelopes made of a thick layer of peptidogluycans, several steps in the protocol were revised to achieve the optimal results. DNA was extracted according to the steps below: (1) Add 1 ml of overnight culture to a 1.5 ml microcentrifuge tube; (2) Centrifuge at 16,000 x g for 2 minutes to pellet the cells. Remove the supernatant; (3) Add 1 ml of overnight culture to a 1.5 ml microcentrifuge tube; (4) Centrifuge at 16,000 x g for another 2 minutes to pellet the cells. Remove the supernatant; (5) Resuspend the cells thoroughly in 480 µl of 50 mM EDTA; (6) Add 120 µl 10 mg/ml lysozyme (Sigma, Cat. No.: L7651) to the resuspended cell pellet, and gently pipet to mix; (7) Incubate the sample at 37 °C for 2 hours. Centrifuge for 2 minutes at 16,000 x g and remove the supernatant; (8) Add 600 µl of Nucleic Lysis Solution. Gently pipet until the cells are resuspended; (9) Incubate at 80 °C for 10 minutes to lyse the cells, then cool to room temperature for 10 minutes; (10) Add 3 µl of RNase Solution to the cell lysate. Invert the tube 5-10 times to mix; (11) Incubate at 37 °C for 2 hours. Cool the sample to room

temperature for 10 minutes; (12) Add 200 µl of Protein Precipitation Solution to the RNase-treated cell lysate. Vortex vigorously at high speed for 30 seconds to mix the Protein Precipitation Solution with the cell lysate; (13) Incubate the sample on ice for 10 minutes; (14) Centrifuge at 16,000 x g for 3 minutes; (15) Transfer the supernatant containing the DNA to a clean 1.5 ml microcentrifuge tube containing 600 µl of room temperature isopropanol; (16) Gently mix by inversion until the thread-like strands of DNA form a visible mass; (17) Centrifuge at 16,000 x g for 2 minutes; (18) Carefully pour off the supernatant and drain the tube on clean absorbent paper. Add 600 µl of room temperature 70 % ethanol and gently invert the tube several times to wash the DNA pellet; (19) Centrifuge at 16,000 x g for 2 minutes. Carefully aspirate the ethanol; (20) Drain the tube on clean absorbent paper and allow the pellet to air-dry for 15 minutes; (21) Add 100 µl of DNA Rehydration Solution to the tube and rehydrate the DNA by incubating the solution overnight at room temperature; (22) Store the DNA at 2-8 °C.

Genome Sequences Analysis

The whole genome sequencing work was conducted at the RI Genomics and Sequencing Center, University of Rhode Island by using a next generation sequencing technology (NGS) instrument (Illumina MiSeq). Next generation sequencing is the generic term for methods that simultaneously sequence millions of small fragments of DNA prepared from an entire genome, transcriptome, or

smaller targeted regions in a single run of the instrument. Genomic DNA was fragmented using a focused ultra-sonicator (Corvaris S220) in order to produce DNA fragments in the ~500 – 700 bp range. The library is then linked to adaptors that enable the fragments to bind to a glass slide or flow cell. Eleven different adaptors were used in this experiment (Table 2). In the Illumina instruments, the immobilized templates are clonally amplified to generate millions of molecular clusters each containing ~1,000 copies of the same template. The clustered templates are then sequenced using Illumina's sequencing-by-synthesis technology. In this process, the addition of fluorescently labeled nucleotides liberates one of four colors that are detected by laser excitation and high-resolution cameras in every run cycle (Mardis 2008).

Table 2 Adaptors used for whole genome sequencing

Adaptor	Sequence (5'-3')	Target
Treseq Universal Adaptor	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T	All
Truseq Adaptor, Index 1	GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC ATC ACG ATC TCG TAT GCC GTC TTC TGC TTG	MZ1
Truseq Adaptor, Index 2	GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC CGA TGT ATC TCG TAT GCC GTC TTC TGC TTG	MZ2
Truseq Adaptor, Index 3	GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC TTA GGC ATC TCG TAT GCC GTC TTC TGC TTG	MZ3
Truseq Adaptor, Index 4	GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC TGA CCA ATC TCG TAT GCC GTC TTC TGC TTG	MZ4
Truseq Adaptor, Index 5	GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC ACA GTG ATC TCG TAT GCC GTC TTC TGC TTG	MZ5
Truseq Adaptor, Index 7	GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC CAG ATC ATC TCG TAT GCC GTC TTC TGC TTG	MZ7
Truseq Adaptor, Index 8	GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC ACT TGA ATC TCG TAT GCC GTC TTC TGC TTG	MZ8
Truseq Adaptor, Index 9	GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC GAT CAG ATC TCG TAT GCC GTC TTC TGC TTG	MZ9
Truseq Adaptor, Index 10	GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC TAG CTT ATC TCG TAT GCC GTC TTC TGC TTG	MZ10
Truseq Adaptor, Index 11	GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC GGC TAC ATC TCG TAT GCC GTC TTC TGC TTG	MZ11
Truseq Adaptor, Index 12	GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC CTT GTA ATC TCG TAT GCC GTC TTC TGC TTG	MZ6

The raw genome sequences were processed with CLC Genomics Workbench v6.0.4 by following the above steps before being analyzed: (1) Sequences Editing: the raw genome sequences were put into a new folder by using the sequencing analysis with Geneious software package. All ABI files of the raw sequences were

edited by deleting the uncertain nucleic acid base, which is denoted as "N" in the raw sequences. Thus all the genome sequences will be composed by only 4 nucleic acid bases afterwards, which are "A", "G", "C" and "T"; (2) Sequences Assembling: both the forward sequence and reverse reads of each sample were assembled by using "De Novo Assembly". A new sequence, which is the assembled sequence, was built; (3) Whole genome summary reports were assembled in CLC.

antiSMASH (http://www.secondarymetabolites.org/) is the first freely available comprehensive software package capable of identifying biosynthetic loci covering the whole range of known secondary metabolite compound classes (polyketides, non-ribosomal peptides, terpenes, aminoglycosides, aminocoumarins, indolocarbazoles, lantibiotics, bacteriocins, nucleosides, beta-lactams, butyrolactones, siderophores, melanins and others). The program aligns the identified regions at the gene cluster level to their nearest relatives from a database containing all other known gene clusters, and integrates or cross-links all previously available secondary-metabolite specific gene analysis methods in one interactive view (Medema et al. 2011). Uploading the assembled sequence to antiSMASH, a report containing some known classes of secondary metabolite biosynthesis gene clusters with detailed NRPS functional annotation and chemical structure of NPRS is generated.

CHAPTER 3 RESULTS

3.1 16S rDNA Sequencing

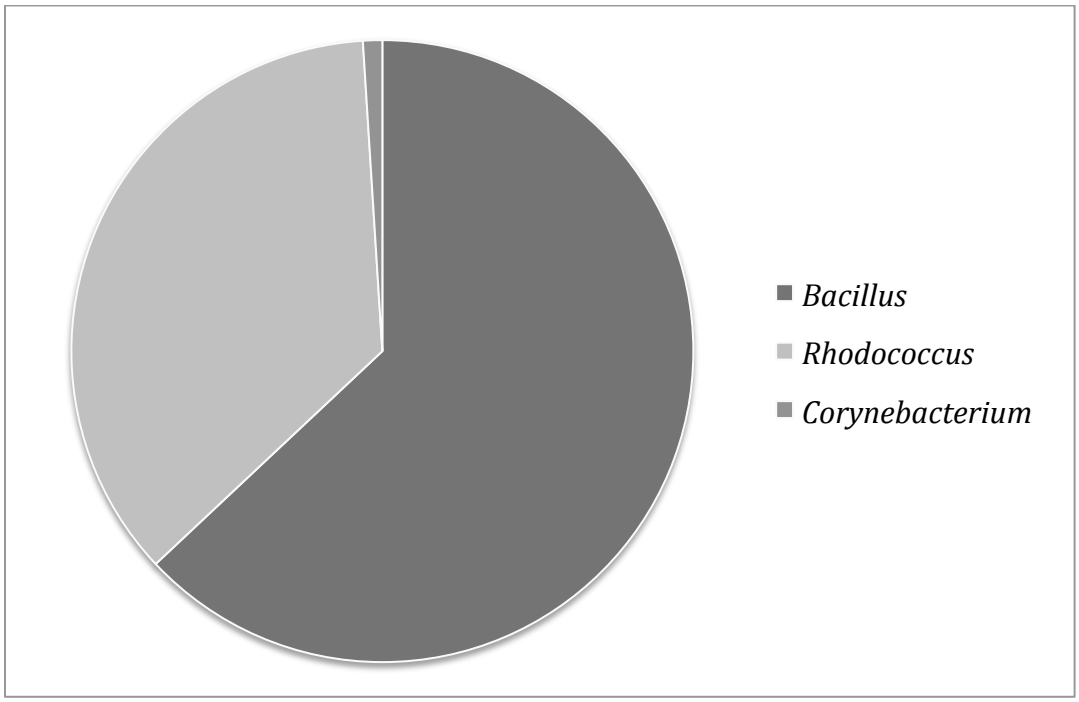


Figure 3 Three genera identified in the 81 strains

According to the partial sequencing of the 16S rDNA, the 81 isolated bacteria were identified as belonging to three genera, *Rhodococcus* (aerobic, nonsporulating, nonmotile, Gram-positive), *Bacillus* (obligate aerobic/facultative aerobic, rod-shaped, Gram-positive) and *Corynebacterium* (rod-shaped, Gram-positive). Most of the strains belonged to *Bacillus* (51 strains) and *Rhodococcus* (29 strains). Only one strain belonged to *Corynebacterium*.

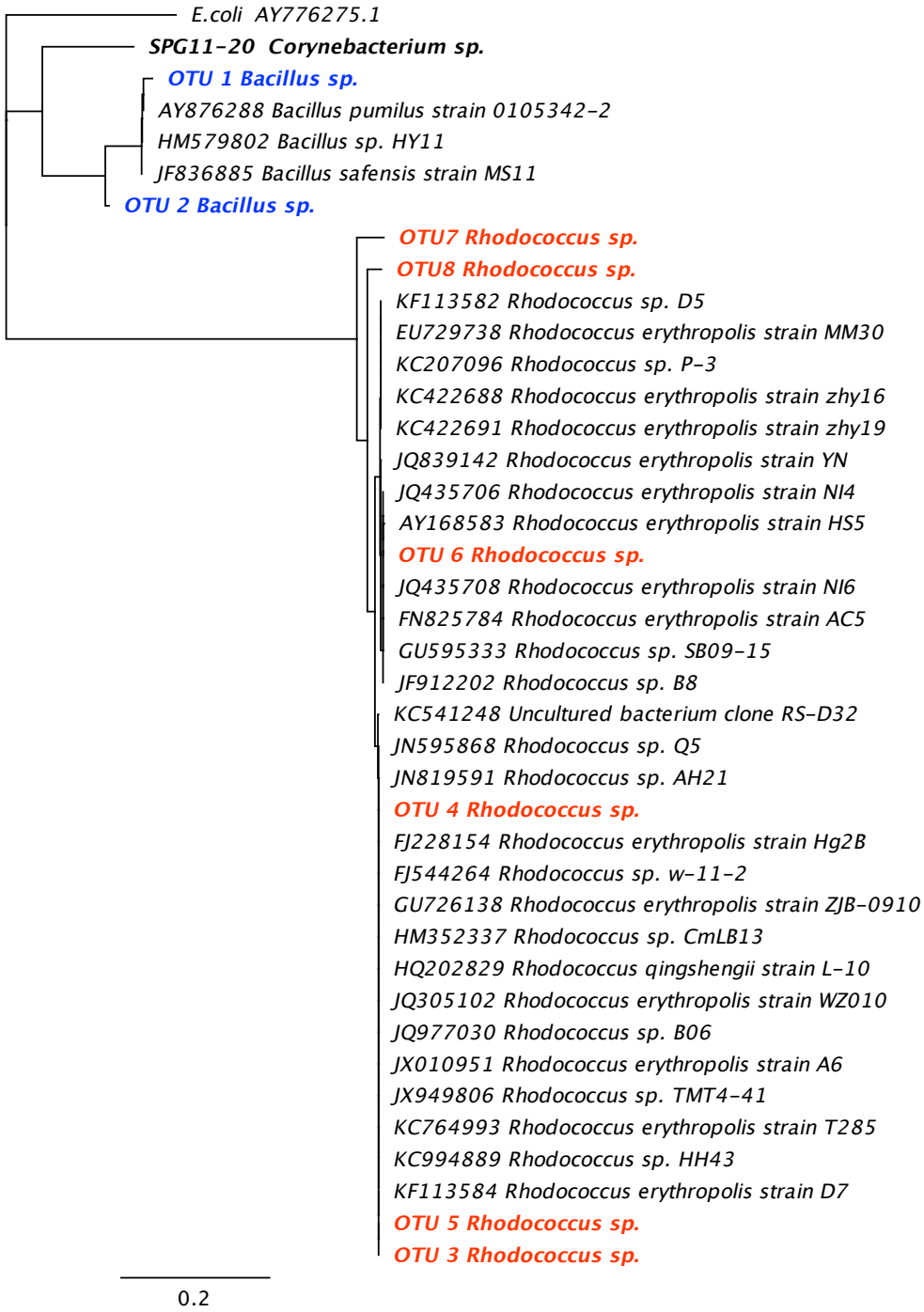


Figure 4 Neighbor-joining tree showing the phylogenetic relationships of 16S rDNA of 81 strains

Ten different operational taxonomic units (OTUs), defined as ≤97 % similarity between sequences were found among the 51 *Bacillus* strains (Appendix A=B) and 29 *Rhodococcus* (Appendix A). Two OTUs – OTU 7 (*Rhodococcus*) and OTU 8 (*Rhodococcus*) may represent new species, since they showed less than 97% similarity to sequences previously submitted to the database/published sequences. Each of these sequences were found in only one strain. Eight OTUs were identified in the *Rhodococcus* genus, which were attributed to species *Rhodococcus erythropolis* and *Rhodococcus qingshengii*. Two OTUs were identified in the *Bacillus* genus, which were attributed to species *Bacillus pumilus* and *Bacillus safensis*.

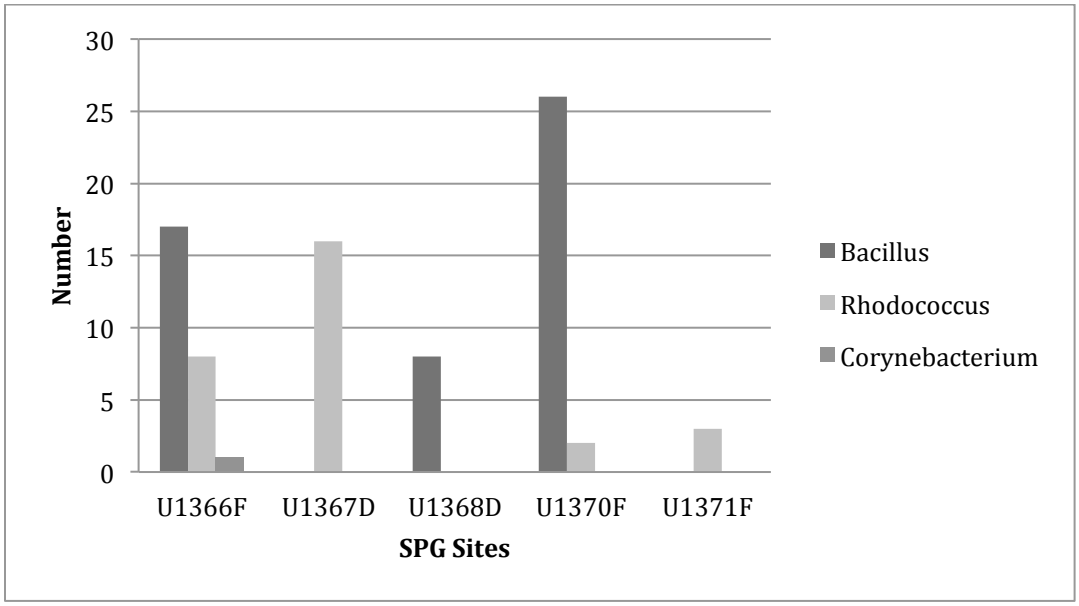


Figure 5 Distribution of 81 strains throughout five SPG sites

All the 81 bacterial isolates were distributed throughout five SPG sites, U1366F, U1367D, U1368D, U1370F and U1371F. The *Corynebacterium* strain was distributed in U1366F. Both U1366F and U1370F had *Bacillus* strains and *Rhodococcus* strains identified. U1367D and U1371F had *Rhodococcus* strains identified only, and U1368D had *Bacillus* strains identified only.

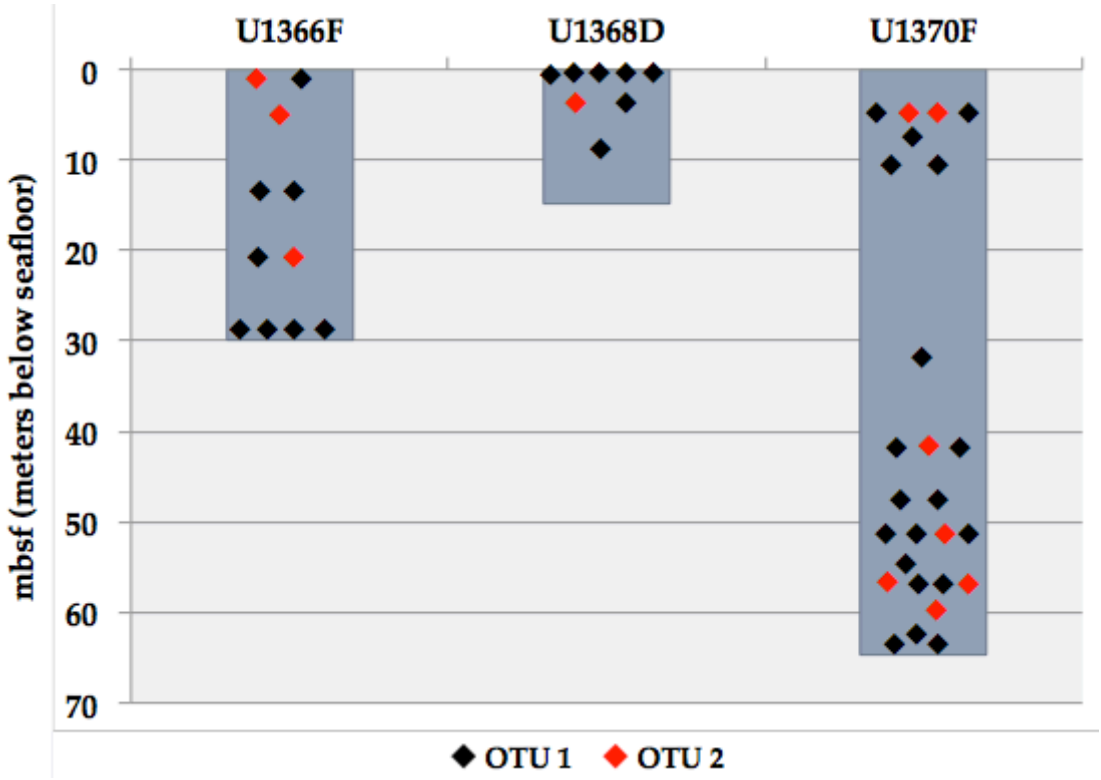


Figure 6 Depth profile of *Bacillus* strains identified

29 *Bacillus* strains from three SPG sites – U1366F, U1368D and U1370F – were identified. No tendencies were observed in *Bacillus* strains, in relation to the depth. Both OTU1 and OTU2 were distributed in each of the three sites.

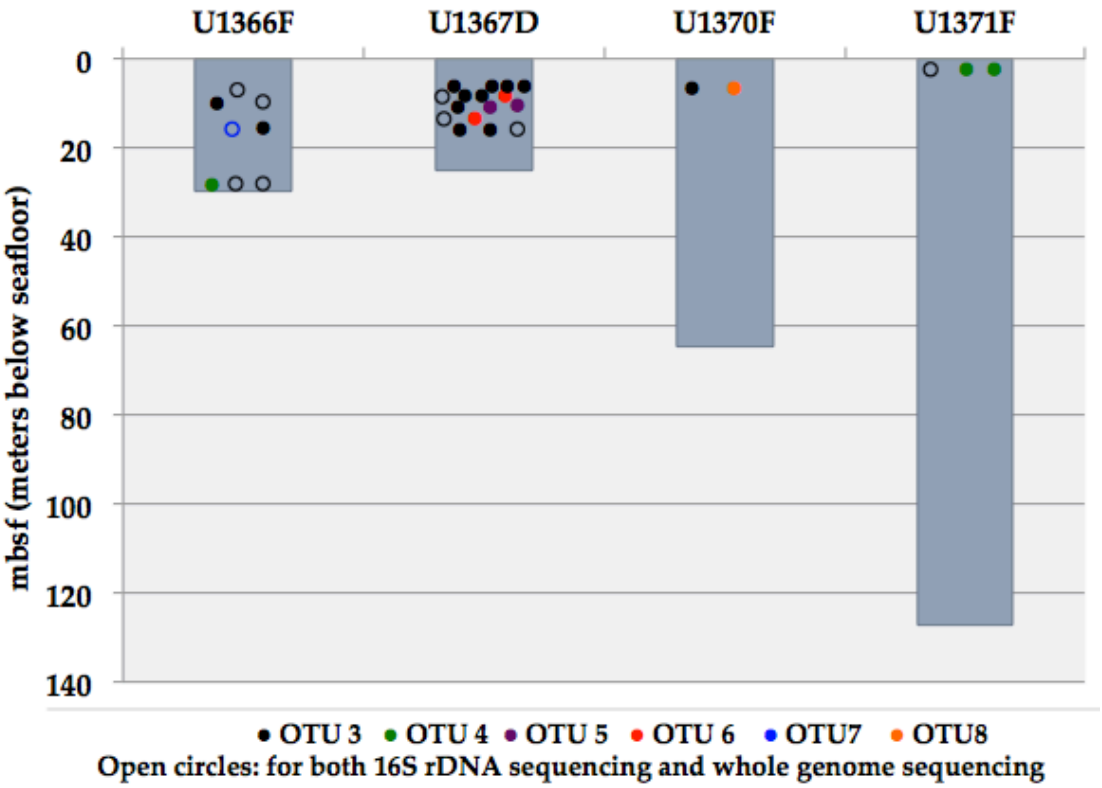


Figure 7 Depth profile of *Rhodococcus* strains identified

51 *Rhodococcus* strains from five SPG sites – U1366F, U1367D, U1370F and U1371F - were identified. *Rhodococcus* strains showed the tendency to distribute in the shallower sediment. No *Rhodococcus* strains were identified in sediment deeper than 30 mbsf. More than two OTUs were distributed in each of the five sites. Two OTUs – OTU3 and OTU4 – were present in Site U1366F. Three OTUs – OTU3, OTU5 and OTU6 – were present in Site U1367D. Two OTUs – OTU3 and OTU8 – were present in Site U1370F. Two OTUs – OTU3 and OTU4 – were present in Site U1371F.

Table 3 Sediment Properties of the sites of 9 Rhodococcus strains

Isolate	Site	Hole	Core	Section	Lat.*	Longti.*	le Depth (mb	Sediment (mer Depth (mb	Basement Age (Ma)*
MZ1	1366	F	2H	6	26°03.0836' S	56°53.6937' W	13.3	8.2	5127.0	100.00
MZ2	1371	F	1H	2	45°57.8502' S	63°11.0369' W	2.7	4.98	5306.0	73.00
MZ3	1366	F	4H	1	26°03.0836' S	56°53.6937' W	24.0	8.2	5127.0	100.00
MZ4	1366	F	4H	1	26°03.0836' S	56°53.6937' W	24.0	8.2	5127.0	100.00
MZ5	1366	F	3H	1	26°03.0836' S	56°53.6937' W	14.2	8.2	5127.0	100.00
MZ6	1367	D	2H	2	26°28.8861' S	37°56.3659' W	9.8	7.24	4285.0	33.50
MZ7	1367	D	3H	1	26°28.8861' S	37°56.3659' W	17.8	7.24	4285.0	33.50
MZ10	1366	F	3H	3	26°03.0836' S	56°53.6937' W	17.5	8.2	5127.0	100.00
MZ11	1367	D	2H	6	26°28.8861' S	37°56.3659' W	15.2	7.24	4285.0	33.50

*: Data from D'Hondt et al, 2009 ** data from LIMS report

3.2 Whole Genome Sequencing

Nine *Rhodococcus* strains' whole genomes were sequenced (Table 3). The nine *Rhodococcus* strains were derived from eight different cores from three SPG sites, U1366F, U1367D and U1371F. Five *Rhodococcus* strains (MZ1, MZ3, MZ4, MZ5 and MZ10) were retrieved from Site U1366F. Three *Rhodococcus* strains (MZ6, MZ7 and MZ11) were retrieved from Site U1367D. Only one *Rhodococcus* strains (MZ2) was retrieved from Site U1371F.

Table 4 Secondary metabolite gene clusters identified

Secondary Metabolite	MZ1	MZ2	MZ3	MZ4	MZ5	MZ6	MZ7	MZ10	MZ11
Terpene	1	1	1	1	1	1	1	1	1
Phenazine	1	0	1	0	0	0	0	0	0
Butyrolactone	1	1	1	1	1	1	1	1	1
Bacteriocin	2	2	2	2	2	2	2	2	2
Ectoine	1	1	1	1	1	1	1	1	1
Amglyccycl	0	1	0	1	1	1	1	1	1
NRPS	10	11	12	11	11	11	11	11	12
T1PKS	2	2	2	2	2	2	2	2	2
Unknown	3	3	4	3	3	3	3	3	2
NRPS-Terpene	1	1	1	1	1	1	1	1	1
Lantipeptide	1	0	0	0	0	0	0	0	0

According to results from the genome analyses using antiSMASH (Table 4), 23~24 secondary metabolites gene clusters, most of which are NRPSs (nonribosomal peptides) and PKSs (Polyketide synthases) were identified from each isolate. Five different NRPS domain patterns (Figure 7) were found in the *Rhodococcus* genomes in this study. The other secondary metabolites gene clusters

that are commonly found in the genome are: ectoine, butyrolactone. phenazine, bacteriocin, and terpene.

NRPs (nonribosomal peptides) are a class of peptide seconday metabolites, usually produced by microorganism such as bacteria and fungi. NRPs are synthesized by nonribosomal peptides, which, unlike ribosomal, are independent of messenger RNA. Each nonribosomal peptide synthetase can synthesize only one type of peptide. Nonribosomal peptides often have a cyclic and/or branched structures, can contain non-proteinogenic amino acids including D-amino acids, carry modifications like N-methyl and N-formyl groups, or are glycosylated, acylated, halogenated, or hydroxylated). Nonribosomal peptides are a very diverse family of natural products with an extremely broad range of biological activities and pharmacological properties. They are often toxins, siderophores, or pigments. Nonribosomal peptide antibiotics, cytostatics, and immunosuppressants are in commercial use.

PKSs (Polyketide synthases) are a family of multi-domain enzymes or enzyme complexes that produce polyketides, a large class of secondary metabolites, in bacteria, fungi, plants, and a few animal lineages. The biosyntheses of polyketides share striking similarities with fatty acid biosynthesis (Khosla et al. 1999, Jenke-Kodama & Dittmann 2005).

Ectoine (1, 4, 5, 6-Tetrahydro-2-methyl-4-pyrimidinecarboxylic acid) is a natural compound found in several species of bacteria. It is a compatible solute produced by aerobic chemoheterotrophic and halophilic/halophilic bacteria, where

it acts as osmoprotectant and effective biomembrane stabilizer, protecting the producing cells from extreme environmental stress. A recent study (Abdel-Aziz et al. 2013) shows that these intestinal barrier stabilizers from natural sources could offer new therapeutic measures for the management of IBD (inflammatory bowel disease).

Butyrolactone (C4H6O2) is a liquid anhydride of butyric acid which is used as a solvent in the manufacture of plastics. Gamma-butyrolactone is a common solvent and reagent in chemistry. In humans it acts as a prodrug for GHB (gamma-hydroxybutric acid) which is a naturally occurring substance found in the human central nervous system. Gamma-butyrolactone is also used as a recreational intoxicant with effects similar to alcohol.

Phenazine (C12H8N2 or C6H4N2C6H4) is also called azophenylene, dibenzopdiazine, dibenzopyrazine, and acridizine. It is a dibenzo annulated pyrazine and the parent substance of many dyestuffs, such as the euhodines, toluylene red, indulines and safranines. Many phenazine compounds are found in nature and are produced by bacteria such as Streptomyces sp.. These phenazine natural products have been implicated in the virulence and competitive fitness of producing organisms. For example, the phenazine pyocyanin produced by Pseudomonas aeruginosa contributes to its ability to colonize the lungs of cystic fibrosis patients. Similarly, phenazine-1-carboxylic acid, produced by a number of Pseudomonas, increases survival in soil environments and has been shown to be essential for the

biological control activity of certain strains (Turner & Messenger 1986, McDonald et al. 2001).

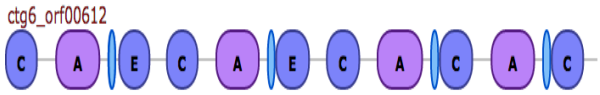
Bacteriocins are proteinaceous toxins produced by bacteria to inhibit the growth of similar or closely related bacterial strains (Farkas-Himsley 1980). Some bacteriocins from LAB (lactic acid bacteria) play an important role in food fermentation by exerting their anti-microbial action by interfering with the cell wall or the membrane of target organisms, either by inhibiting cell wall biosynthesis or causing pore formation, subsequently resulting in death.

Terpenes are a large and diverse class of organic compounds. They are often strong-smelling as many terpenes are aromatic hydrocarbons. A large number of terpenes are present in both terrestrial and marine plants, as well as fungi. However, in the last 5-10 years, it has become evident that terpenes are also produced by numerous bacteria, especially Gram-positive organisms such as Streptomyces and other Actinomycetes (Cane & Ikeda 2012).

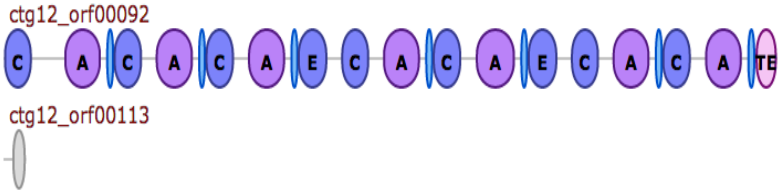
As shown on Figure 8, each major NRPS domain patterns are consisted of the following main domains: A, C, E, AT, KR, TE and PCP. Three modules were found to characterize the structure of NRPS: Module 1 (C-A-PCP), Module 2 (C-A-PCP-E) and Module 3 (C-A-PCP). A domain activates its related amino acid and catalyzed the transfer of the activated substrate to the PCP domain of the same module. C domain is responsible for the formation of the C-N bond between the elongated chain and the activated amino acid. E domain is an auxiliary domain that changes an L-amino acid into a D-amino acid. AT domains are responsible for the incorporation of

malonyl of methylmalonyl-CoA monomers. KR domains perform ketoreduction. The TE domain releases the final peptide product from the enzyme through cyclization or hydrolysis (Nikolouli & Mossialos 2012). Because NRPS clusters terminate with a TE domain, patterns 1 and 5 are incomplete.

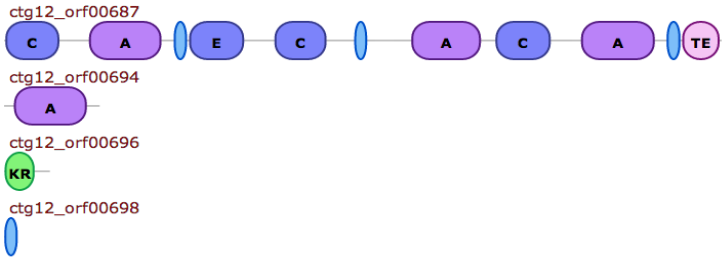
Pattern 1:



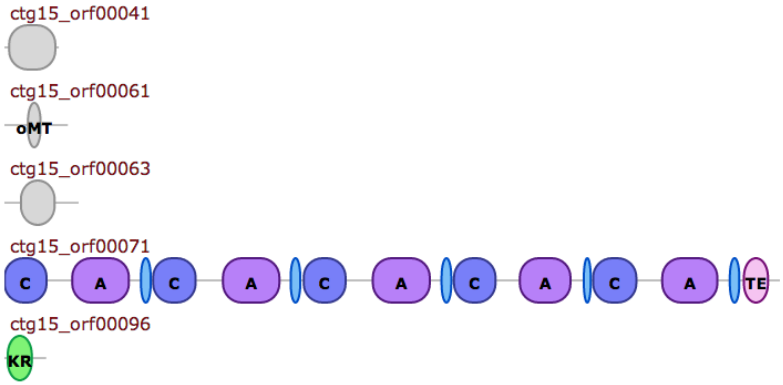
Pattern 2:



Pattern 3:



Pattern 4:



Pattern 5:

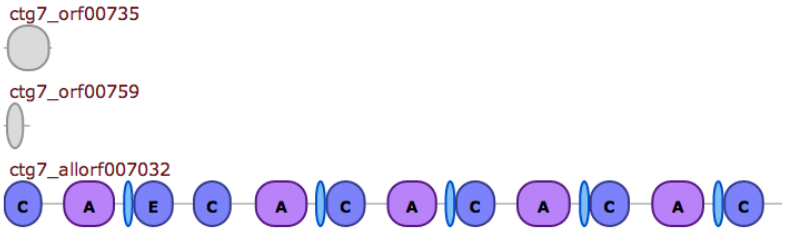


Figure 8 Five NRPS domain patterns in *Rhodococcus* genomes

NRPS domains: A adenylation, C condensation, E epimerization, AT acyltransferase KR ketoreductase, TE thioesterase, O (w/o letter inside) PCP (peptidyl-carrier).

Table 5 Number of NRPS domain patterns in each *Rhodococcus* strain

	P*1	P*2	P*3	P*4	P*5
MZ1	1	1	1	1	0
MZ2	1	1	0	1	1
MZ3	1	1	0	1	0
MZ4	1	1	0	1	1
MZ5	1	1	1	1	1
MZ6	1	1	0	1	1
MZ7	0	1	1	1	1
MZ10	1	1	0	1	1
MZ11	1	1	1	1	1

Five major domain patterns of NRPS in total were found in the *Rhodococcus* genomes, with Pattern 2 and Pattern 4 present in all 11 *Rhodococcus* strains' genomes. Two *Rhodococcus* isolates' genomes - MZ5 (from Site U1366F) and MZ11 (from Site U1367D) - present the gene clusters of all the five NRPS domain patterns.

Table 6 Comparison of sequence assembly results

Rhodococcus	Genome Size (Mb)	N50 (bp)	G+C %	Contigs No.
MZ1	7.29	575,310	62	223
MZ2	6.979	609,266	62	225
MZ3	7.276	446,761	62	194
MZ8	6.491	676,405	62	146
MZ3&8	7.368	419, 494	62	512
MZ4	6.491	676,405	62	146
MZ9	6.52	600,857	62	250
MZ4&9	6.644	676,409	62	594
MZ5	6.589	672,798	62	141
MZ6	6.575	558,109	62	122
MZ7	6.582	600,857	62	126
MZ10	6.583	650,263	62	140
MZ11	6.599	650,186	62	167

In order to test that if higher coverage (40x) could reveal more secondary metabolite gene clusters, two sets of duplicate samples of *Rhodococcus* DNA (MZ8 and MZ9) were run in the Miseq Illumina sequencing system together simultaneously with the other *Rhodococcus* strains. MZ8 is a duplicate of MZ3, and MZ9 is a duplicate of MZ4. The N50 value is the minimum contig length of the 50% of the total contig length. It is calculated by summarizing the lengths of the biggest contigs until 50% of the total contig length is reached. According to the comparison results in Table 4, the indexes varied after combining the duplicates, but just to a small extent.

Table 7 Comparison of secondary metabolite gene clusters identified

Secondary Metabolite	MZ3	MZ8	MZ3&8 Combined	MZ4	MZ9	MZ4&9 Combined
Terpene	1	1	1	1	1	1
Phenazine	1	1	1	0	0	0
Butyrolactone	1	1	1	1	1	1
Bacteriocin	2	2	2	2	2	2
Ectoine	1	1	1	1	1	1
Amglyccycl	0	0	0	1	1	1
NRPS	11	11	11	11	11	11
T1PKS	2	2	2	2	2	2
Unknown	4	3	2	3	3	2
NRPS-Terpene	1	1	1	1	1	1
Lantipeptide	1	0	0	0	0	1

In addition, the combined sequences were analyzed with antiSMASH in order to verify if more or new secondary metabolite gene cluster could be detected with twice higher coverage. According to Table 4, gene clusters of lantipeptide were detected at the coverage of 40x for both *Rhodococcus* strains. The number of unknown gene clusters, which have less than five domains so cannot be denoted as known secondary metabolites, decreased with twice higher coverage. However, the numbers of all the other secondary metabolites gene clusters remained the same even at twice higher coverage.

CHAPTER 4 DISCUSSION

Of the 81 bacterial isolates studied, 51 were identified as Firmicutes, 30 bacterial strains were identified as Actinobacteria with 29 strains identified as *Rhodococcus* and only one strain identified as *Corynebacterium*.

Table 8 Comparison of marine bacteria identified from other marine sediment

Studies	Sediment Origin	Bacteria identified
Biddle et al. 2005	Peru Margin	Gamma Proteobacteria, Cytophaga-flavobacterium-bacteroides
Hongxiang et al. 2008	Northeastern Pacific Ocean	Gamma Proteobacteria, Alpha Proteobacteria, Planctomycetacia, Delta Proteobacteria, Nitrospira, Actinobacteria, Beta Proteobacteria, Acidobacteria, Sphingobacteria, Firmicutes
Bhadra et al. 2008	Indian Ocean	*Brevibacterium ocean*, *Micrbacterium indicum*
da Silva et al. 2013	South Atlantic Ocean	Gamma Proteobacteria, Firmicutes, Actinobacteria
Schauer et al. 2009)	South Atlantic Ocean	Gamma Proteobacteria, Deltaproteobacteria, Actinobacteria
(Jamieson et al. 2013	Crozet Islands, Southern Ocean	Gamma Proteobacteria, Alpha Proteobacteria

Gamma Proteobacteria is the common and sometimes dominant phylum in bacterial communities from marine sediment. In addition, a recent study demonstrated that three Gram- negative bacteria belonging to the genus *Luteimonas* were isolated from sediment collected during EXP 329 in the South Pacific Gyre (Fan et al. 2013). However, both Gamma Proteobacteria and *Luteimonas* were not identified in the isolates in this study. In addition, the three genera identified seem to imply that the diversity of cultivatable bacteria in this deep-sea subseafloor

sedimentary environment is low compared to those isolated from marine sediment in other locations.

It is well known that only a very small minority of bacteria living in marine sediment can currently be isolated in culture. The limitations growth media in agar plates to isolate bacteria was apparent to Morita and ZoBell (Morita & ZoBell 1955) when they isolated organisms from the Philippine Trench. They argued that one recipe for media may support the growth of a subset of the community while another recipe could support another subset. Novel efforts to bring more marine bacteria into pure culture show great promise. Jesen et al. (1996)were able to increase the number of bacteria isolated from the surfaces of marine algae by using low nutrient agar. Zengler et al. (2002)used a combination of microencapsulation of individual cells in agar beads with low nutrient agar to bring marine bacteria into culture in a massively parallel approach. It is clear that including these methods, as well additional innovation in needed to increase the number of marine microbes in pure culture.

It must be noted that even though we strive to increase the diversity of bacterial isolated from marine sediment we still cannot come close to determining either the total biomass or the community diversity via culturing. Based on both microscopic evidence and environmental DNA studies, is very clear that the vast majority of microbes in the marine environment resist our attempts to bring them into isolated cultures. Staley and Konopka (1985) brought used the term “the great plate count anomaly” to describe the difference in orders of magnitude between the

numbers of cells from natural environments that form colonies on agar media and the numbers countable by microscopic examination (Jannasch & Jones 1959). Although these observations apply to all parts of the ocean, the cultivation conditions in the laboratory are even more different from the deep sea sediment environment which is salty, cold and under high pressured. The rapid environmental change may force some of the bacteria to activate different mechanism to protect themselves, such as spore formation. Although *Corynebacterium* and *Rhodococcus* do not produce endospores, *Bacillus* can produce endospores that can stay dormant for extended periods under stressful environment. Different culture media can also produce different colonies on plates selectively. A total of 152 bacteria were isolated in pure culture from the seven SPG sites (Forschner-Dancause 2012), while only 81 bacteria from five out of the seven SPG sites were sequenced in this study. It is possible that some bacteria genera are not sequenced by chance and that the bacteria from the other two sites (U1365 and 1369) have different bacteria genera, thus failed to be identified. Despite the known underestimation of biomass and diversity, the study of marine microbes in pure culture is still very important in this current era of gemonics (Rappé 2013), particularly for drug discovery studies.

16S rDNA phylogeny analysis revealed two *Rhodococcus* OTUs (OTU7 and OTU8, each has only one *Rhodococcus* strain) with less than 97% similarities with known sequences/published sequences (refer to Phylogenetic tree). Unfortunately, none of them was selected for whole genome sequencing since we hadn't got the

phylogeny analysis results by the time we determined to do whole genome sequencing. However, the presence of two unknown *Rhodococcus* strains may be new *Rhodococcus* species, which will reinforces the need for new studies in SPG deep-sea sediment.

Higher coverage revealed the identification of lantipeptide gene clusters, though the assembly results didn't seem to get improved. This triggers the thoughts that whether coverage affects the gene clusters identified and how to determine the priority coverage for the best gene clusters identification results.

Since Tao's literature also focused on *Rhodococcus erythropolis*, a comparison was made in Table 9. A similar number (23) of secondary metabolites gene clusters were found in the *Rhodococcus* (Tao et al. 2011). Four NRPS domain patterns (Pattern 1, 2, 3 and 4) found from the 11 *Rhodococcus* isolates in this study were also present in the *Rhodococcus* genome from Tao's literature (Cluster 9, 15, 13 and 19 respectively). NRPS domain pattern 5 was absent in Tao's literature. Whereas, two different NRPS domain patterns were present in Tao's *Rhodococcus erythropolis* genome. Due to the lack of more published genomic data of *Rhodococcus erythropolis*, no more comparison can be made. However, the comparison reveals that one NRPS domain pattern – Pattern 5 – is unique to this study.

Table 9 antiSMASH result of *Rhodococcus erythropolis*

	Tao 2011
Cluster 1	Butyrolactone
Cluster 2	Amglyccycl
Cluster 3	Lantipeptide
Cluster 4	Bacteriocin
Cluster 5	T1PKS
Cluster 6	T1PKS
Cluster 7	Bacteriocin
Cluster 8	T1PKS
Cluster 9	NRPS
Cluster 10	Bacteriocin
Cluster 11	NRPS
Cluster 12	Other
Cluster 13	NRPS
Cluster 14	NRPS-Terpene
Cluster 15	NRPS
Cluster 16	Bacteriocin
Cluster 17	NRPS
Cluster 18	NRPS
Cluster 19	NRPS
Cluster 20	Terpene
Cluster 21	Ectoine
Cluster 22	Bacteriocin
Cluster 23	Bacteriocin

CHAPTER 5 CONCLUSION

Three genera of bacteria were identified in this study: *Bacillus*, *Rhodococcus* and *Corynebacterium*. Of all the 81 16S rDNA sequenced bacterial isolates, 51 isolates belonging to *Bacillus*, 29 isolates belonged to *Rhodococcus* and one isolate belonging to *Corynebacterium*.

The phylogenetic results further indicate that: (1) all the *Rhodococcus* isolates belonged to *Rhodococcus erythropolis*; (2) 80 bacterial isolates were grouped into eight OTUs, while the *Corynebacterium* strain (SPG11-20) was not grouped. 51 *Baciilus* strains were grouped to 2 OTUs, and 29 *Rhodococcus* strains were grouped to 8 OTUs. 2 *Rhodococcus* OTUs (two *Rhodococcus* strains) showed similarities less than 97% with known/published sequences; (3) 51 *Bacillus* isolates distributed throughout three SPG sites (U1366F, U1368D and U1370F). 29 *Rhodococcus* isolates distributed throughout four SPG sites (U1366F, U1367D, U1370F and U1371F). No bacteria isolates were sequenced from sites U1365 and U1369 in this study by chance. Although the sedimentary environment varied from site to site, the OTUs of both *Bacillus* and *Rhodococcus* seemed to be unaffected. *Bacillus* strains' depth profile revealed that *Bacillus* strains distributed throughout the sediment cores from shallow section to deeper section. However, *Rhodococcus* strains' depth profile revealed that *Rhodococcus* strains tended to distribute in the shallower sections of the sediment cores.

Further study of sequencing the whole genomes of 11 selected *Rhodococcus* isolates throughout three sites (U1366F, U1367D and U1371F) showed that the G+C

content of the genomes was constant with a value of 62%. By annotating the genomes to look for the secondary metabolites gene clusters using antiSMASH, the results further revealed that various secondary metabolites gene clusters were found in the *Rhodococcus* genomes, dominated by NRPS and PKS. Five major domain patterns of NRPS in total were found in all the genomes, while Pattern 2 and Pattern 4 were found in all the 11 *Rhodococcus* genomes. The various secondary metabolites gene clusters, as well as the two potentially new *Rhodococcus* species suggests that SPG deep-sea sediments harbor bacteria with the potential to produce pharmaceutically important secondary metabolites, and more studies in this environment might provide more confidence for scientist to explore new drugs from this environment.

BIBLIOGRAPHY

Abdel-Aziz H, Wadie W, Abdallah DM, Lentzen G & Khayyal MT (2013) Novel effects of ectoine, a bacteria-derived natural tetrahydropyrimidine, in experimental colitis. *Phytomedicine* 20: 585–591.

Ashraf W, Mihdhir A & Colin Murrell J (1994) Bacterial oxidation of propane. *FEMS Microbiol. Lett.* 122: 1–6.

Bhadra B, Raghukumar C, Pindi PK & Shivaji S (2008) Brevibacterium oceani sp. nov., isolated from deep-sea sediment of the Chagos Trench, Indian Ocean. *Int. J. Syst. Evol. Microbiol.* 58: 57–60.

Biddle JF, House C & Brenchley JE Enrichment and Cultivation of Microorganisms from Sediment from the Slope of the Peru Trench (ODP Site 1230). *Proc. ODP, Sci. Results, 201*. http://www-odp.tamu.edu/ publications/201_SR/107/107.htm.

Cane DE & Ikeda H (2012) Exploration and mining of the bacterial terpenome. *Acc. Chem. Res.* 45: 463–472.

Christodoulou J et al. (1994) Barth syndrome: Clinical observations and genetic linkage studies. *Am. J. Med. Genet.* 50: 255–264.

Clardy J, Fischbach MA & Walsh CT (2006) New antibiotics from bacterial natural products. *Nat. Biotechnol.* 24: 1541–1550.

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

D'Hondt S et al. (2009) Subseafloor sedimentary life in the South Pacific Gyre. *Proc. Natl. Acad. Sci.* 106: 11651–11656.

Expedition 329 Scientists (2011) South Pacific Gyre subseafloor life. IODP Prel. Rept., 329.

Fan X, Yu T, Li Z & Zhang X-H (2013) Luteimonas abyssi sp. nov., isolated from deep-sea sediment of South Pacific Gyre. *Int. J. Syst. Evol. Microbiol.*

Farkas-Himsley H (1980) Bacteriocins--are they broad-spectrum antibiotics? *J. Antimicrob. Chemother.* 6: 424–426.

Finnerty WR (1992) The Biology and Genetics of the Genus Rhodococcus. *Annu. Rev. Microbiol.* 46: 193–218.

Fischbach MA & Walsh CT (2009) Antibiotics for Emerging Pathogens. *Science* 325: 1089–1093.

Forschner-Dancause SR (2012) Exploration of deep sea subsurface sediments for microbial diversity and biomedical potential. Ph.D., University of Rhode Island.

Gray KA, Pogrebinsky OS, Mrachko GT, Xi L, Monticello DJ & Squires CH (1996) Molecular mechanisms of biocatalytic desulfurization of fossil fuels. *Nat. Biotechnol.* 14: 1705–1709.

Hongxiang X, Min W, Xiaogu W, Junyi Y & Chunsheng W (2008) Bacterial diversity in deep-sea sediment from northeastern Pacific Ocean. *Acta Ecol. Sin.* 28: 479–485.

Jamieson RE, Heywood JL, Rogers AD, Billett DSM & Pearce DA (2013) Bacterial biodiversity in deep-sea sediments from two regions of contrasting surface water productivity near the Crozet Islands, Southern Ocean. *Deep Sea Res. Part Oceanogr. Res. Pap.* 75: 67–77.

Jannasch HW & Jones GE (1959) Bacterial populations in sea water as determined by different methods of enumeration. *Limnol. Oceanogr.* 4: 128–139.

Jenke-Kodama H & Dittmann E (2005) Combinatorial polyketide biosynthesis at higher stage. *Mol. Syst. Biol.* 1. http://www.nature.com/msb/journal/v1/n1/full/msb4100033.html (Accessed December 31, 2013).

Jensen PR, Kauffman CA & Fenical W (1996) High recovery of culturable bacteria from the surfaces of marine algae. *Mar. Biol.* 126: 1–7.

Kallmeyer J, Pockalny R, Adhikari RR, Smith DC & D'Hondt S (2012) Global distribution of microbial abundance and biomass in subseafloor sediment. *Proc. Natl. Acad. Sci.* 201203849.

Khosla C, Gokhale RS, Jacobsen JR & Cane DE (1999) Tolerance and specificity of polyketide synthases. *Annu. Rev. Biochem.* 68: 219–253.

Kreit J, Lefebvre G & Germain P (1994) Membrane-bound cholesterol oxidase from Rhodococcus sp. cells. Production and extraction. *J. Biotechnol.* 33: 271–282.

Mardis ER (2008) Next-Generation DNA Sequencing Methods. *Annu. Rev. Genomics Hum. Genet.* 9: 387–402.

McDonald M, Mavrodi DV, Thomashow LS & Floss HG (2001) Phenazine biosynthesis in Pseudomonas fluorescens: branchpoint from the primary shikimate biosynthetic pathway and role of phenazine-1,6-dicarboxylic acid. *J. Am. Chem. Soc.* 123: 9459–9460.

Medema MH, Blin K, Cimermancic P, Jager V de, Zakrzewski P, Fischbach MA, Weber T, Takano E & Breitling R (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* gkr466.

Morita RY & ZoBell CE (1955) Occurrence of bacteria in pelagic sediments collected during the mid-Pacific expedition. *Deep Sea Res. 1953* 3: 66–73.

Nikolouli K & Mossialos D (2012) Bioactive compounds synthesized by non-ribosomal peptide synthetases and type-I polyketide synthases discovered through genome-mining and metagenomics. *Biotechnol. Lett.* 34: 1393–1403.

Parkes RJ, Cragg BA & Wellsbury P (2000) Recent studies on bacterial populations and processes in subseafloor sediments: A review. *Hydrogeol. J.* 8: 11–28.

Peters J, Minuth T & Kula M-R (1993) A novel NADH-dependent carbonyl reductase with an extremely broad substrate range from Candida parapsilosis: Purification and characterization. *Enzyme Microb. Technol.* 15: 950–958.

Rappé MS (2013) Stabilizing the foundation of the house that 'omics builds: the evolving value of cultured isolates to marine microbiology. *Curr. Opin. Microbiol.* 16: 618–624.

Schauer R, Bienhold C, Ramette A & Harder J (2009) Bacterial diversity and biogeography in deep-sea surface sediments of the South Atlantic Ocean. *ISME J.* 4: 159–170.

Da Silva MAC et al. (2013) Phylogenetic identification of marine bacteria isolated from deep-sea sediments of the eastern South Atlantic Ocean. *SpringerPlus* 2. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3616218/ (Accessed December 31, 2013).

Staley JT & Konopka A (1985) Measurement of in Situ Activities of Nonphotosynthetic Microorganisms in Aquatic and Terrestrial Habitats. *Annu. Rev. Microbiol.* 39: 321–346.

Takigawa H, Kubota H, Sonohara H, Okuda M, Tanaka S, Fujikura Y & Ito S (1993) Novel Allylic Oxidation of α-Cedrene to sec-Cedrenol by a Rhodococcus Strain. *Appl. Environ. Microbiol.* 59: 1336–1341.

Tao F et al. (2011) Genome Sequence of Rhodococcus erythropolis XP, a Biodesulfurizing Bacterium with Industrial Potential. *J. Bacteriol.* 193: 6422–6423.

Turner JM & Messenger AJ (1986) Occurrence, biochemistry and physiology of phenazine pigment production. *Adv. Microb. Physiol.* 27: 211–275.

Warhurst AM, Clarke KF, Hill RA, Holt RA & Fewson CA (1994) Metabolism of styrene by Rhodococcus rhodochrous NCIMB 13259. *Appl. Environ. Microbiol.* 60: 1137–1145.

Whitman WB, Coleman DC & Wiebe WJ (1998) Prokaryotes: The unseen majority. *Proc. Natl. Acad. Sci.* 95: 6578–6583.

Zengler K, Toledo G, Rappé M, Elkins J, Mathur EJ, Short JM & Keller M (2002) Cultivating the uncultured. *Proc. Natl. Acad. Sci.* 99: 15681–15686.

APPENDICES

Appendix A *Rhodococcus* strains identified in each OTU

	Bacterial Isolates	Site	Top Depth CSF-A (m)*
OTU3 *Rhodococcus*	SPG11-6 = MZ1	1366F-2H-6-30	13.30
	SG11-10 = MZ2	1371F-1H-2-120	2.70
	SPG11-21 =MZ3&8	1366F-4H-1-50	24.00
	SPG11-22	1366F-4H-1-50	24.00
	SPG11-23 = MZ4&9	1366F-4H-1-50	24.00
	SPG11-48	1366F-3H-1-20	14.20
	SPG11-52	1370F-2H-1-110	7.80
	SPG11-53 = MZ5	1366F-3H-1-20	14.20
	SPG11-71 = MZ11	1367D-2H-6-30	15.20
	SPG11-72	1367D-2H-5-20	13.60
	SPG11-73	1367D-3H-1-90	17.80
	SPG11-74	1367D-3H-1-90	17.80
	SPG11-76	1367D-2H-1-40	7.80
	SPG11-77	1367D-2H-1-40	7.80
	SPG11-78 = MZ6	1367D-2H-2-90	9.80
	SPG11-79	1367D-2H-2-90	9.80
	SPG11-81	1367D-2H-2-90	9.80
	SPG11-83	1367D-2H-1-40	7.80
	SPG11-84	1367D-3H-1-90	17.80
	SPG11-86	1367D-2H-1-40	7.80
OTU4 *Rhodococcus*	SPG11-38	1371F-1H-2-120	2.70
	SPG11-39	1366F-3H-3-50	17.50
	SPG11-65	1371F-1H-2-120	2.70
OTU5 *Rhodococcus*	SPG11-68	1367D-2H-5-20	13.60
	SPG11-69	1367D-2H-5-20	13.60
OTU6 *Rhodococcus*	SPG11-70	1367D-2H-6-30	15.20
	SPG11-80	1367D-2H-2-90	9.80
OTU7 *Rhodococcus*	SPG11-29	1366F-3H-3-50	17.50
OTU8 *Rhodococcus*	SPG11-51	1370F-2H-1-110	7.80

* data from LIMS report - data missing MZ: whole genome sequencing

Appendix B *Bacillus* strains identified in each OTU

	Bacterial Isolates	Site	Top Depth CSF-A (m)*
OTU1 *Bacillus*	SPG11-1	1370F-6H-5-30	51.00
	SPG11-2	1370F-6H-5-30	51.00
	SPG11-3	1366F-1H-8-80	-
	SPG11-4	1366F-1H-8-80	-
	SPG11-5	1370F-6H-3-90	48.60
	SPG11-7	1366F-1H-3-80	-
	SPG11-8	1366F-1H-3-80	-
	SPG11-9	1366F-1H-3-80	-
	SPG11-11	1370F-2H-3-90	10.60
	SPG11-12	1370F-2H-3-90	10.60
	SPG11-13	1366F-4H-5-87	29.87
	SPG11-14	1368D-1H-1-20	0.20
	SPG11-15	1368D-1H-1-20	0.20
	SPG11-18	1368D-1H-1-20	0.20
	SPG11-19	1368D-1H-1-20	0.20
	SPG11-24	1366F-4H-5-87	29.78
	SPG11-25	1366F-4H-5-87	29.78
	SPG11-26	1370F-8H-1-90	64.60
	SPG11-27	1370F-7H-6-130	63.00
	SPG11-28	1366F-3H-6-20	21.70
	SPG11-34	1370F-8H-1-90	64.60
	SPG11-35	1366F-1H-1-140	1.40
	SPG11-37	1368D-1H-3-70	3.70
	SPG11-42	1370F-5H-5-70	41.90
	SPG11-43	1370F-5H-5-70	41.90
	SPG11-45	1368D-2H-2-70	9.70
	SPG11-46	1366F-3H-1-20	14.20
	SPG11-49	1370F-7H-2-100	56.70
	SPG11-50	1370F-2H-1-110	7.80
	SPG11-55	1366F-3H-1-20	14.20
	SPG11-57	1370F-7H-3-40	57.60
	SPG11-59	1370F-7H-3-40	57.60
	SPG11-63	1370F-1H-4-50	5.00
	SPG11-64	1370F-1H-4-50	5.00
	SPG11-88	1370F-6H-5-30	51.00
	SPG11-90	1370F-6H-3-90	48.60
	SPG11-93	1366F-4H-5-87	29.78
	SPG11-94	1368D-1H-1-20	0.20
OTU2 *Bacillus*	SPG11-33	1366F-2H-1-50	6.00
	SPG11-36	1366F-1H-1-140	1.40
	SPG11-40	1370F-4H-5-40	32.10
	SPG11-41	1370F-5H-5-70	41.90
	SPG11-44	1370F-7H-4-80	59.50
	SPG11-58	1370F-7H-3-40	57.60
	SPG11-60	1370F-7H-3-40	57.60
	SPG11-61	1370F-1H-4-50	5.00
	SPG11-62	1370F-1H-4-50	5.00
	SPG11-91	1370F-6H-5-30	51.00
	SPG11-92	1366F-3H-6-20	21.70
	SPG11-96	1366D-3H-3-50	-
	SPG11-97	1368D-1H-3-70	3.70

* data from LIMS report - data missing MZ: whole genome sequencing