UNIVERSITY OF CALIFORNIA, SAN DIEGO

Phylogenetic Diversity of Gram-positive Bacteria and Their Secondary Metabolite Genes

A Dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Oceanography

by

Erin Ann Gontang

Committee in charge:

William Fenical, Chair
Douglas H. Bartlett
Bianca Brahamsha
William Gerwick
Paul R. Jensen
Kit Pogliano

2008

The Dissertation of Erin Ann Gontang is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2008

DEDICATION

To John R. Taylor, my incredible partner, my best friend and my love.

*****

To my mom, Janet M. Gontang, and my dad, Austin J. Gontang. Your generous support and unconditional love has allowed me to create my future. Thank you.

*****

To my sister, Allison C. Gontang, who is as proud of me as I am of her. You are a constant source of inspiration and I am so fortunate to have you in my life.

TABLE OF CONTENTS

Published Papers Appended

Oh, D.-C., E. A. Gontang, C. A., Kauffman, P. R. Jensen and W. Fenical. 2008. Salinipyrones and pacificanones, mixed-precursor polyketides from the marine actinomycete, "*Salinispora pacifica*." Journal of Natural Products. 71(4):570-575.

Gontang, E. A., W. Fenical and P. R. Jensen. 2007. Phylogenetic diversity of Gram-positive bacteria cultured from marine sediments. Applied and Environmental Microbiology. 73(10):3272-3282.

Jensen, P. R., E. Gontang, C. Mafnas, T. J. Mincer and W. Fenical. 2005. Culturable marine actinomycete diversity from tropical Pacific Ocean sediments. Environmental Microbiology. 7(7):1039-1048.

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Chapter one, in full, is a reprint of the material as it appears in Applied and Environmental Microbiology, 2007, Gontang, Erin A.; Fenical, William; Jensen, Paul R. The dissertation author was the primary investigator and author of this paper.

Chapter two, in part, will be submitted for publication in PLoS biology, 2008, Penn, Kevin.; Jenkins, Caroline.; Gontang, Erin A.; Jensen, Paul R. The dissertation author will contribute this portion of the dissertation to the publication.

Chapter three, in full, will be submitted for publication as it appears in Applied and Environmental Microbiology, 2008, Gontang, Erin A.; Fenical, William; Jensen, Paul R. The dissertation author was the primary investigator and author of this paper.

I would like to acknowledge the financial support of the Scripps Institution of Oceanography graduate department, the Scripps Environmental Advocates, SeaGrant, the National Institutes of Health's Training Program in Marine Biotechnology, the Claude E. ZoBell Microbiology Fellowship, the Whole Earth Society, the David and Barbara Groce Foundation, and Bill and Gail Zimmerman.

Field collections in Palau were facilitated and supported by Pat and Lori Colin, Emilio Basilius and Matthew Mesubed of the Coral Reef Research Foundation (CRRF). Collection assistance was provided by Chris Kauffman, Wendy Strangman, Koty Sharp and Catherine Sincich. I would also like to acknowledge the invaluable discussions and scientific interactions with Alejandra Prieto-Davó, Sara Kelly, Kelle Freel, Kevin Penn, Lauge Farnaes, Caroline Jenkins, Anna Edlund, Katherine

Maloney, Alessandra Eustaquio, Tracy Mincer, Dong-Chan Oh, Ana-Paula Espindola, Hak Cheol Kwon, Glenroy Dean Martin, Melany Puglisi-Weening, Michelle Moffitt, Dan Udwary, Tona Mendoza-Ponce, Andy Thompson, Juanita Thompson, Rebecca Fenwick, Katie Gagnon, John Holecek, Sarah Glaser, Melissa Soldevilla, Chrisy Mafnas, Sarah White, Lisa Zeigler, Adam Jones, Carla Sorrels, Heather Reed, Jennifer Williams and Nick Haring; colleagues with whom I have grown, who love to learn and who continue to challenge me.

I would also like to acknowledge the mentorship of my scientific career by Jim Hoeffler, John Heyman, Gordon Brown, Lyle Turner, Jere Lipps, Steve Haddock, Kevin Raskoff, Tom Evans, Tom Crowley, David Johnson and Frank White.

I would like to acknowledge the tremendous support and advice of Bill Fenical and the mentorship and guidance of Paul Jensen. I thank the members of my committee, Doug Bartlett, Bianca Brahamsha, Bill Gerwick and Kit Pogliano for their comments, years of encouragement and willingness to be engaged in my graduate career. I would also like to thank the past and present members of the SIO graduate department, especially Becky Burrola, Denise Darling, Ryan Eddings, Josh Reeves, Dawn Huffman and Alice Zheng, and members of the marine biology research division, especially Beth Masek, Pat Fisher, Nora Adderson and Kyle Nakanishi.

Last but definitely not least, I wish to thank my mom, my dad, Allison and John. These are the four people to whom I dedicate my research. They have cheered me on throughout my graduate career and have supported me unconditionally throughout my journey. Thank you and I love you lots!

VITA

May, 2000	B.A. in Molecular and Cellular Biology University of California, Berkeley
2002-2008	Graduate Student Researcher Scripps Institution of Oceanography University of California, San Diego, CA
May, 2008	Ph.D. in Oceanography Scripps Institution of Oceanography University of California, San Diego, CA

PUBLICATIONS

Oh, D.-C., E. A. Gontang, C. A., Kauffman, P. R. Jensen and W. Fenical. 2008. Salinipyrones and pacificanones, mixed-precursor polyketides from the marine actinomycete, "*Salinispora pacifica*." Journal of Natural Products. 71(4):570-575.

Gontang, E. A., W. Fenical and P. R. Jensen. 2007. Phylogenetic diversity of Gram-positive bacteria cultured from marine sediments. Applied and Environmental Microbiology. 73:3272-3282.

Jensen, P. R., E. Gontang, C. Mafnas, T. J. Mincer and W. Fenical. 2005. Culturable marine actinomycete diversity from tropical Pacific Ocean sediments. Environmental Microbiology. 7(7):1039-1048.

Heyman., J. A., Cornthwaite, J., Foncerrada, L., Gilmore, J., Gontang, E., et al. 1999. Genome-scale cloning and expression of individual open reading frames using topoisomerase I-mediated ligation. Genome Research. 9:383-392.

FIELDS OF STUDY

Major Field: Marine Microbiology (Molecular Biology)

Studies in Marine Natural Products Drug Discovery
Professor William Fenical
Director, Center for Marine Biotechnology and Biomedicine
Scripps Institution of Oceanography

Studies in Marine Microbiology and Molecular Biology
Associate Professor Paul R. Jensen
Center for Marine Biotechnology and Biomedicine
Scripps Institution of Oceanography

ABSTRACT OF THE DISSERTATION

Phylogeny and Diversity of Gram-positive Bacteria

and Their Secondary Metabolite Genes

by

Erin Ann Gontang

Doctor of Philosophy in Oceanography

University of California, San Diego, 2008

Professor William Fenical, Chair

To assess the bacterial diversity of marine sediments, a culture-dependent approach was employed to assess the diversity of Gram-positive bacteria in marine sediments collected around the islands of Palau. The survey of the total, aerobic Gram-positive bacterial diversity present in tropical marine sediments resulted in the isolation of a diverse assemblage of Gram-positive isolates. Of the 78 operational taxonomic units cultured (OTUs based on ≥ 98% 16S rRNA gene sequence identity), 52 were determined to be members of the order *Actinomycetales* and 26 were determined to be members of the order *Bacillales*.

Bacterial genome sequencing has provided opportunities to assess the secondary metabolite biosynthetic potential of individual strains. By elucidating entire biosynthetic pathways in the context of all other genes and pathways present, it is also

possible to address questions related to the evolution of gene clusters or individual genes within a cluster. The complete genome sequence of *Salinispora arenicola* strain CNS-205, an actinomycete isolated during the research expedition to Palau, provided an opportunity to investigate the evolution of type I modular polyketide synthase (PKS) gene clusters.

While a tremendous amount can be learned from analyzing the entire genetic blueprint of a microorganism, additional natural product discovery approaches are necessary to gain insight into the biosynthetic potential of unsequenced organisms prior to fermentation, extraction and chemical analyses. A PCR based approach has been used successfully to suggest which representative isolates from the 52 *Actinomycetales* OTUs harbor type I PKS, enediyne PKS and nonribosomal peptide synthetase (NRPS) pathways. Over half of the cultured *Actinomycetales* OTUs were found to possess genes associated with at least one PKS or NRPS biosynthetic pathway and although some of these actinomycetes represent families well known to produce secondary metabolites, the results suggest that novel secondary metabolites can be isolated from both filamentous and unicellular actinomycete families. In addition to determining if these pathways are present, a phylogenetic approach was used successfully to suggest the number and novelty of type I PKS pathways present. By using the phylogenetic approach, the diversity, novelty and identity of type I PKS metabolites can be predicted.

Introductory Chapter.

Although the study of microorganisms that live in the marine environment has been shaped and influenced by the study of terrestrial microorganisms, some of the earliest observations of microorganisms were of marine bacteria (15). In 1677, Anton van Leeuwenhoek described his observations of bacteria in seawater using a "simple" microscope, a single lens instrument with a magnification estimated to be 275 diameters (20, 26). However, regardless of the early observations, an appreciation for the abundance, diversity and ecological roles of microorganisms, particularly marine microorganisms, would take almost two additional centuries. For example, it would be almost 150 years after van Leeuwenhoek's observations that Christian Gottfried Ehrenberg and others performed controlled, laboratory-based investigations of terrestrial and marine bacteria (15).

Even early pioneers in marine science, including Edward Forbes, did not recognize microorganisms as significant contributors to marine ecosystems. Although he made significant contributions during the early 1800's in the areas of marine biology and oceanography, Edward Forbes proposed the "azoic theory" suggesting that life did not exist below an ocean depth of approximately 550m (12). While Forbes' theory discounted organisms both large and small in the marine environment, his hypothesis peaked the curiosities of future marine scientists, including Thomas Henry Huxley. Huxley sought opportunities to examine deep-sea sediments in an effort to find evidence of life deep within the ocean and did so with the help of a medical officer aboard the HMS Cyclops. In 1857, while working with a crew conducting a series of soundings in preparation for the laying of cable, the medical officer aboard the HMS

Cyclops collected sediment samples trapped in the equipment brought up from the bottom of the North Atlantic (12). It was within these sediments that Huxley discovered coccolithophores, the microscopic single-celled organisms distinguished by their calcium carbonate plates known as coccoliths.

Around the same time that Huxley was discovering the coccolithophores, the influential work of microbiologists such as Louis Pasteur and Robert Koch was helping establish techniques and methods by which one could study microorganisms (15). However, even as terrestrial microorganisms began to be recognized as both ubiquitous and diverse, marine microorganisms continued to receive little attention. Regardless of the decades it would take before the study of marine microorganisms truly took off, studies of marine microbiology continued to be conducted. During the Travaillier and Talisman Expeditions of the 1880's, sediment and water samples collected from depths up to 5000m were examined and found by microscopy to contain bacteria (12). A decade later, techniques introduced by Koch were used to culture bacteria from oceanic water samples. Bernard Fischer used Koch's solidified bacteriological media to isolate colonies of bacteria as he crossed the Atlantic on a commercial liner (1).

Although the existence of marine microorganisms would continue to be recognized, their contribution to the marine ecosystem did not fully begin to be appreciated until both Claude E. ZoBell and Lawrence Pomeroy began to investigate how marine bacterioplankton, the bacterial component of the water column, survived in the marine environment (19, 27). The numbers of bacteria and other microorganisms in the ocean had been predicted to be very low and have little impact

on the environment or the cycling of nutrients. It was the pioneering work of marine microbiologists such as ZoBell and Pomeroy that finally drew significant attention to bacterial abundance in the ocean and the important roles those bacteria played in nutrient cycling in the marine environment. For example, in 1974 Pomeroy suggested that microbes are actually responsible for most of the metabolic activity in seawater. By using ultraclean analytical techniques, measuring radiolabeled precursor uptake and incorporation, and by assessing the ATP levels in seawater, Pomeroy described a new paradigm for the ocean's food web (19). With the advent of new and more sophisticated techniques, the roles of microorganism in the ocean continue to be explored and our understanding of the "microbial loop" continues to develop.

Today, marine bacterioplankton represent one of the most thoroughly studied environmental communities on the planet (9) and although there is still much to learn about the diverse microorganisms active in the water column, bacteria inhabiting marine sediments remain largely uncharacterized. In an effort to begin to assess the marine bacterial diversity of marine sediments and further our understanding of the fundamental differences between the bacterial populations inhabiting two major ocean ecosystems, chapter one of this dissertation discusses the culture-dependent approach taken to assess the diversity of Gram-positive bacteria in marine sediments collected around the islands of Palau. While early research estimated that only 5% of the bacteria in the ocean are Gram-positive (27), studies leading up to the present research suggested that the abundance and diversity of Gram-positive strains in sediments could be considerably greater (14, 21, 25). In total, 78 Gram-positive operational taxonomic units (OTUs) were cultured, of which 21 are considered to be new

phylotypes based on sharing < 98% 16S rRNA gene sequence identity with any previously cultured isolate for which sequence data is available. Eight other OTUs, previously observed but not yet formally described, bring the total number of potentially new taxa cultured as a part of the study to 29. The results strongly indicate that considerable Gram-positive diversity can be cultured from marine sediments and reinforces the concept that relatively simple cultivation techniques can be used successfully to isolate many as-yet-undescribed taxa (5, 13, 17).

The Gram-positive bacteria include bacteria from two major phylogenetic subdivisions, the phylum *Actinobacteria*, also described as the high G+C Gram-positives, and the phylum *Firmicutes*, also known as the low G+C Gram-positives, a group that includes well-known genera such as *Bacillus* and *Clostridium* (8). Aerobic Gram-positive bacteria, specifically actinomycetes (defined here as bacteria within the order *Actinomycetales*) and members of the order *Bacillales*, are well-known producers of important secondary metabolites including polyketides, nonribosomal peptides and combinations thereof (e.g. polyketide/nonribosomal peptide hybrids) (7, 11, 22). Many of these natural products possess biomedically relevant properties, such as antibiotic, antitumor, antifungal, and immunosuppressive activities. Over the past several decades, a dramatic drop in the rate of discovery of novel natural products has occurred due in large part to the frequent re-isolation of known metabolites from common terrestrial microorganisms. A rise in the number of drug-resistant pathogens coupled with the limited success of alternative strategies, such as combinatorial chemistry, to generate new natural products of biomedical relevance, has led to the need to find new approaches to natural products discovery.

With the advent of bacterial genome sequencing, researchers have been given the opportunity to assess the secondary metabolite biosynthetic potential of individual strains. For example, the genomic analysis of the well studied strain *Streptomyces coelicolor* strain A3(2) revealed clusters of genes characteristic of biosynthetic pathways responsible for the production of never before isolated small molecules (2). Prior to the sequencing of its genome, *Streptomyces coelicolor* had been studied for over forty years and was known to produce only three secondary metabolites. The genome sequence suggested however, that this strain had the capacity to produce at least 18 additional secondary metabolites. Genome sequencing offers insight into the biosynthetic potential of individual strains and can provide clear evidence for the genetic capacity to produce natural products that have yet to be discovered from that organism. By analyzing biosynthetic pathways in the context of all other genes and pathways present within a single organism, genome sequencing also makes it is possible to ask questions related to the evolution of those pathways or the individual genes within those pathways. The complete genome sequence of *Salinispora arenicola* strain CNS-205, an actinomycete isolated during the research expedition to Palau, offers a tremendous opportunity to investigate the evolution of secondary metabolite gene clusters. Chapter two details the results of an investigation into the type I modular PKS pathways of the *S. arenicola* genome. Investigations of these pathways in the context of the genome sequence and in relation to other sequenced genomes and gene clusters help highlight the genetic rearrangements that drive the evolution of secondary metabolite biosynthetic pathways. By analyzing the type I modular PKS

gene clusters of *S. arenicola*, evidence of gene duplication, gene loss, rearrangement, inversion and recombination were identified.

Genome sequencing to determine the secondary metabolite biosynthetic potential of individual bacterial strains remains cost prohibitive and time consuming. While a tremendous amount can be learned by analyzing the entire genetic blueprint of a microorganism, it remains important to develop additional natural product discovery approaches to gain insight into the biosynthetic potential of an organism prior to fermentation, extraction and chemical analysis.

Chapter three describes a PCR-based approach that can be used to quickly screen novel and diverse microorganisms in order to determine which strains have the genetic capacity to produce secondary metabolites and should be more thoroughly investigated for their chemical compound production. The PCR-based approach was used to screen strains from the 52 *Actinomycetales* OTUs recovered during the research expedition to the Republic of Palau, for the presence of domains associated with PKS pathways (including type I and enediyne PKS pathways) and NRPS pathways. In addition to screening OTUs from families well known for the production of secondary metabolites, the presence of genes associated with secondary metabolite biosynthesis was evaluated in lesser-studied bacteria, including unicellular actinomycetes. In addition to surveying *Actinomycetales* OTUs for domains associated with a variety of secondary metabolite biosynthetic pathways, a phylogenetic analysis of KS domain protein sequences from type I PKS pathways was also employed to evaluate the diversity and novelty of type I PKS pathways present among the cultured strains. The results suggest that over half of the cultured

Actinomycetales OTUs possess at least one secondary metabolite biosynthetic pathway and that by using a phylogentic approach, the diversity, novelty and identity of the secondary metabolites produced by a cultured microorganism can be predicted. Chapter three illustrates the power and utility of molecular tools and phylogenetic prediction, describing a method that can improve the process of secondary metabolite discovery.

In addition to assessing the capacity of actinomycetes to produce secondary metabolites, there is also a great interest in isolating additional strains from rare actinomycete groups known to produce secondary metabolites of biomedical relevance. In chapter four, efforts to identify ways to culture additional MAR2 actinomycetes are discussed. The MAR2 bacteria are generating interest among microbiologists and bacterial taxonomists as they may represent a new genus within the family *Streptomycetaceae*. MAR2 strains are also of interest to chemists as many have yielded an array of secondary metabolites, including compounds with antibiotic and anticancer properties (16). The sheer number and diversity of natural products produced by this one group of strains, in addition to their taxonomic novelty, has encouraged the search for additional MAR2 isolates. In an effort to increase the number of MAR2 strains available for study, a set of PCR primers specific for MAR2 strains was developed to screen strain libraries as well as environmental DNA extracted from marine sediments.

Because the world's oceans cover over 70% of the planet's surface and include some of the most diverse ecosystems on Earth, the marine environment has begun to be recognized as a significant resource of novel microorganisms, including

actinomycetes, with the capacity to produce natural products (3, 4, 6, 18). Culture-dependent investigations of marine sediments suggest that the marine environment is home to a diverse assemblage of microorganisms, many of which have never been detected in the terrestrial environment (10, 23, 24), and these cultured bacteria can now be screened to determine their genetic potential to produce secondary metabolites. The future of natural product drug discovery will undoubtedly continue to be influenced by advances in our ability to culture microorganisms, by increases in our understanding of the molecular genetics of secondary metabolite biosynthesis, and by technological progress in the areas of genome sequencing and other advanced molecular and genetic techniques.

References.

1. **Azam, F.** 2001. Introduction, history, and overview: the "methods" to our madness, p. 1-12. *In* J. H. Paul (ed.), Marine Microbiology, vol. 30. Academic Press, San Diego, C.A.

2. **Bentley, S. D., K. F. Chater, A.-M. Cerdeño-Tárraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C.-H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M.-A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill, and D. A. Hopwood.** 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature. 417:141-147.

3. **Bernan, V. S., M. Greenstein and G. T. Carter**. 2004. Mining marine microorganisms as a source of new antimicrobials and antifungals. Curr. Med. Chem. Anti.-Infect. Agents. 3(3):181-195

4. **Buchanan, G. O., P. G. Williams, R. H. Feling, C. A. Kauffman, P. R. Jensen, and W. Fenical.** 2005. Sporolides A and B: structurally unprecedented

halogenated macrolides from the marine actinomycete *Salinispora tropica*. Org. Lett. 7(13):2731-2734.

5. **Connon, S. A. and S. J. Giovannoni**. 2002. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. Appl. Environ. Microbiol. **68**:3878-3885.

6. **Feling, R. H., G. O. Buchanan, T. J. Mincer, C. A. Kauffman, P. R. Jensen, and W. Fenical.** 2003. Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora*. Angew. Chem. Int. Ed. 42:355-357.

7. **Fischbach, M. A. and C. T. Walsh.** 2006. Assembly-Line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem. Rev. 106:3468-3496.

8. **Garrity, G. M., M. Winters and D. B. Searles.** 2001. Taxonomic outline of the procaryotic genera, p.155-166. *In* G. M. Garrity, D. R. Boone and R. W. Castenholz (ed.), Bergey's manual of systematic bacteriology, 2nd edition, volume 1. Springer-Verlag, New York, N.Y.

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

10**. Gontang, E. A., W. Fenical, and P. R. Jensen.** 2007. Phylogenetic diversity of gram-positive bacteria cultured from marine sediments. Appl. Environ. Microbiol. 73: 3272-3282.

11. **Goodfellow, M., S. T. Williams and M. Mordarski**. 1983. Introduction to and importance of actinomycetes, p. 1-6. *In* M. Goodfellow, M. Mordarski and S. T. Williams (ed.), The biology of the actinomycetes. Academic Press Inc., London, England.

12. **Jannasch, H. W. and C. O. Wirsen.** 1983. Microbiology of the deep sea. *In*: G. T. Rowe (ed.), Deep-sea biology. John Wiley, New York, N.Y.

13. **Janssen, P. H., P. S. Yates, B. E. Grinton, P. M. Taylor and M. Sait**. 2002. Improved culturability of soil bacteria and isolation in pure culture of novel members of the divisions *Acidobacteria, Actinobacteria, Proteobacteria*, and *Verrucomicrobia*. Appl. Environ. Microbiol. **68**:2391-2396.

14. **Jensen, P. R., T. J. Mincer, P. G. Williams and W. Fenical**. 2005. Marine actinomycete diversity and natural product discovery. Antonie van Leeuwenhoek. **87**:43-48.

15. **Karl, D. M. and L. M. Proctor.** 2007. Foundation of Microbial Oceanography. Oceanography. 20(2):16-27.

16. **Kwon, H. C., C. A. Kauffman, P. R. Jensen and W. Fenical.** 2006. Marinomycins A-D, antitumor-antibiotics of a new structure class from a marine actinomycete of the recently discovered genus "*Marinispora*". J. Am. Chem. Soc. 128:1622-1632.

17. **Maldonado, L. A., J. E. M. Stach, W. Pathom-aree, A. C. Ward, A. T. Bull and M. Goodfellow**. 2005. Diversity of cultivable actinobacteria in geographically widespread marine sediments. Antonie van Leeuwenhoek. **87**:11-18.

18. **Oh, D.-C., P. G. Williams, C. A. Kauffman, P. R. Jensen, and W. Fenical.** 2006. Cyanosporosides A and B, unprecedented cyclopenta[a]indene glycosides from the obligate marine actinomycete *Salinispora pacifica*. Org. Lett. 8(6):1021-1024.

19. **Pomeroy, L. R.** 1974. The ocean's food web, a changing paradigm. Bio. Science. 24:499-504.

20. **Porter, J. R.** 1976. Antony van Leeuwenhoek: Tercentenary of his discovery of bacteria. Bacteriological Reviews 40:260–269.

21. **Priest, F. G**. 1989a. Isolation and identification of aerobic endospore-forming bacteria, p.27-56. *In* C. R. Harwood (ed.), *Bacillus*. Plenum Press, New York, N.Y.

22. **Priest, F. G.** 1989b. Products and applications, p. 293-320. *In* C. R. Harwood (ed.), *Bacillus*. Plenum Press, New York, N.Y.

23. **Ravenschlag, K., K. Sahm, J. Pernthaler, and R. Amann.** 1999. High bacterial diversity in permanently cold marine sediments. Appl. Environ. Microbiol. 65(9):3982-3989.

24. **Stach, J. E. M., L. A. Maldonado, D. G. Masson, A. C. Ward, M. Goodfellow, and A. T. Bull.** 2003. Statistical approaches for estimating actinobacterial diversity in marine sediments. Appl. Environ. Microbiol. 69(10):6189-6200.

25. **Stach, J. E. M. and A. T. Bull**. 2005. Estimating and comparing the diversity of marine actinobacteria. Antonie van Leeuwenhoek. **87**:3-9.

26. **van Leeuwenhoek, A.** 1677. Observations…concerning little animals observed in rain-, well-, sea- and snow-water, pp. 8–9. *In*: T. D. Brock (ed.), Milestones in Microbiology. Prentice-Hall, Inc., Englewood Cliffs, N.J.

27. **ZoBell, C. E.** 1946. Marine Microbiology: A Monograph on Hydrobacteriology. Chronica Botanica Company, Waltham, MA, 240 pp.

Chapter One: The Phylogenetic Diversity of Gram-Positive Bacteria Cultured from Marine Sediments

Abstract.

Major advances in our understanding of marine bacterial diversity have been gained through studies of bacterioplankton, the vast majority of which appear to be Gram-negative. Less effort has been devoted to studies of bacteria inhabiting marine sediments, yet there is evidence to suggest that Gram-positive bacteria comprise a relatively large proportion of these communities. To further expand our understanding of the aerobic Gram-positive bacteria present in tropical marine sediments, a culture-dependent approach was applied to sediments collected in the Republic of Palau at depths from intertidal-500 m. This investigation resulted in the isolation of 1,624 diverse Gram-positive bacteria spanning 22 families, including many that appear to represent new taxa. Phylogenetic analysis of 189 representative isolates, based on 16S ribosomal DNA (rDNA) sequence data, indicated that 124 (65.6%) belonged to the class *Actinobacteria* while the remaining 65 (34.4%) were members of the class *Bacilli*. Using a sequence identity value of ≥ 98%, the 189 isolates grouped into 78 operational taxonomic units (OTUs) of which 29 (37.2%) are likely to represent new taxa. The high degree of phylogenetic novelty observed during this study highlights the fact that a great deal remains to be learned about the diversity of Gram-positive bacteria in marine sediments.

Introduction.

Gram-positive bacteria can be divided into two major subdivisions, the phylum *Actinobacteria*, also described as the high G+C Gram-positives, and the phylum *Firmicutes*, also known as the low G+C Gram-positives, a group that includes well-known genera such as the *Bacillus* and *Clostridium* (21). Gram-positive bacteria typically have a cell wall consisting of a thick layer of peptidoglycan (19) while a few rather unusual genera lack a cell wall entirely (42). Many in this large group of primarily chemoorganotrophic bacteria are also known to produce spores in response to starvation or harsh chemical or physical conditions (18, 40, 50). Aerobic Gram-positive bacteria, specifically actinomycetes (defined here as bacteria within the order *Actinomycetales*) and members of the order *Bacillales*, are generally saprophytic and include well-known producers of important secondary metabolites (23, 53).

While the most thoroughly studied Gram-positive bacteria include human pathogens (e.g., *Mycobacterium tuberculosis*, *Bacillus anthracis*) and soil derived, antibiotic producing actinomycetes (2), relatively little is known about the diversity and distribution of Gram-positive bacteria in the marine environment. This lack of information persists despite the fact that Gram-positive bacteria have been cultured from the ocean for decades (5, 26, 31, 43, 68) and consistently appear in culture-independent studies (e.g., 62, 66), including the report of a new and as-yet-uncultured order within the class *Actinobacteria* (54). Gram-positive bacteria are likely to play important microbiological roles in the marine environment yet without a fundamental understanding of their diversity and ecophysiology, it is difficult to assess the

ecological significance of this relatively overlooked component of the marine bacterial community.

Although Gram-positive bacteria have been cultivated from seawater, marine invertebrates and other sample types (25, 27, 29, 47, 69), marine sediments (31, 33, 45, 48, 64), including deep-sea sediments (39, 56, 68), are the primary oceanic habitat from which they have been recovered (1). While probable that some marine-derived Gram-positive bacteria are terrigenous microorganisms, washed or blown into the marine environment, species occurring exclusively in the sea have been described (25, 26, 69). The recovery of Gram-positive bacteria that require seawater for growth, including several *Bacillus* species (24, 28, 56, 71) and the recently described actinomycete genus *Salinispora* (44), suggests that additional, obligate marine taxa reside in marine sediments.

Encouraged by recent work that clearly demonstrated how improved, selective cultivation methods are an effective means of isolating significant new bacterial diversity (36, 55, 57, 70), we performed a series of culture-dependent experiments designed to assess the diversity of Gram-positive bacteria in marine sediments. The present study revealed a diverse assemblage of bacteria spanning 22 Gram-positive families, including many that appear to represent new taxa.

Materials & Methods.

Sediment collection and bacterial isolation.

A total of 225 sediment samples were collected at depths ranging from intertidal-500 m during a research expedition to the Republic of Palau (7°30'N, 134°30'E), from

6-17 March 2004. Sediment samples were collected either by SCUBA divers or using a modified, surface deployed sediment sampler (Kahlsico, El Cajon, Calif. model #214WA110). Following collection, samples were placed into sterile 50-ml plastic Whirl-Pak bags (NASCO, Modesto, Calif.) and kept cool until processed (within 4 h) by one or more of the following four selective methods.

The first processing method involved drying 10 ml of wet sediment overnight in a laminar-flow hood before stamping onto agar media. The method was performed as described previously (33) with the exception that a polyester fiber tipped sterile swab (Fisher Scientific, Hampton, NH) was used to press the dried sediment onto the agar surface 35 to 40 times creating a serial dilution effect. The second processing method involved adding 0.5 g of sediment (dried overnight) to 4 ml of autoclaved, 0.2-μm filtered seawater (AFSW) either with or without kanamycin (5 μg/ml final concentration). After vigorous shaking for 30 s, the sediment was allowed to settle for five minutes before 50 μl were inoculated onto agar media and spread with an alcohol-sterilized glass rod. For the third processing technique, wet sediment was diluted (1:4) in AFSW and then heated for 6 min at 55^{o}C. The diluted sample was then vigorously shaken for 30 s, further diluted (1:4), and then 50 μl of each dilution were plated onto agar media. Finally, pour plates were prepared by adding 0.5 g of wet sediment to 25 ml of autoclaved, molten (~42^{o}C), 100% seawater agar amended with cycloheximide (100 μg/ml) and rifampicin (5 μg/ml).

Processed samples were inoculated onto one or more of 11 different isolation media (A1 to A11). All agar media were prepared with 0.2-μm filtered, deionized (DI) water and/or natural seawater and were amended with 0.2-μm filtered

cycloheximide (100 μg/ml) and a second antibiotic (if noted), after autoclaving. Isolation media consisted of the following: A1, 18 g agar, 10 g starch, 4 g yeast extract, 2 g peptone, 1 liter natural seawater and rifampicin (5 μg/ml); A2 (10% A1), 18 g agar, 1 g starch, 0.4 g yeast extract, 0.2 g peptone, 1 liter natural seawater; A3, 18 g agar, 2.5 g starch, 1 g yeast extract, 0.5 g peptone, 0.2 g glycerophosphate (disodium pentahydrate), 750 ml natural seawater and 250 ml DI water; A4 (100% seawater agar), 18 g agar and 1 liter natural seawater; A5 (75% seawater agar), 18 g agar, 750 ml natural seawater and 250 ml DI water; A6-A9, 18 g agar, 1 liter natural seawater and one of the following antibiotics respectively: polymixin B sulfate (5 μg/ml), kanamycin (5 μg/ml), novobiocin (25 μg/ml) or rifampicin (5 μg/ml); A10, 8 g noble agar, 0.5 g mannitol, 0.1 g peptone, 1 liter natural seawater and rifampicin (5 μg/ml); A11 [Munz media (49)], 18 g agar, 1 g KNO_3, 0.1 g $MgSO_4$-$7H_2O$, 2 g Na_2HPO_4-$7H_2O$, 0.14 g KH_2PO_4, 1 g NaCl, 1 liter DI water and 5 ml light, liquid paraffin (added after autoclaving).

Inoculated plates were incubated at 25-28^{o}C for up to 12 weeks and all well-separated bacterial colonies, observed by eye or using a stereomicroscope at up to 64X magnification (Leica Microscopy Systems Ltd., Heerbrugg, Switzerland), were removed from the original isolation plate and sub-cultured on A1. The Gram reaction of all pure cultures was determined via the nonstaining (KOH) method (6). The majority of the Gram-positive strains possessed morphological features characteristic of the recently described actinomycete genus *Salinispora* (44). Multiple strains from each *Salinispora*-like morphotype were cryopreserved at –80^{o}C along with all of the remaining Gram-positive strains. All strains were grouped according to colony color,

morphology and pigment production and representatives from each phenotype were subjected to phylogenetic analysis.

Nucleic acid extraction, 16S rRNA gene (16S rDNA) amplification and sequencing.

Genomic DNA was extracted according to the DNeasy protocol (Qiagen Inc., Valencia, Calif.) with the following modifications. After RNase A (2 mg/ml) was added to the enzymatic lysis buffer, the resuspended bacterial pellet was incubated for two hours at 37°C. Following the addition of proteinase K and buffer AL, the sample was held for one hour at 70°C. Genomic DNA was eluted from the spin column with 100 μl of buffer AE for immediate use or storage at –20°C.

The 16S rRNA genes were PCR amplified from genomic DNA using the primers FC27 (5'-AGAGTTTGATCCTGGCTCAG-3') and RC1492 (5'-TACGGCTACCTTGTTACGACTT-3'). The 50 μl PCR reaction mixture contained 20 to 50 ng of DNA, 250 pmol of each primer, ThermoPol Buffer (New England BioLabs Inc., Beverly, Mass.), 2.5 U of TaqDNA polymerase (New England BioLabs Inc., Beverly, Mass.), and 100 μM deoxynucleoside triphosphate mixture. The PCR program consisted of 30 cycles of 94°C for 1 min, 60°C for 1 min, and 72°C for 1 min, followed by a final extension step at 72°C for 7 minutes. Amplification products were examined by agarose gel electrophoresis and purified using Qiagen's QIAquick PCR cleanup kit according to the manufacturer's suggested protocol (Qiagen Inc., Valencia, Calif.). A partial consensus sequence (*E. coli* numbering 20-531) for each isolate was obtained using the primers FC27 and R530 (5'-CCGCGGCTGCTGGCACGTA-3'). Nearly complete sequences were obtained for select 16S rDNA amplicons (*E. coli*

numbering 20-1392) using four additional primers: RC1492, R936 (5'-GTGCGGGCCCCCGTCAATT-3'), F514 (5'-GTGCCAGCAGCCGCGGTAA-3'), and F1114 (5'-GCAACGAGCGCAACCC-3'). Sequencing reactions were carried out with an ABI 3100 DNA sequencer at the DNA Sequencing Shared Resource, UCSD Cancer Center (funded in part by the NCI Cancer Center Support Grant #2 P30CA23100-18).

Phylogenetic analyses and diversity estimates.

All nucleotide sequences were assembled, analyzed and manually edited using the Sequencher software package (version 4.5, Gene Codes Co., Ann Arbor, Mich.) and compared to sequences within the NCBI database (http://www.ncbi.nlm.nih.gov/) using the Basic Local Alignment Search Tool (BLAST). All partial 16S rRNA gene sequences sharing a phylogenetic affiliation with either the *Actinobacteria* or the *Firmicutes* were imported into ARB (41) and aligned. Aligned partial 16S rRNA gene sequences (*E. coli* numbering 20-531) were analyzed using the Clusterer program (http://www.bugaco.com/bioinf) and the number of operational taxonomic units (OTUs) calculated using sequence identity values ranging from ≥ 90% to 100%. For at least one representative of each OTU generated using the ≥ 98% sequence identity value, a nearly complete 16S rRNA gene sequence was obtained. Phylogenetic analyses were performed using PAUP (63) and trees drawn using distance neighbor joining and UPGMA methods and maximum parsimony.

In order to estimate the taxonomic novelty of the bacteria cultured, strains within OTUs sharing a sequence identity value of ≥ 98% were subjected to further analysis. An OTU was considered a new phylotype if all strains within the OTU shared < 98%

sequence identity with any previously cultured bacterium for which sequence data was available (as determined by a BLAST search); otherwise, the OTU was designated a known (previously cultured) phylotype. In addition to determining whether the members of each OTU had been previously cultured, an OTU's taxonomic novelty was assessed using the OTU's nearest type strain (http://www.bacterio.cict.fr/). If all isolates within an OTU shared < 98% sequence identity with the nearest type strain, as calculated using the ARB distance matrix, the OTU was considered to have a high probability of representing a new taxon. OTUs calculated using a sequence identity value of ≥ 98% were further used to estimate Gram-positive bacterial diversity using the abundance-based coverage estimator (ACE, 9) and Chao's richness estimator (8) implemented in EstimateS (version 7; R.K. Colwell; available at http://viceroy.eeb.uconn.edu/estimates).

Effects of seawater on growth.

Select isolates were screened to determine whether they required seawater for growth. Using a sterile loop, cells from a single colony were streaked onto A1 prepared with natural seawater and A1 prepared with DI water. Plates were incubated at 25-28^{o}C for six to eight weeks and growth was monitored at up to 64X magnification. Strains that grew on the medium prepared with seawater but not on the medium prepared with DI water were scored as requiring seawater for growth.

Nucleotide sequence accession numbers.

16S rRNA gene sequences have been deposited in the GenBank database (http://www.ncbi.nlm.nih.gov/Genbank/index.html) under the accession numbers DQ092624, DQ224159 and DQ448693-DQ448806.

Results.

From a total of 225 sediment samples, 1,624 Gram-positive bacteria were isolated. Interestingly, 1,353 (83.3%) of these strains possessed morphological features characteristic the genus *Salinispora* (44). Four hundred seven of the *Salinispora*-like strains, along with the remaining 271 Gram-positive strains, were cryopreserved at –80^{o}C. Of these 678 strains, 199 were chosen for phylogenetic analysis based on colony color and morphology. These 199 isolates included 25 *Salinispora*-like strains and 64.2% (174) of the remaining Gram-positive strains cultured. NCBI nucleotide BLAST searches using the partial 16S rRNA gene sequences of these 199 strains revealed that 189 (95.0%) of the isolates were Gram-positive and shared a phylogenetic affiliation with either the *Actinomycetales* or *Bacillales*. These results further validate the KOH method (6) as a rapid and effective means to determine the cell wall type of an isolate. Appendix A provides supplemental information on the 189 Gram-positive isolates, including collection depth, isolation method and medium, seawater requirement, and nearest type strain. Even though sediment-processing methods were not applied equally to all samples, stamping dried sediments onto low nutrient agar proved to be a highly successful method to cultivate Gram-positive bacteria. In fact, over 70% of the Gram-positive strains were cultured on low nutrient media, in particular media A4-A6.

A phylogenetic analysis of 25 of the 1,353 strains that shared morphological similarities with the genus *Salinispora* (44) revealed that 23 shared > 99% 16S rRNA gene sequence identity with members of this taxon. The other two strains belonged to

the closely related genus *Micromonospora*. Of the 23 *Salinispora* strains, 16 (69.9%) shared 100% sequence identity with *S. arenicola*, further supporting the pantropical distribution and lack of intraspecies 16S rRNA gene diversity within this taxon (34). None of the 23 strains claded with *S. tropica* which, to date, has only been reported from the Bahamas. The remaining seven (30.4%) strains belonged to a new phylotype for which the name "*S. pacifica*" has been proposed (34).

The diversity of Gram-positive bacteria cultured in this study was estimated by performing cluster analyses using the 189 partial 16S rRNA gene sequences. The number of operational taxonomic units (OTUs) calculated using various levels of sequence identity were as follows: ≥ 90% = 8 OTUs; ≥ 91% = 9 OTUs; ≥ 92% = 15 OTUs; ≥ 93% = 18 OTUs; ≥ 94% = 35 OTUs; ≥ 95% = 43 OTUs; ≥ 96% = 49 OTUs; ≥ 97% = 63 OTUs; ≥ 98% = 78 OTUs; ≥ 99% = 95 OTUs; 100% = 116 OTUs. Of the 116 distinct Gram-positive sequences identified, 70 (60.3%) were phylogenetically affiliated with the order *Actinomycetales* (Figure 1.1). These actinomycetes are most closely related to 25 different genera that fall within 18 separate family level groupings and span eight of the 10 sub-orders within the *Actinomycetales*.

The remaining 46 (39.7%) OTUs calculated using 100% sequence identity shared a phylogenetic affiliation with the order *Bacillales* (Figure 1.2). The majority of these OTUs (67.4%) formed a highly diverse clade, all of whose members are most closely related to the genus *Bacillus*. The remaining 15 OTUs were most closely related to the genera *Exiguobacterium*, *Halobacillus*, *Laceyella*, *Paenibacillus*, *Pontibacillus*, or *Staphylococcus*. Contrary to formal taxonomic assignment, the single *Staphylococcus* strain (CNJ-924) and the *Exiguobacterium* strains (CNJ-771 and CNJ-781) did not

clade with their respective families when partial 16S rRNA gene sequences were used (Figure 1.2). However, when using nearly complete 16S rRNA gene sequences, these relationships were rectified. CNJ-924 grouped with its appropriate family, the *Staphylococcaceae* (data not shown), and the *Exiguobacterium* strains, while deeply rooted, grouped with the *Bacillaceae* (Figure 1.3).

While 100% 16S rRNA gene sequence identity was used to highlight the 16S rRNA diversity of the cultured isolates, the 52 *Actinomycetales* OTUs (Table 1.1) and 26 *Bacillales* OTUs (Table 1.2) generated using the more conservative identity value of ≥ 98% were used to estimate the phylogenetic novelty of the culture collection. When all strains within these OTUs shared < 98% sequence identity with all previously cultured bacteria for which comparable sequence data was available, the OTU was considered a new phylotype. Using this criterion, 12 of the 52 *Actinomycetales* OTUs (23.1%) and nine of the 26 *Bacillales* OTUs (34.6%) represent new phylotypes (Table 1.3). Thus, 21 of the 78 observed phylotypes (26.9%) have not been previously cultured and reported using 16S sequence-based methods. Of the 40 known *Actinomycetales* OTUs and 17 known *Bacillales* OTUs, 35 and 10, respectively, had not been previously reported from marine sources. Given that the samples were collected close to shore, (within 10 km), these 45 strains represent bacteria that appear to be adapted to both marine and non-marine environments. Using the abundance-based coverage estimator (ACE) and Chao's richness estimator, it can be predicted that the numbers of cultivable OTUs (≥ 98% sequence identity) in these sediments are 131 and 130 respectively. Relative to the 78 OTUs detected, these

estimators suggest that further processing of the sediment samples would yield additional Gram-positive bacterial diversity.

The 21 OTUs identified as new (not previously cultured) phylotypes have a high probability of representing new taxa. Additionally, six previously observed *Actinomycetales* and two previously observed *Bacillales* OTUs may also represent new taxa as all strains within these OTUs shared < 98% 16S rRNA gene sequence identity with their nearest type strain. Thus in total, 29 of the 78 (37.2%) OTUs cultured as part of this study have the potential to be described as new taxa. The nearly full 16S rRNA gene sequences of representative isolates from each of these OTUs were used to construct a phylogenetic tree (Figure 1.3). Also included in this tree are representative isolates from the 12 OTUs that were not considered new but were most closely related to a type strain isolated from a marine source.

Of the 144 strains tested, 57 required seawater for growth while the remainder either grew poorly (24 strains) or equally well (63 strains) when seawater was replaced with deionized water in the growth medium. Forty-five of the 57 seawater-requiring strains were divided among 14 OTUs (≥ 98% sequence identity) that were comprised solely of seawater requiring strains. These strains either belonged to a new OTU or an OTU most closely related to a type strain isolated from a marine source (Figure 1.3). Ten additional seawater-requiring strains fell into seven previously observed OTUs that contained from one to five strains that did not require seawater for growth. The final two seawater requiring strains, each the single member of a separate OTU, belonged to known OTUs that had not been previously described as seawater requiring.

Thirty-three of the 57 seawater-requiring strains belonged to the *Actinomycetales*. In addition to strains related to the known seawater requiring genus *Salinispora*, seawater requiring actinomycetes were also most closely related to the genera *Dietzia*, *Kocuria*, *Kytococcus*, *Marmoricola*, *Microbacterium*, *Mycobacterium* and *Pseudonocardia*. Outside of the genus *Salinispora*, these strains are among the first seawater-requiring actinomycetes to be reported. Also requiring seawater were 24 strains within the class *Bacilli*. While the majority of these seawater-requiring strains were most closely related to *Bacillus* species, seawater-requiring strains related to the *Halobacillus*, *Laceyella* and *Paenibacillus* species were also cultivated.

Discussion.

Marine bacterioplankton represent one of the most thoroughly studied environmental communities on the planet (22), yet bacteria inhabiting marine sediments remain largely uncharacterized. This lack of information hinders an effective assessment of marine bacterial diversity and limits our understanding of the fundamental differences between the bacterial populations inhabiting two major ocean ecosystems. One apparent yet relatively unexplored difference between seawater and sediment bacterial communities is the relative abundance of Gram-positive bacteria. While early research estimated that only 5% of the bacteria in the ocean are Gram-positive (72), more recent studies suggest that the abundance and diversity of Gram-positive strains in sediments may be considerably greater (32, 52, 59). The present study employed cultivation-dependent methods to assess the diversity of Gram-positive bacteria in marine sediments collected around the islands of Palau. In total,

78 Gram-positive OTUs were cultured, of which 21 are considered to be new phylotypes based on sharing < 98% 16S rRNA gene sequence identity with any previously cultured isolate for which sequence data is available. Eight other OTUs, previously observed but not yet formally described, bring the total number of potentially new taxa cultured as a part of this study to 29. These results indicate that considerable Gram-positive diversity can be cultured from marine sediments and reinforces the concept that relatively simple cultivation techniques can be used successfully to isolate many as-yet-undescribed taxa (13, 35, 43).

The frequent use of high nutrient media in previous studies of bacterial diversity may explain why some Gram-positive bacteria have gone uncultured. During the present study, the majority of isolates were obtained using low nutrient media (e.g., seawater agar, see Appendix A). In fact, 24 of the 29 OTUs for which formal taxonomic descriptions are not yet available were isolated exclusively from low nutrient media. While all of the cultured strains were ultimately capable of growth on a high nutrient medium (e.g., A1), our observations support previous studies (13, 14, 60) which suggest that lower nutrient concentrations improve the initial isolation and recovery of diverse microorganisms as they help avoid contamination and overgrowth by fast growing strains.

The identification of 21 new Gram-positive phylotypes, despite extensive culture-independent investigations of seawater, might suggest that seawater and sediment communities are significantly different. The fact that the number of new phylotypes falls by three to 18 when the results of culture-independent analyses are included in the comparison (data not shown) supports this possibility. Alternatively, biases

associated with culture-independent methods (17, 61, 67) may have contributed to the underestimation of specific groups of Gram-positive bacteria that occur both in seawater and sediments. This may particularly apply in the case of spore-forming Gram-positive bacteria, as it is known that even when specific steps are taken to lyse spores, these bacteria are underrepresented in environmental clone libraries when spore counts are $\leq 10^3$/ml sediment (15, 46, 48). Although culture-dependent approaches also have well-known biases (38, 51, 65), these methods may prove to be the most effective way to detect certain groups of marine bacteria. In addition, cultured strains can be subjected to taxonomic characterization and their physiology, ecology and biotechnological potential explored.

While the number of OTUs was reported using multiple 16S rRNA gene sequence identity values, only those clusters generated using a value of ≥ 98% were subjected to additional diversity analyses. This value was chosen based on the relationship between percent DNA-DNA reassociation and 16S rRNA gene similarity, where 70% DNA relatedness is expected to correspond to > 98% 16S rRNA gene sequence identity (16). Although Stach et al. (58) suggested that a 16S rDNA sequence identity value of ≥ 99% could be used to define an OTU, that study was focused solely on delineating actinobacterial OTUs. The use of a sequence identity value of ≥ 98% may not provide the most conservative estimate of OTU numbers however, even at this value, it is probable that diversity will be underestimated.

Members of the actinomycete families *Micromonosporaceae, Nocardiaceae* and *Streptomycetaceae* have dominated previous studies of terrestrial and marine-derived *Actinobacteria* (11, 12, 26, 43), and isolates most closely related to members of each

of these three families were cultured during the present study. Based on morphological characterization, the majority of the isolates recovered were identified as *Micromonosporaceae*, supporting the hypothesis that these bacteria are among the dominant actinomycetes cultivable from marine sediments (30, 68). Also readily cultured from marine sediments were actinomycetes of the *Nocardiaceae* and *Streptomycetaceae* families. While surprised not to recover *Rhodococcus* isolates, which are among the most common members of the *Nocardiaceae* recovered from marine samples (11, 12, 26), our processing methods clearly did not select against other mycolate actinomycetes including strains most closely related to *Corynebacterium*, *Dietzia*, *Gordonia*, *Mycobacterium* and *Nocardia*. Within the *Streptomycetaceae*, a diverse assemblage of filamentous, spore-forming actinomycetes grouped into 15 OTUs. Five of those *Streptomycetaceae* OTUs shared < 98% 16S rRNA gene sequence identity with the most closely related type strain and thus considerable new taxonomic diversity appears to have been cultured within this well studied family.

The phylogenetic identification of what appear to be new taxa within the *Actinomycetales* and *Bacillales* confirmed previous observations that marine sediments harbor new diversity within these groups (11, 26, 43, 44). These two orders are responsible for almost 50% of the known bioactive microbial metabolites discovered to date, including many well-known antibiotics (2). Although marine microorganisms have only recently become a target for natural product drug discovery, it has become increasingly clear that Gram-positive strains are a rich source of new structures that possess promising antimicrobial and anti-cancer activities (3, 4,

37), and that a better understanding of microbial diversity will provide important insight into how to devise intelligent strategies for natural product discovery (7). The present study helps to establish a fundamental understanding of the diversity of Gram-positive bacteria in the marine environment and provides a diverse, marine-derived assemblage of cultured Gram-positive bacteria whose chemical and biosynthetic diversity can be investigated.

In addition to the isolation of actinomycetes from the *Micromonosporaceae*, *Nocardiaceae* and *Streptomycetaceae* families, spore-forming strains from the *Pseudonocardiaceae* and *Thermomonosporaceae*, as well as a large and diverse assemblage of unicellular and/or non-spore-forming Gram-positive bacteria, were cultured. While a diverse assemblage of bacteria within the *Actinomycetales* was cultured, no strains from other orders within the *Actinobacteria* were isolated despite the fact that bacteria from other orders have been identified in the marine environment using culture-independent methods (54).

Within the actinomycetes, the highest level of sequence divergence was observed within the *Nocardioidaceae* (Table 1.1), with all strains sharing < 98% sequence identity to currently described species. CNJ-780 and CNJ-872 were most closely related to *Marmoricola auranticus*, the only described species within the genus *Marmoricola*. Their percent identities with the type strain (94.8% and 97.0% respectively), suggest they may represent new species and, perhaps in the case of CNJ-872, a new genus within the *Nocardioidaceae*. Significant phylogenetic novelty was also observed among strains most closely related to the genera *Bacillus*, *Pontibacillus*, *Paenibacillus* and *Laceyella*. These strains appear to represent multiple new species

and, in the case of the *Paenibacillus* and *Laceyella* strains, which share only 91.9% and 92.3% sequence identity with their respective nearest type strains, possibly higher level taxa.

Of the potential new taxa observed, seven of the 11 *Bacillales* OTUs and three of the 18 *Actinomycetales* OTUs required seawater for growth (Figure 1.3). While possible that strains belonging to these OTUs also occur in non-marine environments, it is equally plausible that the seawater requiring OTUs represent obligate marine taxa. Both the number and phylogenetic distribution of these seawater requiring actinomycete and *Bacillales* strains was intriguing as they were clearly scattered throughout the phylogenetic tree (Figure 1.3). Thus, it remains possible that the requirement of seawater for growth either evolved rapidly and independently in these groups, was acquired by horizontal gene transfer, or represents a highly plastic phenotype.

The most remarkable intraclade diversity observed in the present study occurred within the genus *Bacillus*. This genus has been generally recognized to be among the most heterogeneous within the bacterial domain and in need of division into multiple genera (10). The present study recovered 45 strains most closely related to 17 described *Bacillus* species. These strains shared in some cases < 88% 16S rRNA gene sequence identity, far outside the sequence diversity associated with most bacterial genera. While a taxonomic reevaluation of the genus *Bacillus* in the near future is improbable, the results clearly indicate that considerable *Bacillus* diversity can be readily cultured from marine sediments.

Another noteworthy observation from this study was the recovery of 11 strains from six separate OTUs that share 100% 16S rRNA sequence identity with a type strain. While it was not surprising to culture *Salinispora arenicola* and *Serinicoccus marinus*, species previously reported from marine sediments and seawater respectively, the recovery of a strain with 100% sequence identity to *Kocuria palustris*, isolated originally from a cattail rhizopshere sampled at the Soroksar tributary of the Danube river, Hungary (Table 1.1), suggests that some bacterial strains exhibit remarkably broad geographical and environmental distributions.

There is currently much to learn about Gram-positive bacteria in marine sediments. Like their terrestrial relatives, marine Gram-positive bacteria may have a significant role in the breakdown of recalcitrant organic matter and may therefore play an integral role in the ocean's biogeochemical cycle. Additionally, even as spores, marine Gram-positive bacteria have the capacity to impact their surrounding chemical environment as evidenced by their capacity to oxidize metals (20). It is clear from this single survey that considerable new Gram-positive bacterial diversity can be readily cultured from marine sediments. The continued use of cultivation-dependent techniques will undoubtedly lead to the discovery of additional Gram-positive diversity, and provide a direct means to learn more about their ecophysiology and applications in biotechnology.

References.

1. **Austin, B**. 1988. Marine microbiology. Cambridge University Press, New York, N.Y.

2. **Bérdy, J**. 2005. Bioactive Microbial Metabolites. J. Antibiot. **58**:1-26.

3. **Bernan, V. S., M. Greenstein and G. T. Carter**. 2004. Mining marine microorganisms as a source of new antimicrobials and antifungals. Curr. Med. Chem. – Anti-Infective Agents. **3**:181-195.

4. **Blunt, J. W., B. R. Copp, M. H. G. Munro, P. T. Northcote and M. R. Prinsep**. 2006. Marine natural products. Nat. Prod. Rep. **23**:26-78.

5. **Bonde, G. J**. 1981. *Bacillus* from marine habitats: allocation to phena established by numerical techniques, p. 181-215. *In* R. C. W. Berkeley and M. Goodfellow (ed.), The aerobic endospore-forming bacteria: classification and identification. Academic Press Inc., New York, N.Y.

6. **Buck, J. D**. 1982. Nonstaining (KOH) method for determination of Gram reactions of marine bacteria. Appl. Environ. Microbiol. **44**:992-993.

7. **Bull, A. T**. 2004. Microbial diversity: the resource preamble, p. 13-14. *In* A.T. Bull (ed.), Microbial diversity and bioprospecting. ASM Press, Washington, D.C.

8. **Chao, A**. 1984. Non-parametric estimation of the number of classes in a population. Scand. J. Stat. **11**:265-270.

9. **Chazdon, R. L., R. K. Colwell, J. S. Denslow and M. R. Guariguata**. 1998. Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of N.E. Costa Rica, p. 285-309. *In* F. Dallmeier and J. A. Comiskey (ed.), Forest biodiversity research, monitoring and modeling: conceptual background and Old World case studies. Parthenon Publishing, Paris, France.

10. **Claus, D. and D. Fritze**. 1989. Taxonomy of *Bacillus*, p. 5-26. *In* C.R. Harwood (ed.), *Bacillus*. Plenum Press, New York, N.Y.

11. **Colquhoun, J. A., J. Mexson, M. Goodfellow, A. C. Ward, K. Horikoshi and A. T. Bull**. 1998. Novel rhodococci and other mycolate actinomycetes from the deep sea. Antonie van Leeuwenhoek. **74**:27-40.

12. **Colquhoun, J. A., S. C. Heald, L. Li, J. Tamaoka, C. Kato, K. Horikoshi and A. T. Bull**. 1998. Taxonomy and biotransformation activities of some deep-sea actinomycetes. Extremophiles. **2**:269-277.

13. **Connon, S. A. and S. J. Giovannoni**. 2002. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. Appl. Environ. Microbiol. **68**:3878-3885.

14. **Crawford, D. L., J. M. Lynch, J. M. Whipps and M. A. Ousley**. 1993. Isolation and characterization of actinomycete antagonists of a fungal root pathogen. Appl. Environ. Microbiol. **59**:3899-3905.

15. **Cresswell, N., V. A. Saunders, and E. M. H. Wellington**. 1991. Detection and quantification of *Streptomyces violaceolatus* plasmid DNA in soil. Lett. Appl. Microbiol. **13**:193-197.

16. **Devereux, R., S.-H. He, C. L. Doyle, S. Orkland, D. A. Stahl, J. LeGall and W. B. Whitman**. 1990. Diversity and origin of *Desulfovibrio* species: phylogenetic definition of a family. J. Bacteriol. **172**:3609-3619.

17. **Dunbar, J., S. Takala, S. M. Barns, J. A. Davis and C. R. Kuske**. 1999. Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning. Appl. Environ. Microbiol. **65**:1662-1669.

18. **Doi, R. H**. 1989. Sporulation and germination, p.169-215. *In* C.R. Harwood (ed.), *Bacillus*. Plenum Press, New York, N.Y.

19. **Foster, S. J. and D. L. Popham**. 2002. Structure and synthesis of cell wall, spore cortex, teichoic acids, S-layers, and capsules, p.21-42. *In* A. L. Sonenshein, J. A. Hoch and R. Losick (ed.), *Bacillus subtilis* and its closest relatives. ASM Press, Washington, D.C.

20. **Francis, C. A. and B. M. Tebo**. 2002. Enzymatic manganese (II) oxidation by metabolically dormant spores of diverse *Bacillus* species. Appl. Environ. Microbiol. **68**:874-880.

21. **Garrity, G. M., M. Winters and D. B. Searles.** 2001. Taxonomic outline of the procaryotic genera, p.155-166. *In* G. M. Garrity, D. R. Boone and R. W. Castenholz (ed.), Bergey's manual of systematic bacteriology, 2nd edition, volume 1. Springer-Verlag, New York, N.Y.

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

23. **Goodfellow, M., S. T. Williams and M. Mordarski**. 1983. Introduction to and importance of actinomycetes, p. 1-6. *In* M. Goodfellow, M. Mordarski and S. T. Williams (ed.), The biology of the actinomycetes. Academic Press Inc., London, England.

24. **Gugliandolo, C., T. L. Maugeri, D. Caccamo and E. Stackebrandt**. 2003. *Bacillus aeolius* sp. nov. a novel thermophilic, halophilic marine *Bacillus* species from Eolian Islands (Italy). Syst. Appl. Microbiol. **26**:172-176.

25. **Han, S. K., O. I. Nedashkovskaya, V. V. Mikhailov, S. B. Kim and K. S. Bae**. 2003. *Salinibacterium amurskyense* gen. nov., sp. nov., a novel genus of the family *Microbacteriaceae* from the marine environment. Int. J. Syst. Evol. Microbiol. **53**:2061-2066.

26. **Helmke, E. and H. Weyland**. 1984. *Rhodococcus marionescens* sp. nov., an actinomycete from the sea. Int. J. Syst. Bacteriol. **34**:127-138.

27. **Hill, R. T**. 2004. Microbes from marine sponges: a treasure trove of biodiversity for natural products discovery, p. 177-190. *In* A. T. Bull (ed.), Microbial diversity and bioprospecting. ASM Press, Washington, D.C.

28. **Imada, C., K. Hotta and Y. Okami**. 1998. A novel marine *Bacillus* with multiple amino acid analog resistance and selenomethionine-dependent antibiotic productivity. J. Mar. Biotechnol. **6**:189-192.

29. **Ivanova, E.P., M. V. Vysotskii, V. I. Svetashez, O. I. Nedashkovskaya, N. M. Gorshkova, V. V. Mikhailov, N. Yumoto, Y. Shigeri, T. Taguchi and S. Yoshikawa**. 1999. Characterization of *Bacillus* strains of marine origin. Int. Microbiol. **2**:267-271.

30. **Jensen, P. R., R. Dwight and W. Fenical**. 1991. Distribution of actinomycetes in near-shore tropical marine sediments. Appl. Environ. Microbiol. **57**:1102-1108.

31. **Jensen, P. R. and W. Fenical**. 1995. The relative abundance and seawater requirements of Gram-positive bacteria in near-shore tropical marine sediments. Microb. Ecol. **29**:249-257.

32. **Jensen, P. R., T. J. Mincer, P. G. Williams and W. Fenical**. 2005. Marine actinomycete diversity and natural product discovery. Antonie van Leeuwenhoek. **87**:43-48.

33. **Jensen, P. R., E. Gontang, C. Mafnas, T. J. Mincer and W. Fenical**. 2005. Culturable marine actinomycete diversity from tropical Pacific Ocean sediments. Environ. Microbiol. **7**:1039-1048.

34. **Jensen, P. R. and C. M. Mafnas**. 2006. Biogeography of the marine actinomycete *Salinispora*. Environ. Microbiol. **8**:1881-1888.

35. **Janssen, P. H., P. S. Yates, B. E. Grinton, P. M. Taylor and M. Sait**. 2002. Improved culturability of soil bacteria and isolation in pure culture of novel members of the divisions *Acidobacteria, Actinobacteria, Proteobacteria*, and *Verrucomicrobia*. Appl. Environ. Microbiol. **68**:2391-2396.

36. **Joseph, S. J., P. Hugenholtz, P. Sangwan, C. A. Osborne and P. H. Janssen**. 2003. Laboratory cultivation of widespread and previously uncultured soil bacteria. Appl. Environ. Microbiol. **69**:7210-7215.

37. **Kwon, H. C., C. A. Kauffman, P. R. Jensen and W. Fenical**. 2006. Marinomycins A-D, antitumor-antibiotics of a new structure class from a marine actinomycete of the recently discovered genus "*Marinispora*." J. Am. Chem. Soc. **128**:1622-1632.

38. **Laiz, L., G. Piñar, W. Lubitz and C. Saiz-Jimenez**. 2003. Monitoring the colonization of monuments by bacteria: cultivation versus molecular methods. Environ. Microbiol. **5**:72-74.

39. **Li, L., J. Guezennec, P. Nichols, P. Henry, M. Yanagibayashi and C. Kato**. 1999. Microbial diversity in Nankai Trough sediments at a depth of 3,843 m. J. Oceanography. **55**:635-642.

40. **Locci, R. and G. P. Sharples**. 1983. Morphology, p.165-199. *In* M. Goodfellow, M. Mordarski and S. T. Williams (ed.), The biology of the actinomycetes. Academic Press Inc., London, England.

41. **Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Förster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lüßmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode and K.-H. Schleifer**. 2004. ARB: a software environment for sequence data. Nucleic Acids Res. **32**:1363-1371.

42. **Madigan, M. T., J. M. Martinko and J. Parker**. 2003. Brock biology of microorganisms 10th ed. Pearson Education Inc., Upper Saddle River, N.J.

43. **Maldonado, L. A., J. E. M. Stach, W. Pathom-aree, A. C. Ward, A. T. Bull and M. Goodfellow**. 2005. Diversity of cultivable actinobacteria in geographically widespread marine sediments. Antonie van Leeuwenhoek. **87**:11-18.

44. **Maldonado, L. A., W. Fenical, P. R. Jensen, C. A. Kauffman, T. J. Mincer, A. C. Ward, A. T. Bull and M. Goodfellow**. 2005. *Salinispora arenicola* gen. nov., sp. nov. and *Salinispora tropica* sp. nov., obligate marine actinomycetes belonging to the family *Micromonosporaceae*. Int. J. Syst. Evol. Microbiol. **55**:1759-1766.

45. **Mincer, T. J., P. R. Jensen, C. A. Kauffman and W. Fenical**. 2002. Widespread and persistent populations of a major new marine actinomycete taxon in ocean sediments. Appl. Environ. Microbiol. **68**:5005-5011.

46. **Mincer, T. J., W. Fenical and P. R. Jensen**. 2005. Culture-dependent and culture-independent diversity within the obligate marine actinomycete genus *Salinispora*. Appl. Environ. Microbiol. **71**:7019-7028.

47. **Montalvo, N. F., N. M. Mohamed, J. J. Enticknap and R. T. Hill**. 2005. Novel actinobacteria from marine sponges. Antonie van Leeuwenhoek. **87**:29-36.

48. **Moran, M. A., L. T. Rutherford and R. E. Hodson**. 1995. Evidence for indigenous *Streptomyces* populations in a marine environment determined with a 16S rRNA. Appl. Environ. Microbiol. **61**:3695-3700.

49. **Nesterenko, O. A., S. A. Kasumova and E. I. Kvasnikov**. 1978. Microorganisms of the *Nocardia* genus and the '*rhodochrous'* group in soils of the Ukranian SSR. Mikrobiologiya. **47**:866-870.

50. **Nicholson, W. L., N. Munakata, G. Horneck, H. J. Melosh and P. Setlow**. 2000. Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments. Microbiol. Mol. Bio. Rev. **64**:548-572.

51. **Nold, S. C., E. D. Kopczynski and D. M. Ward**. 1996. Cultivation of aerobic chemoorganotrophic proteobacteria and Gram-positive bacteria from a hot spring microbial mat. Appl. Environ. Microbiol. **62**:3917-3921.

52. **Priest, F. G**. 1989. Isolation and identification of aerobic endospore-forming bacteria, p.27-56. *In* C.R. Harwood (ed.), *Bacillus*. Plenum Press, New York, N.Y.

53. **Priest, F. G.** 1989. Products and applications, p. 293-320. *In* C. R. Harwood (ed.), *Bacillus*. Plenum Press, New York, N.Y.

54. **Rappé, M. S., D. A. Gordon, K. L. Vergin and S. J. Giovannoni**. 1999. Phylogeny of actinobacteria small subunit (SSU) rRNA gene clones recovered from marine bacterioplankton. Syst. Appl. Microbiol. **22**:106-112.

55. **Rappé, M. S., S. A. Connon, K. L. Vergin and S. J. Giovannoni**. 2002. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature. **418**:630-633.

56. **Ruger, H. J., D. Fritze and C. Sproer**. 2000. New psychrophilic and psychrotolerant *Bacillus marinus* strains from tropical and polar deep-sea sediments and emended description of the species. Int. J. Syst. Bacteriol. **50**:1305-1313.

57. **Sait, M., P. Hugenholtz and P. H. Janssen**. 2002. Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. Environ. Microbiol. **4**:654-666.

58. **Stach, J. E. M., L. A. Maldonado, D. G. Masson, A. C. Ward, M. Goodfellow and A. T. Bull**. 2003. Statistical approaches for estimating actinobacterial diversity in marine sediments. Appl. Environ. Microbiol. **69**:6189-9200.

59. **Stach, J. E. M. and A. T. Bull**. 2005. Estimating and comparing the diversity of marine actinobacteria. Antonie van Leeuwenhoek. **87**:3-9.

60. **Stevenson, B. S., S. A. Eichorst, J. T. Wertz, T. M. Schmidt and J. A. Breznak**. 2004. New strategies for cultivation and detection of previously uncultured microbes. Appl. Environ. Microbiol. **70**:4748-4755.

61. **Suzuki, M. T. and S. J. Giovannoni**. 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. **62**:625-630.

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

63. **Swofford, D. L**. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Software version 4.0b 10. Sinauer Associates, Sunderland, MA.

64. **Takizawa, M., R. R. Colwell and R. T. Hill**. 1993. Isolation and Diversity of Actinomycetes in Chesapeake Bay. Appl. Environ. Microbiol. **59**:997-1002.

65. **Tamaki, H., Y. Sekiguchi, S. Hanada, K. Nakamura, N. Nomura, M. Matsumura and Y. Kamagata**. 2005. Comparative analysis of bacterial diversity in freshwater sediment of a shallow eutrophic lake by molecular and improved cultivation-based techniques. Appl. Environ. Microbiol. **71**:2162-2169.

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

67. **von Wintzingerode, F., U. B. Göbel and E. Stackebrandt**. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol. Rev. **21**:213-229.

68. **Weyland, H**. 1969. Actinomycetes in North Sea and Atlantic Ocean sediments. Nature **223**:858.

69. **Yi, H., P. Schumann, K. Sohn and J. Chun**. 2004. *Serinicoccus marinus* gen. nov., sp. nov., a novel actinomycete with L-ornithine and L-serine in the peptidoglycan. Int. J. Syst. Evol. Microbiol. **54**:1585-1589.

70. **Zengler, K., G. Toledo, M. Rappé, J. Elkins, E. J. Mathur, J. M. Short and M. Keller**. 2002. Cultivating the uncultured. Proc. Natl. Acad. Sci. U. S. A. **99**:15681-15686.

71. **Zhuang, W.-Q., J.-H. Tay, A. M. Maszenan and S. T. L. Tay**. 2002. *Bacillus naphthovorans* sp. nov. from oil-contaminated tropical marine sediments and its role in naphthalene biodegredation. Appl. Microbiol. Biotechnol. **58**:547-553.

72. **ZoBell, C. E**. 1946. Marine microbiology: A monograph on hydrobacteriology. Chronica Botanica Co., Waltham, Mass.

Table 1.1: List of isolates representing the 52 Actinomycetales OTUs generated using a 16S rRNA percent identity value of ≥ 98%. For each OTU, the nearest type strain and its source are identified. When an isolate and its nearest type strain share < 98% sequence identity, the percent sequence identity is in bold.

Phylogenetic group (family)	Representative isolate (accession number)	Sequence length (bp)	No. of strains in OTU	Nearest type strain (accession number)	Sequence identity (%) a	Source of nearest type strain
Brevibacteriaceae	CNJ737 (DQ448693)	1480	1	*Brevibacterium epidermidis* (X76565)	99.8	skin
Corynebacterium	CNJ954 (DQ448694)	1480	1	*Corynebacterium variabile* (AJ222815)	98.6	animal fodder
Dermacoccaceae	CNJ855 (DQ448695)	1476	3	*Kytococcus sedentarius* (X87755)	99.6	skin
Dietziaceae	CNJ898 (DQ448696)	1469	8	*Dietzia maris* (X79290)	99.9	marine sediment
Geodermatophilaceae	CNJ868 (DQ448697)	1471	1	*Blastococcus saxobsidens* (AJ316570)	98.8	stone surfaces
	CNJ793 (DQ448698)	1472	2	*Modestobacter multiseptatus* (Y18646)	98.1	soil
Gordoniaceae	CNJ756 (DQ448699)	1473	5	*Gordonia bronchialis* (X79287)	98.4	human sputum
	CNJ863 (DQ448700)	1475	1	*Gordonia nitida* (AF148947)	100	industrial wastewater
	CNJ754 (DQ448701)	1473	3	*Gordonia polyisoprenivorans* (Y18310)	98.4	deteriorated automobile tire
	CNJ752 (DQ448702)	1475	4	*Gordonia terrae* (X81922)	100	soil
Intrasporangiaceae	CNJ824 (DQ448703)	1477	3	*Ornithinimicrobium humiphilum* (AJ277650)	**97.2**	soil
	CNJ927 (DQ448704)	1495	1	*Serinicoccus marinus* (AY382898)	100	sea water
Microbacteriaceae	CNJ745 (DQ448705)	1479	2	*Agromyces aurantiacus* (AF389342)	98.2	soil
	CNJ930 (DQ448706)	1477	2	*Microbacterium flavescens* (Y17232)	**97.3**	soil
	CNJ743 (DQ448707)	1475	1	*Microbacterium imperiale* (X77442)	**97.9**	moth's alimentary tract
	CNJ797 (DQ448708)	1474	1	*Microbacterium schleiferi* (Y17237)	99	activated sludge
Micrococcaceae	CNJ723 (DQ448709)	1377	2	*Kocuria marina* (AY211385)	**96.6**	marine sediment
	CNJ900 (DQ448710)	1479	1	*Kocuria palustris* (Y16263)	100	cattail rhizosphere
	CNJ770 (DQ448711)	1479	2	*Kocuria rosea* (X87756)	98.2	soil
	CNJ719 (DQ448712)	1475	13	*Micrococcus luteus* (AF542073)	99.3	wall painting
Micromonosporaceae	CNS326 (DQ448713)	1468	1	*Micromonospora endolithica* (AJ560635)	**97.8**	Antarctic sandstone
	CNJ878 (DQ448714)	1469	1	*Micromonospora endolithica* (AJ560635)	98.6	Antarctic sandstone
	CNS051 (DQ448715)	1468	16	*Salinispora arenicola* (AY040619)	100	marine sediment
	CNS143 (DQ092624)	1468	7	*Salinispora tropica* (AY040617)	99.6	marine sediment
Mycobacteriaceae	CNJ859 (DQ448716)	1472	3	*Mycobacterium brisbanense* (AY012577)	98.5	human
	CNJ823 (DQ448717)	1472	7	*Mycobacterium porifera* (AF480589)	99.9	Finland sponge
Nocardiaceae	CNS044 (DQ448718)	1471	2	*Nocardia arthritidis* (AB108781)	99.3	human
Nocardioidaceae	CNJ889 (DQ448719)	1475	1	*Aeromicrobium erythreum* (AF005021)	**94.8**	soil
	CNJ780 (DQ448720)	1471	1	*Marmoricola auranticus* (Y18629)	**94.8**	marble statue
	CNJ872 (DQ448721)	1472	1	*Marmoricola auranticus* (Y18629)	**97**	marble statue
	CNJ892 (DQ448722)	1468	1	*Nocardioides ganghwensis* (AY423718)	**97.9**	tidal flat sediment
Nocardiopsaceae	CNR923 (DQ448723)	1485	1	*Nocardiopsis lucentensis* (X97888)	99	salt marsh soil
Promicromonosporaceae	CNJ734 (DQ448724)	1474	1	*Promicromonospora sukumoe* (AJ272024)	**97.3**	soil
Pseudonocardiaceae	CNJ888 (DQ448725)	1483	1	*Pseudonocardia antarctica* (AJ576010)	**94.4**	soil
	CNS139 (DQ448726)	1472	1	*Pseudonocardia yunnanensis* (AJ252822)	**97.7**	soil
	CNS004 (DQ448727)	1474	1	*Pseudonocardia zijingensis* (AF325725)	98.7	soil
Streptomycetaceae	CNR884 (DQ448728)	1478	1	*Streptomyces arenae* (AJ399485)	99.5	soil
	CNR926 (DQ448729)	1470	2	*Streptomyces aureofaciens* (AY289116)	**97.7**	soil
	CNR881 (DQ448730)	1476	1	*Streptomyces bikiniensis* (X79851)	98.7	soil
	CNR918 (DQ448731)	1480	1	*Streptomyces caviscabies* (AF112160)	99.5	potato lesion
	CNR924 (DQ448732)	1478	1	*Streptomyces chartreusis* (AJ399468)	99.4	soil
	CNR875 (DQ448733)	1478	1	*Streptomyces galilaeus* (AB045878)	98.7	soil
	CNR872 (DQ448734)	1497	1	*Streptomyces hebeiensis* (AY277529)	**95.5**	soil
	CNR880 (DQ448735)	1475	1	*Streptomyces koyangensis* (AY079156)	99.2	soil
	CNS177 (DQ448736)	1481	1	*Streptomyces lydicus* (Y15507)	99.2	potato scab
	CNJ962 (DQ448737)	1495	1	*Streptomyces sampsonii* (D63871)	**95.5**	potato scab
	CNR887 (DQ448738)	1477	1	*Streptomyces sampsonii* (D63871)	99	potato scab
	CNR885 (DQ448739)	1481	3	*Streptomyces tendae* (D63873)	98.9	potato scab
	CNR877 (DQ448740)	1423	1	*Streptomyces thermocarboxydovoran* (U94489)	**95.6**	soil
	CNR940 (DQ448741)	1482	2	*Streptomyces thermocoprophilus* (AJ007402)	**96.8**	poultry feces
	CNR925 (DQ448742)	1480	2	*Streptomyces thermocoprophilus* (AJ007402)	**97.3**	poultry feces
Thermomonosporaceae	CNU125 (DQ448743)	1472	1	*Actinomadura cremea* (AF134067)	99.1	soil

a Sequence identity shared between the representative isolate and its nearest type strain.

Table 1.2: List of isolates representing the 26 Bacillales OTUs generated using a 16S rRNA percent identity value of ≥ 98%. For each OTU, the nearest type strain and its source are identified. When an isolate and its nearest type strain share < 98% sequence identity, the percent sequence identity is in bold.

Phylogenetic group (family)	Representative isolate (accession number)	Sequence length (bp)	No. of strains in OTU	Nearest type strain (accession number)	Sequence identity (%) a	Source of nearest type strain
Bacillaceae	CNJ803 (DQ448744)	1502	7	*Bacillus algicola* (AY228462)	99.8	*F. evanescens* thallus
	CNJ796 (DQ448745)	1505	3	*Bacillus aquimaris* (AF483625)	**96.1**	seawater
	CNJ733 (DQ448746)	1504	3	*Bacillus aquimaris* (AF483625)	99.5	seawater
	CNJ815 (DQ448747)	1502	3	*Bacillus barbaricus* (AJ422145)	99.4	experimental wall painting
	CNJ826 (DQ448748)	1501	1	*Bacillus bataviensis* (AJ542508)	**96.9**	soil
	CNJ732 (DQ448749)	1504	2	*Bacillus cereus* (AE017013)	100	air
	CNJ816 (DQ448750)	1504	1	*Bacillus cohnii* (X76437)	98.2	horse meadow soil
	CNJ828 (DQ448751)	1503	5	*Bacillus decoloationis* (AJ315075)	**97.9**	mural painting
	CNJ958 (DQ448752)	1444	1	*Bacillus endophyticus* (AF295302)	**95**	cotton plant inner tissue
	CNJ905 (DQ448753)	1502	1	*Bacillus firmus* (AJ717384)	**97.8**	nonsal. alkaline groundwater
	CNJ933 (DQ448754)	1503	2	*Bacillus firmus* (AJ717384)	99.7	nonsal. alkaline groundwater
	CNJ759 (DQ448755)	1504	1	*Bacillus horikoshii* (X76443)	99.5	soil
	CNJ775 (DQ448756)	1462	3	*Bacillus humi* (AJ627210)	**96.4**	soil
	CNJ782 (DQ448757)	1490	2	*Bacillus indicus* (AJ583158)	99.9	aresenic polluted sand
	CNJ778 (DQ448758)	1504	6	*Bacillus megaterium* (X60629)	99.9	soil
	CNJ748 (DQ448759)	1503	1	*Bacillus methanolicus* (AB112727)	**96.3**	sugar beet waste water facility
	CNJ742 (DQ448760)	1500	3	*Bacillus pumilus* (AY876289)	99.4	soil
	CNJ771 (DQ448761)	1514	4	*Exiguobacterium aestuarii* (AY594264)	99.7	seawater, Korea
	CNJ915 (DQ448762)	1517	3	*Halobacillus litoralis* (X94558)	99.2	salt marsh soil
	CNJ895 (DQ448763)	1518	2	*Halobacillus salinus* (AF500003)	99.9	East Sea coast salt lake
	CNJ812 (DQ448764)	1475	3	*Halobacillus trueperi* (AJ310149)	98	saline sediment, salt lake
	CNJ912 (DQ448765)	1515	1	*Pontibacillus chungwhensis* (AY553296)	**97.1**	Korean solar saltern
Paenibacillaceae	CNJ934 (DQ448766)	1508	1	*Paenibacillus turicensis* (AF378697)	**91.9**	cerebrospinal fluid shunt
Staphylococcaceae	CNJ924 (DQ448767)	1503	1	*Staphylococcus capitis* (AY688040)	99.8	human skin
Thermoactinomycetaceae	CNR949 (DQ448768)	1497	2	*Laceyella sacchari* (AF138737)	**91.9**	soil
	CNJ795 (DQ448769)	1469	3	*Laceyella sacchari* (AF138737)	**92.3**	soil

a Sequence identity shared between the representative isolate and its nearest type strain.

Table 1.3: Number of OTUs, generated using a 16S rRNA gene sequence identity of ≥ 98%, and strains belonging to new and known phylotypes. All new phylotypes share < 98% 16S rRNA gene sequence identity with cultured and sequenced bacteria. Known phylotypes that are most closely related to a type strain isolated from a marine source are in parentheses.

Order	New		Known (Marine)		Total	
	OTUs	Strains	OTUs	Strains	OTUs	Strains
Actinomycetales	12	14	40 (5)	110 (39)	52	124
Bacillales	9	16	17 (7)	49 (22)	26	65

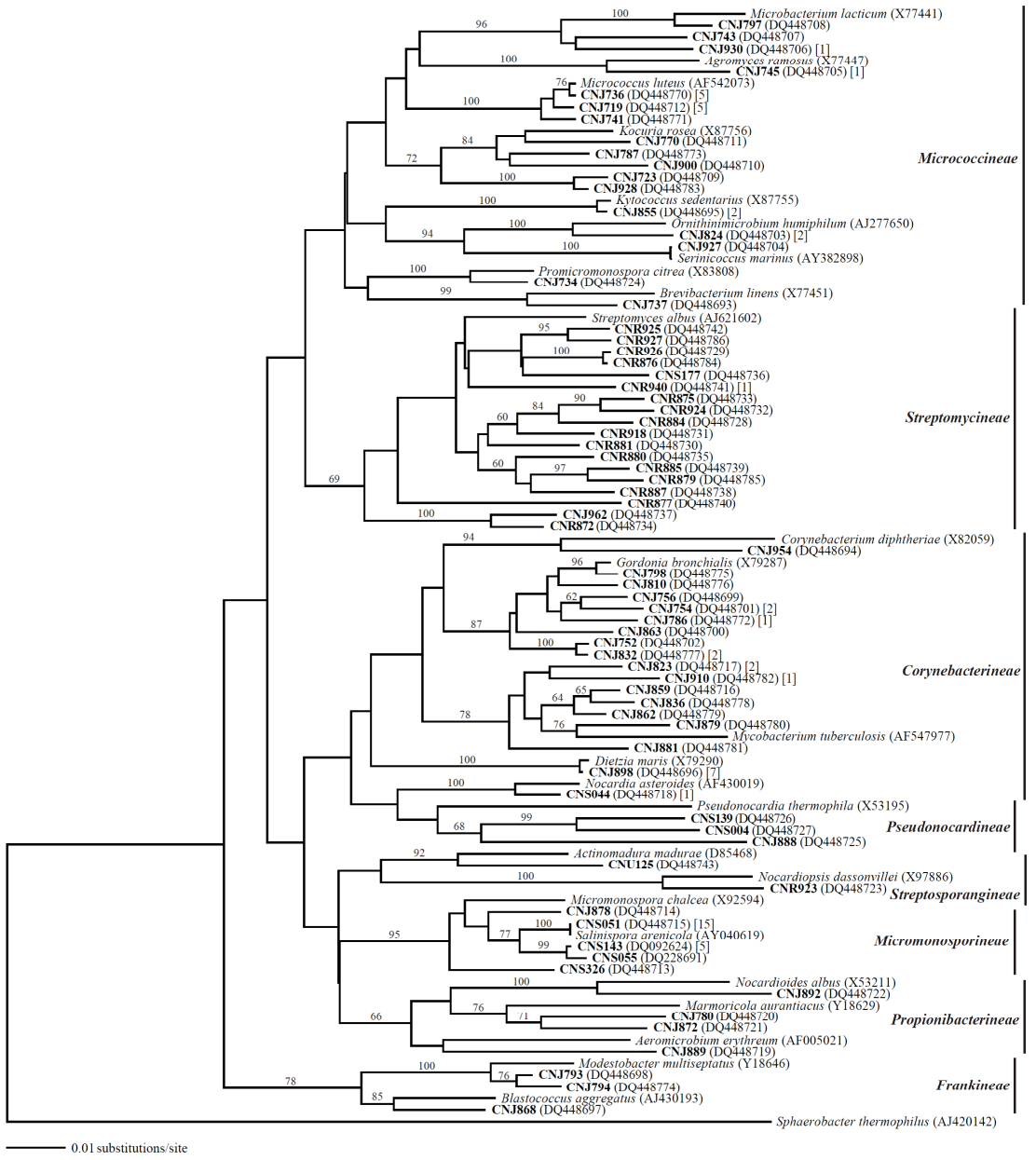

Figure 1.1: Neighbor-joining distance tree constructed in PAUP (63) using the aligned, partial 16S rRNA gene sequences (512 nucleotide positions) of strains representing each of the 70 *Actinomycetales* OTUs (generated using a sequence identity value of 100%) and the type strains of the most closely related genera. Sequences from this study are shown in bold and GenBank accession numbers are given in parentheses. Bootstrap values (in percent) calculated from 1000 re-samplings using the neighbor-joining method are shown at their respective nodes for values ≥ 60%. *Sphaerobacter thermophilus* was used to position the root. When multiple strains shared an identical, partial 16S rRNA gene sequence, the number of additional isolates is presented in brackets. The sub-orders to which the strains belong are presented on the right. Family and genus level affiliations were maintained when distance UPGMA and maximum parsimony treeing methods were applied although some within family branching patterns changed.

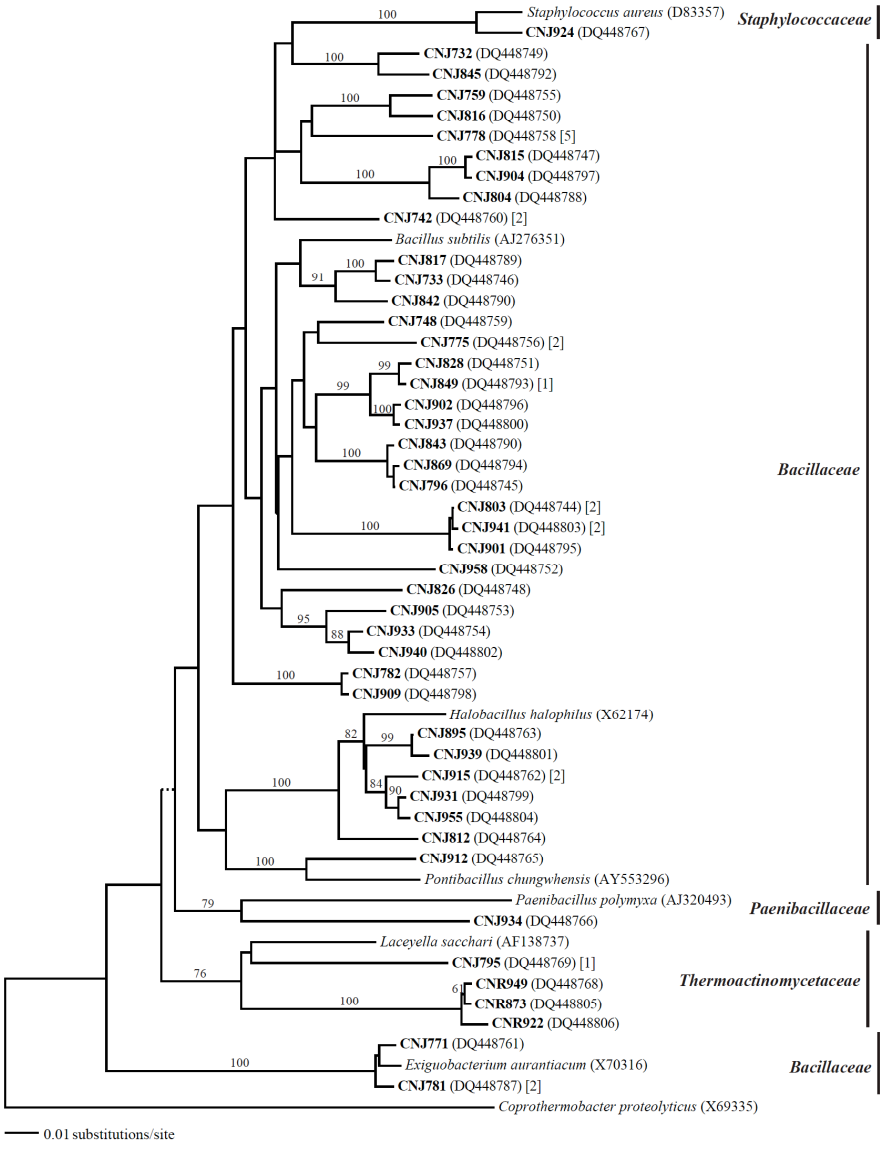

Figure 1.2: Neighbor-joining distance tree constructed in PAUP (63) using the aligned, partial 16S rRNA gene sequences (512 nucleotide positions) of strains representing each of the 46 *Bacillales* OTUs (generated using a sequence identity value of 100%) and the type strains of the most closely related genera (with the exception of *Exiguobacterium aurantiacum* and *Halobacillus halophilus*, for which alternative sequences were used). Sequences from this study are shown in bold and GenBank accession numbers are given in parentheses. Bootstrap values (in percent) calculated from 1000 re-samplings using the neighbor-joining method are shown at their respective nodes for values ≥ 60%. *Coprothermobacter proteolyticus* was used to position the root. When multiple strains shared an identical, partial 16S rRNA gene sequence, the number of additional isolates is presented in brackets. The families to which the strains belong are presented on the right. Family and genus level affiliations were maintained when distance UPGMA and maximum parsimony treeing methods were applied although some within family branching patterns changed.

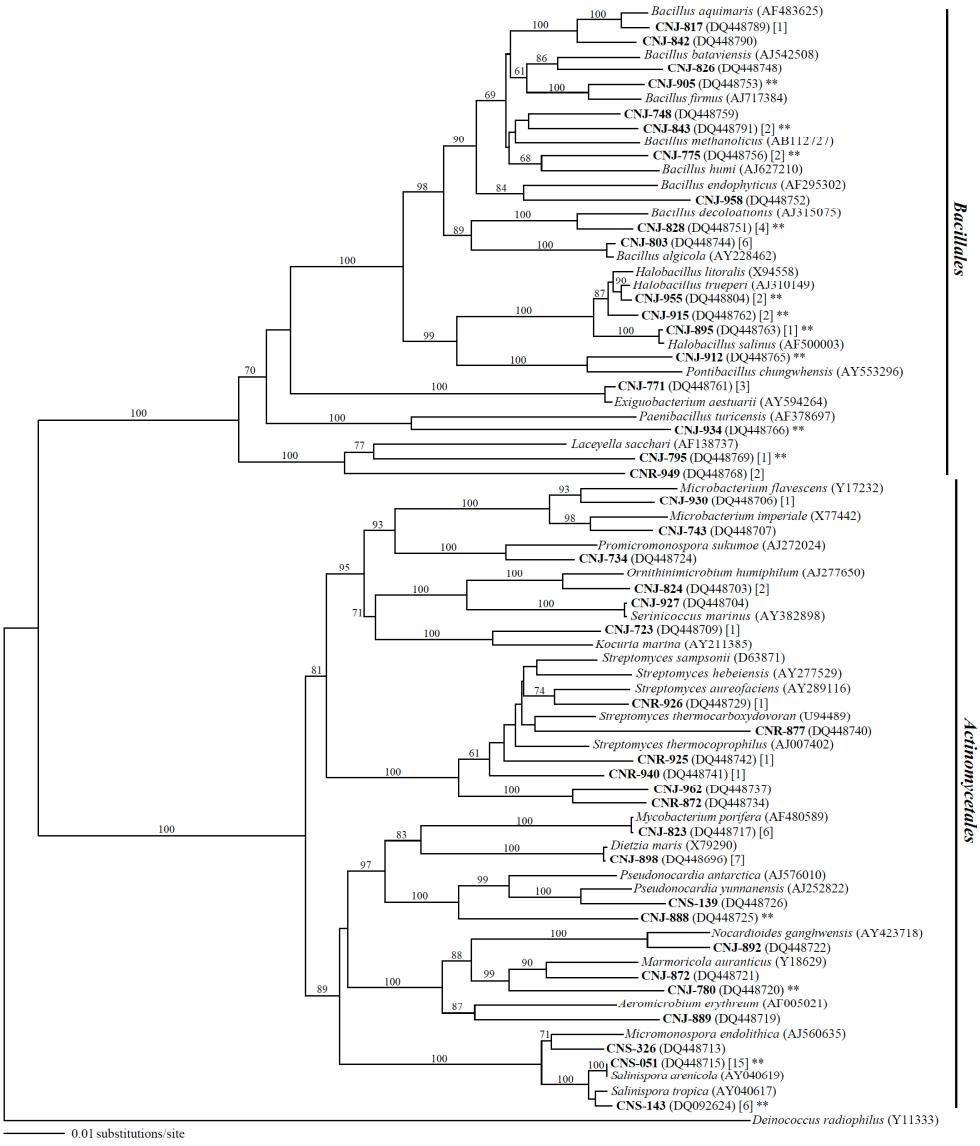

Figure 1.3: Neighbor-joining distance tree based on the nearly complete and aligned 16S rRNA gene sequences of 41 *Actinomycetales* and *Bacillales* OTUs observed in this study and their nearest type strains. The strains used to construct this tree represent the 29 OTUs that have not yet been formally described and the 12 OTUs whose nearest type strain was isolated from a marine source. A total of 1,367 nucleotide positions were included in the analysis and *Deinococcus radiophilus* was used to position the root. GenBank accession numbers are given in parentheses following the strain identification (in bold). Bootstrap values (in percent) calculated from 1000 re-samplings using the neighbor-joining method are shown at their respective nodes for values ≥ 60%. The number of additional isolates within each OTU (calculated using a sequence identity value of ≥ 98%) is presented in brackets. Double asterisks (**) indicate OTUs in which all of the tested isolates required seawater for growth. The topology of the distance neighbor-joining tree is supported by distance UPGMA and maximum likelihood treeing methods.

Chapter Two: An Investigation of the Mechanisms Driving the Evolution of Type I Modular PKS Pathways in *Salinispora arenicola*

Introduction.

In 1991, the distribution of cultivable actinomycetes in near-shore tropical marine sediments collected in the Bahamas was described (20). From those sediments, isolates from the *Streptomycetaceae* and *Micromonosporaceae* families were cultured (20). Among the *Micromonosporaceae* cultured were marine obligate isolates that were later designated MAR1 bacteria (21). Populations of MAR1 bacteria have since been found to be persistent and widespread in ocean sediments, cultured from the Pacific and Atlantic oceans as well as the Red Sea and the Sea of Cortez (32). In 2005, the MAR1 bacteria were recognized as a new actinomycete taxon and described as the genus *Salinispora* (30). The genus *Salinispora* is currently comprised of two species, *Salinispora arenicola* and *Salinispora tropica*. However, since the description of these first two species, a third species ("*Salinispora pacifica*") has been discovered and its formal description is being prepared.

Since 1991, thousands of *Salinispora* strains have been collected from all over the globe and the distinct biogeographical patterns in the distribution of the three species documented (22). *S. arenicola* isolates are the most abundant species cultured and have been described as having a pantropical distribution whereas *S. tropica* has a limited distribution, having only been isolated from sediments collected in the Bahamas. "*S. pacifica*" isolates have been cultured from sediments collected in the Red Sea, Guam and the Republic of Palau, however this species is isolated with less frequency than *S. arenicola* (22).

Like other filamentous, spore forming actinomycetes, *Salinispora* strains produce a wealth of diverse secondary metabolites (Figure 2.1). Interestingly, *Salinispora* strains produce secondary metabolites in species-specific patterns (23), contradicting traditional paradigms of secondary metabolite strain specificity. For example, all *S. tropica* isolates examined synthesize salinosporamide A (9), and to date, this metabolite is only observed from *S. tropica*. While most compounds elucidated from the genus can be classified as species specific, it has been observed that on occasion a strain will produce a new compound in addition to the core metabolites associated with that species. It has been approximated that almost 10% of the *Salinispora* isolates evaluated produce "accessory" secondary metabolites (23). In an effort to better understand and explore the biosynthetic potential of the genus *Salinispora*, the genomes of *S. arenicola* strain CNS-205 and *S. tropica* strain CNB-440 were sequenced.

In 2007, the genomes of *S. arenicola* strain CNS-205 and *S. tropica* strain CNB-440 were sequenced and annotated by the Joint Genome Institute (JGI), in collaboration with the genome analysis and system modeling group of the life sciences division of Oak Ridge National Laboratory (ORNL). Both the genome of *S. arenicola* strain CNS-205 and the genome of *S. tropica* strain CNB-440 revealed a single circular chromosome with no plasmids. The genome of CNS-205 has a length of 5,786,361 nucleotides and is estimated to have 4,917 protein coding genes whereas the genome of CNB-440 is 5,183,331 nucleotides in length and is estimated to have 4,536 protein coding genes. Both genomes have an average G+C content of 69.5%. Putative clusters of genes associated with secondary metabolism were identified by

sequentially searching each translated protein sequence using BLASTP against a library of natural product domains and genes (47, 48). The analysis of the secondary metabolite biosynthetic gene clusters of both *S. arenicola* and *S. tropica* identified 25 and 17 secondary metabolite biosynthetic pathways, respectively (Table 2.1). By comparing and contrasting the pathways present in each species, the fermentation-based observation that certain metabolites are species-specific is found to have a genetic basis (23). The analyses also reveal that the biosynthetic potential of these two species has yet to be completely realized.

The sequencing of a bacterial genome provides an opportunity to assess the biosynthetic potential of an individual strain. For example, the genomic analysis of the long studied *Streptomyces coelicolor* strain A3(2) revealed clusters of genes characteristic of biosynthetic pathways responsible for the production of never before isolated small molecules (3). Prior to the sequencing of its genome, *Streptomyces coelicolor* had been studied for over forty years and was known to produce only three secondary metabolites, an aromatic polyketide known as actinorhodin, a non-ribosomal peptide known as CDA, and a type II polyketide derived grey spore pigment. The genome sequence of *Streptomyces coelicolor* strain A3(2) suggested however, that the strain had the capacity to produce at least 18 additional secondary metabolites. By comparing the sequenced clusters with similar, previously characterized biosynthetic pathways and by applying sequence analysis tools, Bentley and his colleagues were able to predict the probable structure of end products for several of the pathways (3). Moreover, by mining the genome of *Streptomyces coelicolor*, Challis and his colleagues were able to predict the production of a specific

secondary metabolite and through the design of the appropriate fermentation medium, demonstrate that their prediction was largely correct (25). Genome sequencing offers insight into the biosynthetic potential of strains and provides clear evidence that many natural products have yet to be discovered. It is possible that secondary metabolite pathways are only active under specific physical, chemical and biological conditions and therefore have not been identified by natural product discovery programs because strains have not yet been cultured under the correct environmental conditions. Natural product discovery programs may need to consider growing strains in the presence of physical (desiccation, high/low temperature), chemical (low nutrient, low iron) and/or biological (competition) stresses, and observe cultures at different stages of their development, in order to target or capture the production of many as yet unelucidated compounds.

Like the genome of *Streptomyces coelicolor* strain A3(2), the genome sequences of *Salinispora arenicola* strain CNS-205 and *Salinispora tropica* strain CNB-440 suggest that the secondary metabolite biosynthetic potential of these two species has not yet been fully recognized. Both *S. arenicola* and *S. tropica* harbor a wealth of diverse and novel secondary metabolite gene clusters including polyketide synthase (PKS), nonribosomal peptide synthetase (NRPS) and hybrid PKS/NRPS biosynthetic gene clusters (47, 48). Although several of the compounds biosynthesized by the PKS, NRPS and PKS/NRPS hybrid gene clusters have already been elucidated, the genome sequences suggest *Salinispora* strains possess a largely untapped potential to produce secondary metabolites.

During a comparative genomic analysis of the *Salinispora* strains, it was discovered that many of the secondary metabolite biosynthetic gene clusters are located on distinct genomic islands (personal communication K. Penn). Genomic islands describe regions within a genome dominated by genes that are absent in the most closely related strains (8). For many years the genes of genomic islands and their evolutionary implications have been discussed. Among the earliest genomic islands studied were pathogenicity islands, or regions within bacterial genomes that harbor virulence genes. Pathogenicity islands were identified as the only genetic difference between some virulent and avirulent bacterial strains (16, 27), allowing bacteria harboring the pathogenicity island to survive within a host and cause disease. Over the past decade, studies have also begun to investigate the genomic islands of non-pathogenic bacterial species and the effect genes of these genomic islands have on the relative fitness of a strain or species (8, 17). For example, an investigation of *Prochlorococcus* suggested that genomic islands have played an important role in the evolution these cyanobacteria, contributing to or allowing niche differentiation among strains or species (7). To date, it has been suggested that genetic information present on genomic islands can increase the overall fitness of a bacterial strain and that the genes encoded on the genomic island may contribute to the evolution of both pathogenic and non-pathogenic bacteria (8).

Of the *S. arenicola* genomic islands identified, several harbor pathways associated with secondary metabolite biosynthesis. On two such genomic islands, *S. arenicola* possesses gene clusters associated with type I modular polyketide biosynthesis. On the genomic island that stretches from gene Sare1204 through gene Sare1314, a type I

modular PKS pathway, with a striking resemblance to the rifamycin biosynthetic pathway of *Amycolatopsis mediterranei* (GenBank accession number AF040570), was identified. A second type I modular PKS pathway was identified on the genomic island that spans from gene Sare3138 through gene Sare3170. While the product of this latter pathway remains unknown, the pathway is predicted to biosynthesize a macrolide.

The identification of as yet uncharacterized gene clusters is significant because many PKS, NRPS and hybrid PKS/NRPS derived natural products possess biomedically relevant properties, including antibiotic, antitumor, antifungal, and immunosuppressive activities. In fact, the PKS/NRPS derived salinosporamide A, a small molecule produced by *S. tropica* and discovered in 2003 (9), is currently in phase I clinical trials as an anticancer drug (38). PKS, NRPS and PKS/NRPS natural products have proven to be one of the primary sources of new chemical scaffolds around which pharmaceutical agents have been developed (33, 34) and recent advances in the areas of small molecule detection, isolation and structure elucidation, genome sequencing and molecular techniques that improve our understanding of the molecular genetics of secondary metabolite biosynthesis, are rekindling an interest in discovery platforms aimed at realizing the biosynthetic capacity of rare and novel microorganisms.

Polyketides, nonribosomal peptides and polyketide/nonribosomal peptide hybrids are synthesized by the coordinated action of enzymatic assembly lines (PKS, NRPS and PKS/NRPS hybrid pathways respectively) which conduct the iterative chemical condensation of monomeric units, including acyl-CoA and/or amino acid monomers

(6, 10, 49). The order and identity of each domain within a given assembly line specifies the sequence of monomer activation and incorporation, the chemistry that occurs at each step in the assembly line, and the length and functionality of the product released from the assembly line (10). The condensation of a carboxylic acid monomer to a growing acyl chain is accomplished via ketosynthase (KS), acyltransferase (AT) and acyl carrier protein (ACP) domains within PKSs. Following condensation, the presence and functionality of ketoreductase (KR), dehydratase (DH) and enoylreducase (ER) domains determines whether the oxidation state of the beta-carbonyl remains either a keto group or is reduced to a hydroxyl, methine or methylene (19). The incorporation of amino acid monomers into non-ribosomal peptides is accomplished via the condensation (C), adenylation (A) and peptidyl carrier protein (PCP) domains associated with NRPSs. When present, epimerization (E), methyltransferase (MT) and oxidase (Ox) domains tailor the amino acid being added (10). The immense structural complexity and functional diversity of polyketide, nonribosomal peptide or hybrid compounds is generated both by the intricately coordinated enzyme organization of these assembly lines and specific tailoring enzymes responsible for post-assembly modifications (6, 10).

The genome sequence of *S. tropica* revealed the most diverse assemblage of polyketide biosynthetic pathways ever observed in a single genome (47). With the recent publication of the *S. arenicola* genome, it has become clear that *S. arenicola* harbors an equally diverse and arguably richer collection of polyketide biosynthetic pathways. Both genomes harbor modular and iterative type I PKS pathways as well as type II, type III and enediyne PKS pathways

Polyketides assembled in a manner similar to the type I fatty acid synthases (FAS) of fungi and animals are known as type I PKS pathways. At present, there exist iterative type I PKS pathways and modular type I PKS pathways with either cis-ATs or trans-ATs (Figure 2.2). Iterative type I PKS pathways consist of a single protein (also known as a module) capable of coordinating multiple acyl-CoA condensation reactions. The single protein contains multiple catalytic domains that perform the appropriate enzymatic activities in an iterative fashion to produce a final product. Modular type I PKS pathways with either cis-ATs or trans-ATs are comprised of multiple proteins, also known as modules, and each module is responsible for a single acyl-CoA condensation reaction (10). The most well-known and well-studied group of type I PKS pathways are the modular type I PKSs with cis-ATs. These PKS pathways, responsible for the production of biomedically relevant compounds such the antibiotic erythromycin and the anthelmintic avermectin, are composed of multiple modules, each with the complete set of catalytic domains necessary to coordinate an acyl-CoA condensation reaction (a minimal module has the KS, AT and ACP domains). For type I PKS pathways with cis-ATs, the number of modules present generally corresponds to the number of acyl units combined to create the final product (42). The close correlation has become known as the colinearity rule and has enabled polyketide structures to be predicted from genomic sequences and vice versa (35). Recently, an alternative modular type I PKS pathway architecture has been described where the AT domains are absent from each module. The absence of an AT in each module is complimented by an AT domain outside of the pathway. Although investigations of these modular type I PKS pathways with trans-ATs are in their

infancy, significant differences between the trans-AT and cis-AT PKS architecture and natural product biosynthesis are already being discussed. Interestingly, trans-AT PKSs have been identified in proteobacteria, myxobacteria and bacilli whereas cis-AT PKSs are associated with actinomycetes, myxobacteria and cyanobacteria (35). The asymmetric distribution has been hypothesized to be associated with biases in codon usage and differences in metabolism and lateral gene transfer mechanisms (35). Investigations of type I modular pathways offers insight into the biosynthetic mechanisms responsible for polyketide production and provides an opportunity to investigate the evolutionary history of these gene clusters.

In addition to a number of type I PKS pathways, both *Salinispora* genomes possess gene clusters associated with type II, type III and enediyne PKS derived polyketides. Like iterative type I PKS pathways, type II PKS pathways are also iterative. However, unlike type I PKS pathways, the catalytic domains of type II PKS pathways are each found on a separate protein. Type II PKS architecture is similar to that of the type II FASs of bacteria and plants. Only a single KS, a single AT and a single ACP are involved in the elongation process and while the growing acyl chain remains tethered to the single ACP domain, the other domains act iteratively. A chain length factor (CLF), which forms a heterodimer with the KS domain in type II PKS systems, has been suggested to determine the length of the final acyl chain; the volume of the polyketidyl binding pocket of the CLF may play a role in determining when the acyl chain is cleaved from the ACP (10). Type III PKS are distinguished from type I and type II PKS pathways in that they are relatively small dimeric proteins that carry out iterative condensation reactions with free CoA substrates such as malonyl-CoA.

Type III PKS pathways are involved in the biosynthesis of a variety of heterocyclic natural products and as more of these pathways have been discovered, it has been suggested that these type III pathways are able to utilize a broader range of starter and extender units than had been previously recognized (41). Enediyne PKS genes are highly conserved and analogous to iterative type I PKS pathways. These enediyne PKSs are believed to be responsible for the biosynthesis of the enediyne core unit or "warhead", a moiety characteristic of the family of antibiotics in which the enediyne acts as a potent DNA-damaging agent. Enediynes contain two acetylenic groups conjugated to a double bond or incipient double bond within a 9- or 10-membered ring and in the presence of DNA, the enediyne cycloaromatizes resulting in DNA cleavage (1, 28). Enediynes with the characteristic warhead structure include calicheamicin, esperamicin, dynemicin and neocarzinostatin. With the recent structure elucidation of the cyanosporasides and sporolides from "*S. pacifica*" and *S. tropica* strains respectively, it has been proposed that enediyne precursors biosynthesized by enediyne PKS pathways can undergo spontaneous Bergman cyclization to yield a final product (4, 36). As more genomes are sequenced and as more polyketide biosynthetic pathways are discovered, it is likely that the structural and functional diversity of polyketides will only be rivaled by the structural diversity and complexity of the pathways responsible for their production.

Genome sequences reveal entire biosynthetic pathways in the context of all other genes and pathways present within a single organism. With the entire, precise blueprint of an individual strain in hand, it is possible to begin to rigorously analyze the gene clusters that are responsible for the production of secondary metabolites and

ask questions related to the evolution of those gene clusters or even more simply, individual genes within a given cluster. The complete genome sequences of *S. arenicola* and *S. tropica*, offer an unprecedented opportunity to investigate the evolution of the secondary metabolite gene clusters of two closely related species. The present study investigates the type I modular PKS pathways of the *S. arenicola* genome. Investigations of these pathways in the context of the genome sequence and in relation to other sequenced genomes and gene clusters help highlight the genetic rearrangements that drive the evolution of secondary metabolite biosynthetic pathways. Analysis of the rifamycin biosynthetic gene cluster and the gene cluster for an as yet unidentified macrolide, has yielded evidence of gene duplication, gene loss, rearrangement, inversion and recombination.

Materials and Methods.

In 2007, the genome of *Salinispora arenicola* strain CNS-205 was sequenced and annotated by the Joint Genome Institute (JGI), in collaboration with the life sciences division of Oak Ridge National Laboratory (ORNL). From the *S. arenicola* genome, the nucleotide sequences of the two type I modular PKS biosynthetic pathways were obtained (nucleotides 1392183 through 1483610 and 3583210 through 3651608). From GenBank, the nucleotide sequence of the characterized rifamycin biosynthetic gene cluster of *Amycolatopsis mediterranei* (GenBank accession number AF040570) was also obtained. All three biosynthetic pathways were uploaded individually into Vector NTI version 7 (Invitrogen Co., Carlsbad, Calif.) and the open reading frames (ORFs) or genes designated by using the annotated JGI or GenBank

sequence files. Vector NTI was then used to compare and contrast the rifamycin biosynthetic pathways of *S. arenicola* and *A. mediterranei*. BLASTP analyses of the *A. mediterranei* genes against the *S. arenicola* genome (http://genome.ornl.gov/cgi-bin/Blast/blastform.cgi?blastorgs=sare:18jul07) were used to determine the amino acid sequence identity of homologous genes. For each gene within the two *S. arenicola* biosynthetic pathways, paralogs were identified and evidence of horizontal gene transfer (HGT) evaluated (personal communication K. Penn).

For each of the type I PKS genes of both *S. arenicola* clusters and the rifamycin biosynthetic pathway of *A. mediterranei*, the domain architecture of each module was determined. Using the specialized BLAST two sequences tool (bl2seq) available on the NCBI website (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi), the KS, AT, DH, ER, KR and ACP domains of the type I PKS modules were identified. Through sequential alignment with a well characterized domain from the fourth module of the type I modular PKS pathway responsible for the biosynthesis of erythromycin, the domains of each type I PKS pathway were defined. The domain architecture of the characterized rifamycin biosynthetic pathway corresponded to that published (2).

The KS, AT, DH, KR and ACP domains of the two *S. arenicola* type I PKS pathways and the rifamycin PKS pathways were aligned using ClustalX version 1.83 (46) and manually edited using MacClade version 4.07 (29). Phylogenetic trees were constructed using the distance neighbor-joining method in PAUP (43). One set of phylogenetic trees was constructed using aligned KS, AT, KR and ACP domain sequences from Sare3156, Sare3154, Sare3153, Sare3152 and Sare3151 while a

second set of phylogenetic trees was constructed using aligned KS, AT, DH, KR and ACP domain sequences from the characterized rifamycin biosynthetic pathway of *A. mediterranei* and the type I modular PKS pathway of *S. arenicola* predicted to biosynthesize a rifamycin-like natural product. For each phylogenetic tree, the appropriate domain from the fourth module of the type I PKS pathway involved in erythromycin biosynthesis was used to position the root.

Results and Discussion.

A bioinformatic analysis of the *Salinispora arenicola* genome revealed two type I modular PKS pathways. While one of these biosynthetic pathways is predicted to biosynthesize an as yet unknown macrolide, the other biosynthetic pathway is responsible for the production of a rifamycin compound. Investigations of these pathways in the context of the genome sequence and in relation to other sequenced genomes and gene clusters helped highlight the genetic rearrangements that drive the evolution of secondary metabolite biosynthetic pathways. By analyzing the gene clusters, evidence of gene duplication, gene loss, rearrangement, inversion and recombination were identified.

The gene cluster predicted to encode the biosynthesis of the unknown macrolide is predicted to include 16 genes, Sare3148 through Sare3163. Among these 16 genes are five type I modular PKS genes that encode a total of ten modules (Figure 2.3). All five genes were determined to be paralogous, genes predicted to have arisen within a single species through gene duplication events (personal communication K. Penn). The modules of type I PKS pathways have been hypothesized to evolve via gene

duplication followed by sequence divergence (18) and therefore it is not surprising that the five PKS genes encoding a modular type I PKS share high sequence and structural similarity and were suggested to be paralogous. The paralog designation however, belies the evolutionary history of these type I PKS genes and masks their uniqueness. The paralog designation refers only to the gene and without taking a closer look at the modular and domain architecture of the individual type I PKS genes, in the context of the pathway, how each module within each gene may have evolved and what each gene's specific contribution is to the overall structure of the final product cannot be fully appreciated.

The five PKS genes predicted to biosynthesize the acyl chain of an as yet unidentified macrolide are arranged on the reverse strand of the genome. The genes are arranged in order starting with Sare3156 and are only interrupted by a putative cytochrome P450 monooxygenase (Sare3155) between the first and second PKS genes (Figure 2.3). The first PKS gene, Sare3156, is comprised of five modules. The first or loading module is a minimal module with only a KS domain, an AT domain and an ACP domain. The second and third modules each have five domains, the KS, AT, DH, KR and ACP domains whereas the fourth and fifth modules have only four domains, the KS, AT, KR and ACP domains. Based on conserved amino acid sequences within the AT domain, the identity of the extender unit added at each module can be predicted. According to the AT domain sequence analysis, the third module of Sare3156 is specific for the addition of a methylmalonyl-CoA whereas the other four modules are predicted to extend the acyl chain using malonyl-CoA. The second PKS gene, Sare3154, is comprised of only a single module. However, this

module contains six domains, the KS, AT, DH, ER, KR and ACP domains. Sare3153, the third PKS gene, is composed of two modules, each with five domains, the KS, AT, DH, KR and ACP domains. Interestingly, the architecture of these two modules is virtually identical to the architecture of the second and third modules of Sare3156, the first PKS gene. In fact, according to the AT domain sequence analysis, the second module of Sare3153, like the third module of Sare3156, is specific for the addition of a methylmalonyl-CoA whereas the first module of Sare3153, like the module preceding the third module of Sare3156, is predicted to extend the acyl chain using malonyl-CoA. Genes Sare3152 and Sare3151 are each comprised of a single module. However, while Sare3151, like Sare3154, contains all six domains, the KS, AT, DH, ER, KR and ACP domains, Sare3152, like the final domain of Sare3156 contains only four domains, the KS, AT, KR and ACP domains.

An intriguing pattern emerges when the first two PKS genes, Sare3156 and Sare3154, are compared to the final three PKS genes, Sare3153, Sare3152 and Sare3151 (Figure 2.3). The architecture of the final four modules of the PKS pathway mirrors that of modules two, three, four or five and six. While the exact evolutionary history of the PKS modules cannot be reconstructed with absolute confidence, it is possible to formulate a hypothesis as to how this pathway may have evolved based on the emergent modular pattern and the PKSs evolutionary relatedness to type I FASs. The basic modular structure and architecture of type I FASs and type I PKSs are very similar and there is strong evidence to suggest that these pathways have shared a long joint evolutionary process (19, 37). Analysis of the present PKS modular architecture, in light of gene duplication and domain loss, could support a variety of paths by which

evolutionary events transpired, leading to the type I modular PKS pathway discovered within the *S. arenicola* genome. The following series of events however, offer a single hypothesis by which the type I modular PKS pathway responsible for the putative biosynthesis of an as yet unidentified macrolide could have evolved. This hypothesis is formulated solely by evaluating the domain architecture of the type I PKS modules and proposing the most parsimonious explanation for module duplication followed by sequence divergence.

The proposed evolution begins with Sare3151, a complete module, replete with all six domains and a thioesterase domain, used to terminate the elongation of an acyl chain. A gene duplication event of Sare3151 coupled with the loss of the thioesterase domain may account for the evolution of the loading module. Over time, selection pressure on the loading module may have led to the loss of all three optional domains. Unless there is a beta-carbonyl on the starting unit that could be reduced, optional domains present within the first PKS module would have no function and could be lost over time without altering the final product. Prior to losing all of the optional domains however, it is possible to propose that the module was duplicated after losing only the ER domain. The new module could have then duplicated twice and while in one of the subsequent modules a DH domain was lost, in the other module the AT domain evolved to extend the growing acyl chain with methylmalonyl-CoA instead of malonyl-CoA. The four modules downstream of the loading domain might then have all been duplicated creating a PKS pathway with nine modules. Only one additional duplication event would have had to occur and it is proposed that the newly duplicated module with only the KS, AT, KR and ACP domains was duplicated to yield the ten

modules. At some point just before or at some point after the final gene duplication event that yielded the ten modules, a cytochrome P450 gene is proposed to have been inserted into the pathway in between the first five modules and the final five modules. In order to look for evidence supporting the hypothetical evolutionary pathway, the amino acid sequences of the KS AT, KR and ACP domains were aligned and used to draw neighbor-joining phylogenetic trees (Figure 2.4). In theory, if a module arose through a gene duplication event, the domains of the duplicate module and the parent module would be more similar to one another than to any other domains of the PKS pathway. All four of the phylogenetic trees drawn indicate that the KS, AT, KR and ACP domains of the fourth and fifth modules share the highest sequence identity. The high sequence identity between modules four and five strongly suggests that they arose through a duplication event and represent the most recent duplication event in the PKS pathway. Strong additional evidence supporting the proposed evolutionary history of the PKS pathway is difficult to identify using the phylogenetic trees. For example, while the KR domains of Sare3154 and Sare3151 are most closely related (as would be expected given the proposed evolutionary pathway), phylogenetic analyses using the KS and AT domains suggest that the KS and AT domains of Sare3151 are most similar to the corresponding domains in Sare3152. While likely that the proposed evolutionary pathway is not entirely correct, the phylogenetic analyses make it clear that over time, one or more other forces are influencing the sequence of the PKS pathway. In addition to gene duplication, I propose that intermodular recombination of domains may play a significant role in the evolution of type I modular PKS pathways. At any time, the switching of one or more domains

between modules could provide the cohesive force that helps maintain a functional PKS pathway as well as provides a mechanism by which to create structural variation of the final secondary metabolite product.

While it is important to critically think about the evolutionary processes that influence the biosynthesis of secondary metabolites, it is also important to use an understanding of the modular architecture and enzymatic capabilities of each module in order to predict a partial structure of an as yet unelucidated secondary metabolite. With a partial structure in hand, it is possible to focus an investigation on targeting the isolation and elucidation of such a compound. By taking into consideration the number of modules present, the domains present in each module and by using the colinearity rule, it is possible to predict the partial structure of the product generated by PKS pathway. The unknown macrolide may include the partial structure illustrated in Figure 2.5.

Although the product of one of the type I modular PKS gene clusters is not yet known, the other type I modular PKS gene cluster present within the *S. arenicola* genome likely produces a rifamycin compound or analog. Rifamycins are a family of ansamycin antibiotics that have been used to treat tuberculosis, leprosy, and AIDS-related mycobacterium avium complex (MAC) infections (13). Rifamycins are effective against both Gram-positive and some Gram-negative bacteria due to their ability to inhibit DNA-dependent RNA synthesis (5). Rifamycins were first elucidated in 1959 (40) from a terrestrial actinomycete originally described as *Streptomyces mediterranei* (31). The *Streptomyces* species was later renamed *Nocardia mediterranei* (45) before being recognized as the first species of a new genus,

Amycolatopsis mediterranei, in 1986 (26). Since the discovery of the first rifamycins, numerous investigations have focused on understanding their biosynthesis (12). While many groups are interested in learning more about type I modular PKS pathways through the study of rifamycin biosynthesis, a number of other groups have been particularly interested in characterizing the enzymes responsible for the formation of rifamycin's unusual starter unit, 3-amino-5 hydroxybenzoic acid, also known as AHBA (14, 24). A large number of research groups have worked to better understand AHBA and rifamycin production and great strides have been made over the past decade since the sequencing of the roughly 100 kb *A. mediterranei* gene cluster responsible for both AHBA formation and rifamycin biosynthesis (2, 39, 44). The full sequence of the *A. mediterranei* rifamycin biosynthetic gene cluster revealed a type I modular PKS pathway with ten modules, genes encoding the eight enzymes responsible for AHBA formation and a large number of regulatory genes and genes encoding for post-translational modifications to the PKS product (12).

In 2007, Jensen *et al.* suggested that *S. arenicola* had the unique capacity to produce rifamycin relative to other Salinispora phylotypes following a PCR-based screen for the AHBA synthase gene, *rifK* (23). Their results indicated that the *rifK* gene was only present in *S. arenicola* and not in either of the other two *Salinispora* phylotypes. The *rifK* gene amplified was most closely related to the *rifK* gene of *A. mediterranei* as opposed to other homologous genes in more closely related actinomycete families. Chemical analyses of *S. arenicola* confirmed the presence of rifamycin-like compounds but complete structure elucidation of the rifamycins has yet to be completed. Genome sequencing of *S. arenicola* and *S. tropica* strains confirmed

that a gene cluster, with striking similarity to the rifamycin biosynthetic gene cluster of *A. mediterranei*, was present within the *S. arenicola* genome yet absent from the *S. tropica* genome. The similarity of the *S. arenicola* and *A. mediterranei* pathways led to the suggestion that the gene cluster had been subject to horizontal gene transfer (HGT). With the *S. arenicola* genome now available for further study, evidence of horizontal gene transfer could be sought. Additionally, using the *A. mediterranei* as a reference, the effect of other evolutionary forces on each biosynthetic gene clusters could be investigated.

The putative rifamycin biosynthetic gene cluster of *Salinispora arenicola* is proposed to contain 36 genes, Sare1242 through Sare1277, whereas the *Amycolatopsis mediterranei* biosynthetic gene cluster contains 43 genes (Table 2.2). Similar to the rifamycin biosynthetic pathway of *A. mediterranei*, the *S. arenicola* pathway can be divided into four regions (Figure 2.6). The first region is composed of Sare1242 through Sare1245. Within the first region, all four genes are homologous to the *rifS*, *rifT*, *orf35* and *orf0* genes respectively, of *A. mediterranei*. The second region, Sare1246 through Sare1251, encodes five type I modular PKS genes and a chain-terminating amide synthase. The five type I modular PKS genes share a very similar domain architecture with the characterized rifamycin biosynthetic gene cluster. Interestingly, the domain architecture of the second module of PKS gene Sare1246 is different from the domain architecture of the second rifamycin PKS module in *A. mediterranei*. In *S. arenicola* this encodes KS, AT, DH and ACP domains, whereas in *A. mediterranei* the DH domain is missing. The third region of the pathway is predicted to be primarily responsible for the formation of the AHBA starter unit.

While the original hypothesis proposed that four or five genes would be involved in AHBA synthesis (2), a total of eight genes were determined to be required for AHBA formation in *A. mediterranei*. All eight genes, *rifG* through *rifN*, were determined to be required and sufficient for AHBA production following heterologous expression of all eight genes in *Streptomyces coelicolor* (12). Again, an interesting difference between the two clusters is observed with respect to AHBA formation. Although all eight genes determined to be necessary for AHBA formation are present in the *S. arenicola* gene cluster, the organization of the genes in the two clusters is markedly different. All eight *Amycolatopsis mediterranei* genes, with the exception of *rifJ*, are clustered together following the type I modular PKS genes. The *rifJ* gene is located downstream, roughly 26 kb from the other seven genes. Within the *S. arenicola* gene cluster, five of the eight genes cluster together following the type I modular PKS genes while the other three are clustered together downstream, almost 20 kb from the other genes proposed to form AHBA (Table 2.2). Of the 18 genes that make up the fourth region of the *S. arenicola* cluster, all but four are homologous to genes present within the *A. mediterranei* rifamycin biosynthetic cluster. Like those genes described for *A. mediterranei*, the genes of the *S. arenicola* biosynthetic pathway are proposed to be a series of regulatory genes and genes encoding enzymes responsible for the post-synthetic modification of the natural product. It is among these regulatory and modifying enzymes that genomic modification events are readily observed (Figure 2.7). Comparison of the two gene clusters allows an inversion event to be recognized. Sare1263 and Sare1264 appear to be inverted when compared to genes *orf6* and *orf7* of the *A. mediterranei* rifamycin pathway. Several single gene rearrangements are

also apparent. Sare1259 and its homolog *orf13*, Sare1260 and its homolog *orf15b*, Sare1262 and its homolog *orf20*, Sare1274 and its homolog *rifI*, and Sare1275 and its homolog *rifN* are each positioned in a significantly different location relative to the other operon. Genes unique to each cluster are also identified when the two clusters are compared. While the *S. arenicola* cluster appears to have four genes that are not shared between the two gene clusters, the *A. mediterranei* cluster possesses eleven genes that are not shared between the two clusters. Of the four unique genes present within the *S. arenicola* gene cluster, three are homologous to genes found elsewhere in the *S. arenicola* genome suggesting they may have been involved in a gene duplication event. Of the eleven unique genes identified in the *A. mediterranei* gene cluster, nine are also homologous to *S. arenicola* genes found outside of its type I modular PKS pathway. The existence of homologs found amongst the unique genes is likely a reflection of their involvement in regulatory activities or in the modification of many different cellular metabolites. For example, *orf4* and *orf5* are unique to the *A. mediterranei* pathway yet encode cytochrome P450 monooxygenases. Cytochrome P450 monooxygenases are involved in a number of other described pathways and therefore it is not a surprise to have found homologs for *orf4* and *orf5* in *S. arenicola* or other actinomycetes. It is likely that the unique genes of each cluster arose through the duplication of a gene within the biosynthetic gene cluster or elsewhere in the respective genomes, an event that was subsequently followed by the translocation of the gene to a new position within the respective type I modular PKS gene cluster. Both *orf4* and *orf5* may have evolved through gene duplication events. Both genes share very high sequence identity with Sare1259 and yet *orf13* was determined to

share the highest sequence identity with Sare1259. It could be hypothesized that *orf4* and *orf5* are duplicates of *orf13*, which was later moved downstream during a rearrangement event. Although the true order of events that shaped both type I modular PKS gene clusters cannot be known, clear evidence of rearrangements, gene acquisitions and inversion is observed when the two pathways are compared.

According to analyses of individual genes for evidence of horizontal gene transfer (HGT), only three of the genes within the *S. arenicola* pathway suggest that HGT played a role in their acquisition. The three genes include Sare1244, Sare1248 and Sare1249 (personal communication C. Jenkins). Again, the fact that *S. arenicola* harbors a very large gene cluster with strikingly homology to the characterized rifamycin biosynthetic pathway, coupled with the fact that the closely related *S. tropica* does not contain any part of the pathway, led to a proposal that the entire gene cluster was acquired via HGT. Although evidence of horizontal gene transfer could only be identified for three of the 36 genes, the most parsimonious explanation as to how the *S. arenicola* rifamycin biosynthetic pathway was acquired is via horizontal gene transfer of the entire gene cluster. One explanation as to why very few of the S. arenicola rifamycin genes were flagged as genes potentially acquired by horizontal gene transfer, relates to one of the primary tests used to identify horizontally transferred genes. GC content is often used to help identify genes or gene clusters that differ significantly from the overall GC content of the organism. The lack of evidence supporting horizontal gene transfer of the *S. arenicola* rifamycin biosynthetic pathway likely stems from the acquisition of the pathway from an organism with similar GC content. *S. arenicola*, *A. mediterranei* and a large number of other actinomycetes have

a GC content around 70%. If *S. arenicola* were to have acquired the pathway from *Amycolatopsis mediterranei* for example, the GC content of the acquired pathway and the S. arenicola would be too similar and as a result the genes of the acquired pathway would not be flagged as HGT candidates. As the genomes of other actinomycetes known to produce rifamycin and other ansamycin compounds are sequenced, it is likely a more clear understanding of where and how this pathway originated and was passed among members of the same phylum will become available.

By comparing entire pathways it is possible to suggest whether genes within the pathway or the entire pathway itself may have been acquired horizontally. By comparing the orientation of genes within similar pathways, as was done for the *S. arenicola* and *A. mediterranei* pathways, it is possible to identify genetic modifications and gain insight into the forces driving secondary metabolite evolution. However, it is only upon closer investigation of the genes themselves that some of the smallest forces driving the evolution of secondary metabolite biosynthetic pathways can be observed. Within both the *S. arenicola* and *A. mediterranei* pathways there are five type I modular PKS genes and these five genes share a very similar but not identical domain architecture (Figure 2.8). The first of the type I PKS genes encodes three modules and interestingly, the second module of *S. arenicola* is different from the second module of the characterized rifamycin PKS pathway. The second module of *S. arenicola* encodes four domains, including KS, AT, DH and ACP domains, whereas the second *A. mediterranei* module contains only three domains, lacking the DH domain entirely. The significance of the subtle difference can be appreciated in a number of ways. First, the presence of an optional domain in the module of one but

not both PKS pathways could lead to the biosynthesis of two subtly different polyketides. A relatively small change in the modular architecture of a type I PKS could therefore result in the generation of structural diversity and the production of compound analogues. Although the presence of a DH domain in the absence of a KR domain generally does not reduce the beta-carbonyl, the *S. arenicola* module is one KR domain acquisition away from a pathway that might make a compound different from the predicted rifamycin-like product. Whether the DH domain was acquired in *S. arenicola* or lost in *A. mediterranei* cannot be determined at this time however it raises the question as to whether domains and perhaps modules, in addition to entire genes, can be recombined within the same organism or acquired from another organism in order to generate structurally diverse polyketides.

An awareness of the impact of recombination events on the evolution of type I modular PKS pathways has only just begun to emerge. Historically, the hypothesis with respect to type I PKS modules is that they arose by sequential gene duplication events. The modules of type I PKS pathways have been hypothesized to evolve via gene duplication followed by sequence divergence (18) and yet recently published studies are beginning to draw attention to examples suggesting exceptions to the long-standing hypothesis (15, 19). While some phylogenetic analyses of KS and AT domains strongly support the evolution of a pathway through gene duplication only, such as the oligomycin biosynthetic pathway of *Streptomyces avermitilis* (19), phylogenetic analyses of the KS domains from the rapamycin and ascomycin pathways suggest that horizontal gene transfer of individual modules may be involved in the evolution of secondary metabolites (15). Recombinational exchange of

domains, domain acquisition and domain loss are also likely to be particularly important in generating polyketide diversity and as more bacterial genomes are sequenced, it will likely become possible to better understand the role intra- and intergenomic recombination play in the evolution of bacterial type I modular PKS pathways.

In an attempt to detect intermodule domain recombination events within their respective PKS genes, phylogenetic analyses of the KS, AT, DH, KR and ACP domains from both *S. arenicola* and *A. mediterranei* were performed (Figure 2.9). If the assumption is made that the two type I modular PKS pathways were acquired horizontally and intermodule domain recombination was not a factor in the evolution of the pathways, each domain would be most closely related to the domain of the corresponding module in the other pathway. In affect, the domains of each pathway would be interleaved. If intermodule domain recombination does occur within and between PKS genes and has influenced each pathway since they shared a common ancestor, then domains would not be perfectly interleaved but rather more closely related to another domain from their own pathway or another type I modular PKS pathways present within the given organism. By comparing the various domains of the two PKS pathways, it is possible to suggest that recombination does occur but only for specific domains. Intermodule domain recombination does not appear to occur for the KS domains of the *S. arenicola* and *A. mediterranei* pathways as they are perfectly interleaved. Each of the KS domains was most closely related to the KS domain of the corresponding module in the other pathway. While no recombination of KS domains appears to occur, AT domains that encode for the addition of malonyl-CoA to the

growing acyl chain appear to recombine frequently. For each pathway, all AT domains responsible for selecting a malonyl-CoA extender unit were clustered together. The phylogenetic analysis suggests that three of the *S. arenicola* AT domains underwent recent recombination as their sequences were almost identical. On the other hand, the AT domains of both modules two and nine, both responsible for selecting methylmalonyl-CoA extender units, were interleaved. This pattern suggests that these domains have not experienced recombination since the acquisition of the pathway. Analyses of the DH, KR and ACP domains suggest that these domains are also subject to intermodular recombination, however not to the extent of the recombination amongst AT domains. In addition to recombination, domain acquisition may also be playing a role in the evolution of the biosynthetic pathways. A DH domain is present in module two of the *S. arenicola* pathway but not in the second module of the characterized rifamycin PKS pathway. The DH domain is unique to the *S. arenicola* genome, not having any phylogentic relatedness to any other DH domain present within the *S. arenicola* genome. The presence of a unique domain suggests that small, functional pieces of DNA may be acquired from other microorganisms with similar pathways. Although this particular DH domain did not BLAST to any previously sequenced DH domain, future sequencing efforts may reveal which PKS pathway this domain was picked up from.

Gene duplication, intermodule domain recombination, domain acquisition and domain loss are all likely to influence the evolution of type I modular PKS pathways. Gene duplication is likely to play the most significant role in the generation of new modules. As genes and or modules are duplicated and as sequences diverge,

intermodule domain recombination likely helps maintain the functional integrity of the pathway. Recombination is also likely to result in the creation of diverse new natural products. The evolution of type I modular PKS pathways of filamentous actinomycetes, such *Salinispora*, *Amycolatopsis* and *Streptomyces*, is likely to be influenced frequently by recombination and as additional sequence data becomes available and as homologous pathways from different families are discovered, further insight into the roles these various forces play on the generation of polyketide diversity will be realized.

A genome sequence provides a snapshot of a microorganism. Genome sequences capture secondary metabolite pathways as they are at this moment and by comparing genomes and pathways and by looking closely at the modules and the domain architecture of type I PKS pathways, predictions regarding the evolutionary forces affecting secondary metabolite biosynthetic pathways can be made (11). By analyzing the type I modular PKS pathways of *S. arenicola*, evidence of gene duplication, gene loss, rearrangement, inversion, domain acquisition and recombination were identified.

References.

1. **Ahlert, J., E. Shepard, N. Lomovskaya, E. Zazopoulos, A. Staffa, B. O. Bachmann, K. Huang, L. Fonstein, A. Czisny, R. E. Whitman, C. M. Farnet and J. S. Thorson.** 2002. The calicheamicin gene cluster and its iterative type I enediyne PKS. Science. 297:1173-1176.

2. **August, P. R., L. Tang, Y.J. Yoon, S. Ning, R. Müller, T.-W. Yu, M. Taylor, D. Hoffmann, C.-G. Kim, X. Zhang, C.R. Hutchinson, and H.G. Floss.** 1998. Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the rif biosynthetic gene cluster of *Amycolatopsis mediterranei*. Chem. Biol. 5(2):69-79.

3. **Bentley, S. D., K. F. Chater, A.-M. Cerdeño-Tárraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C.-H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M.-A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill, and D. A. Hopwood.** 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature. 417:141-147.

4. **Buchanan, G. O., P. G. Williams, R. H. Feling, C. A. Kauffman, P. R. Jensen, and W. Fenical.** 2005. Sporolides A and B: structurally unprecedented halogenated macrolides from the marine actinomycete *Salinispora tropica*. Org. Lett. 7(13):2731-2734.

5. **Calvori, C. L. Frontali, L. Leoni and G. Tecce.** 1965. The effect of rifamycin on protein synthesis. Nature. 207:417.

6. **Cane, D. E., C. T. Walsh, and C. Khosla.** 1998. Harnessing the biosynthetic code: combinations, permutations, and mutations. Science. 282:63-68.

7. **Coleman, M. L., M. B. Sullivan, A. C. Martiny, C. Steglich, K. Barry, E. F. DeLong and S. W. Chisholm.** 2006. Genomic Islands and the Ecology and Evolution of *Prochlorococcus*. Science. 311(5768):1768-1770.

8. **Dobrindt, U., B. Hochhut, U. Hentschel and J. Hacker.** 2004. Genomic islands in pathogenid and environmental microorganisms. Nature Reviews Microbiology. 2:414-424.

9. **Feling R. H., G. O. Buchanan, T. J. Mincer, C. A Kauffman, P. R. Jensen and W. Fenical.** 2003. Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora*. Angew. Chem. Int. Ed. Engl. 42(3):355-7.

10. **Fischbach, M. A. and C. T. Walsh.** 2006. Assembly-Line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem. Rev. 106:3468-3496.

11. **Fischbach, M. A., C. T. Walsh and J. Clardy.** 2008. The eveolution of gene collectives: how natural selection drives chemical innovation. Proc. Natl. Acad. Sci. U.S.A. 105(12):4601-4608.

12. **Floss, G. H. and T.-W. Yu.** 1999. Lessons from the rifamycin biosynthetic gene cluster. Curr. Opin. Chem. Biol. 3:592-597.

13. **Floss, G. H. and T.-W. Yu.** 2005. Rifamycin: mode of action, resistance, and biosynthesis. Chem. Rev. 105:621-632.

14. **Ghisalba, O and J. Nüesch.** 198A genetic approach to the biosynthesis of the rifamycin chromophore in Nocardia mediterranei. IV. Identification of 3-amino-5-hydroxybenzoic acid as a direct precursor of the seven carbon amino starter unit. J. Antibiot. 34:64-71.

15. **Ginolhac, A., C. Jarrin, P. Robe, G. Perriere, T. M. Vogel, P. Simonet, and R. Nalin.** 2005. Type I polyketide synthases may have evolved through horizontal gene transfer. J. Mol. Evol. 60:716-725.

16. **Hacker, J., G. Blum-Oehler, I. Muhldorfer, and H. Tschape.** 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. Mol. Microbiol. 23:1089-1097

17. **Hentschel, U. and J. Hacker.** 2001. Pathogenicity islands: the tip of the iceberg. Microbes. Infect. 3(7):545-548.

18. **Hopwood, D. A.** 1997. Genetic contributions to understanding polyketide synthases. Chem. Rev. 97:2465-2497.

19. **Jenke-Kodama, H., A. Sandmann, R. Muller, and E. Dittmann.** 2005. Evolutionary implications of bacterial polyketide synthases. Mol. Biol. Evol. 22(10):2027-2039.

20. **Jensen, P. R., R. Dwight and W. Fenical.** 1991. Distribution of Actinomycetes in Near-Shore Tropical Marine Sediments. Appl. Environ. Microbiol. 57(4):1102-1108.

21. **Jensen, P. R. and W. Fenical.** 1995. The Relative Abundance and Seawater Requirements of Gram-Positive Bacteria in Near-Shore Tropical Marine Sediments. Microbial Ecology. 29:249-257.

22. **Jensen, P. R. and C. Mafnas.** 2006. Biogeography of the actinomycete genus *Salinispora*. Environ. Microbiol. 8:1881-1888.

23. **Jensen, P. R., P. G. Williams, D.-C. Oh, L. Zeigler, and W. Fenical.** 2007. Species-specific secondary metabolite production in marine actinomycetes of the genus *Salinispora*. Appl. Environ. Microbiol. 73(4):1146-1152.

24. **Kibby, J. J., I. A. McDonald and R. W. Rickards.** 1980. 3-amino-5-hydroxybenzoic acid as a key intermediate in ansamycin and maytansinoid biosynthesis. J. Chem. Soc. Chem. Commun. 1980:768-769.

25. **Lautru, S., R. J. Deeth, L. M. Bailey and G. L. Challis.** 2005. Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. Nature Chemical Biology. 1(5):265-269.

26. **Lechevalier, M. P., H. Prauser, D. P. Labeda and J.-S. Ruan.** 1986. Two new genera of nocardioform actinomycetes: *Amycolata* gen. nov. and *Amycolatopsis* gen. nov. Int. J. Syst. Bacteriol. 36:29-37.

27. **Liang, X., X.-Q. T. Pham, M. V. Olson and S. Lory.** 2001. Identification of a Genomic Island Present in the Majority of Pathogenic Isolates of *Pseudomonas aeruginosa*. J. Bact. 183(3):843-853.

28. **Liu, W., J. Ahlert, Q. Gao, E. Wendt-Pienkowski, B. Shen and J. S. Thorson.** 2003. Rapid PCR amplification of minimal enediyne polyketide synthase cassettes leads to a predictive familial classification model. Proc. Natl. Acad. Sci. U.S.A. 100(21):11959-11963.

29. **Maddison, D. R. and W. P. Maddison.** 2001. Macclade 4. Sunderland, MA, USA: Sinauer Associates.

30. **Maldonado, L. A., W. Fenical, P. R. Jensen, C. A. Kauffman, T. J. Mincer, A. C. Ward, A. T. Bull and M. Goodfellow.** 2005. *Salinispora arenicola* gen. nov., sp. nov. and *Salinispora tropica* sp. nov., obligate marine actinomycetes belonging to the family *Micromonosporaceae*. Int. J. Syst. Evol. Microbiol. 55:1759-1766.

31. **Margalith, P. and G. Beretta.** 1960. Rifomycin. XI. Taxonomic study on *Streptomyces mediterranei* nov. sp. Mycopathol. Mycol. Appl. 12:189–195.

32. **Mincer, T. J., P. R Jensen, C. A. Kauffman and W. Fenical.** 2002. Widespread and persistent populations of a major new marine actinomycete taxon in ocean sediments. Appl. Environ. Microbiol. 68(10):5005-5011.

33. **Newman, D. J., G. M. Cragg and K. M. Snader.** 2000. The influence of natural products upon drug discovery. Nat. Prod. Rep. 17:215-234.

34. **Newman, D. J., G. M. Cragg and K. M. Snader.** 2003. Natural products as sources of new drugs over the period 1981-2002. J. Nat. Prod. 66:1022-1037.

35. **Nguyen, T., K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertwick and J. Piel.** 2008. Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. Nature Biotech. 26(2):225-233.

36. **Oh, D.-C., P. G. Williams, C. A. Kauffman, P. R. Jensen, and W. Fenical.** 2006. Cyanosporosides A and B, unprecedented cyclopenta[a]indene glycosides

from the obligate marine actinomycete *Salinispora pacifica*. Org. Lett. 8(6):1021-1024.

37. **Ridley C. P., H. Y. Lee and C. Khosla.** 2008. Evolution of polyketide synthases in bateria. Proc. Natl. Acad. Sci. U.S.A. 105(12):4595-4600.

38. **Ruiz, S., Y. Krupnik, M. Keating, J. Chandra, M. Palladino and D. McConkey.** 2006. The proteasome inhibitor NPI-0052 is a more effective inducer of apoptosis than bortezomib in lymphocytes from patients with chronic lymphocytic leukemia. Mol. Cancer Ther. 5(7):1836-43.

39. **Schupp, T., C. Toupet, N. Engel and S. Goff.** 1998. Cloning and sequence analysis of the putative rifamycin polyketide synthase gene cluster from *Amycolatopsis mediterranei*. FEMS Microbiol. Lett. 159:201-207.

40. **Sensi, P., P. Margalith and M. T. Timbal.** 1959. Rifamycin, a new antibiotic: preliminary report. Farmaco. Ed. Sci. 14:146-147.

41. **Song, L., F. Barona-Gomez, C. Corre, L. Xiang, D. W. Udwary, M. B. Austin, J. P. Noel, B. S. Moore and G. L. Challis.** 2006. Type III polyketide synthase beta-ketoacyl-ACP starter unit and ethylmalonyl-CoA extender unit selectivity discovered by *Streptomyces coelicolor* genome mining. J. Am. Chem. Soc. 128:14754–14755.

42. **Staunton, J. and K. J. Weissman.** 2001. Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. 18:380-416.

43. **Swofford, D. L**. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Software version 4.0b 10. Sinauer Associates, Sunderland, MA.

44. **Tang, L., Y. J. Yoon, C. Y. Choi and C. R. Hutchinson.** 1998. Characterization of the enzymatic domains in the modular polyketide synthase involved in rifamycin B biosynthesis in *Amycolatopsis mediterranei*. Gene. 216:255-265.

45. **Thiemann J. E., G. Zucco G and G. Pelizza.** 1969. A proposal for the transfer of *Streptomyces mediterranei* Margalith and Beretta 1960 to the genus *Nocardia* as *Nocardia mediterranea* (Margalith and Beretta) comb. nov. Arch. Mikrobiol. 67(2):147–155.

46. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876-4882.

47. **Udwary, D. W., L. Zeigler, R. N. Asolkar, V. Singan, A. Lapidus, W. Fenical, P. R. Jensen and B. S. Moore.** 2007. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. Proc. Natl. Acad. Sci. U.S.A. 104(25): 10376-10381.

48. **Udwary, D. W., L. Zeigler, R. N. Asolkar, V. Singan, A. Lapidus, W. Fenical, P. R. Jensen and B. S. Moore.** In preparation. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora arenicola*.

49. **Walsh, C. T.** 2004. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. Science. 303:1805-1810.

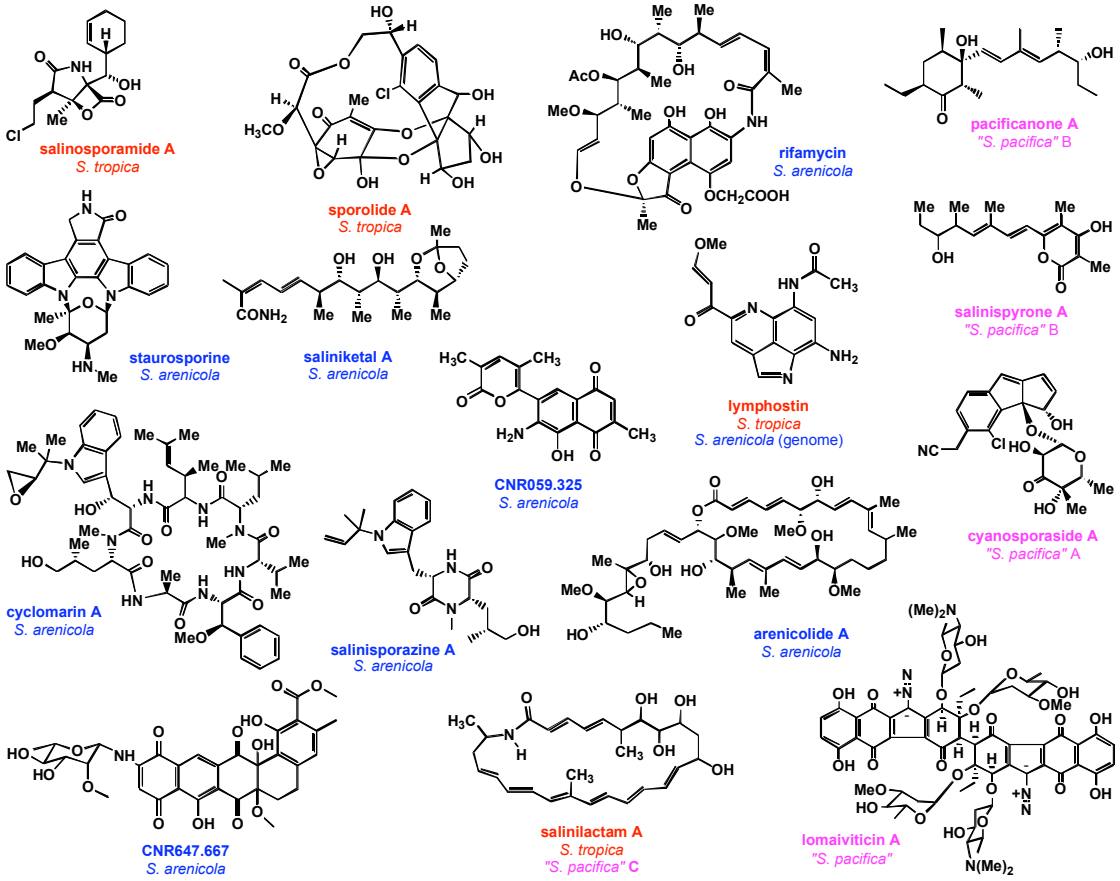

Figure 2.1: Structures of representative natural products biosynthesized by *Salinispora arenicola*, *Salinispora tropica* and "*Salinispora pacifica*". Figure used with permission by P.R. Jensen.

Table 2.1: List of the secondary metabolite biosynthetic gene clusters from the *Salinispora tropica* and *Salinispora arenicola* genomes (47, 48). The gene cluster location, type, size, actual or predicted product and whether the pathway is shared by the two genomes is given.

Salinispora tropica

Gene Cluster Location	Type	Size (kb)	Actual** or Predicted Secondary Metabolite	In Both Genomes
stro0586-0610	enediyne PKS	30	10-membered enediyne	no
stro0672-0699	NRPS	31	unknown dipeptide	no
stro1012-1043	PKS/NRPS	41	salinosporamide**	no
stro2204-2235	type II PKS	35	glycosylated decaketide	no
stro2340-2346	aminocyclitol	8	unknown aminocyclitol	yes
stro2486-2509	type II PKS	23	spore pigment polyketide	yes
stro2539-2554	NRPS	18	desferrioxamine**	yes
stro2636-2660	PKS/NRPS	28	yersiniabactin-like siderophore	yes
stro2695-2733	enediyne PKS	50	sporolide**	no
stro2757-2781	type I PKS	80	salinilactam**	no
stro2786-2813	PKS/NRPS	30	unknown siderophore	no
stro2814-2842	NRPS	44	coelibactin-like siderophore	no
stro3042-3066	PKS/NRPS	33	lymphostin**	yes
stro3244-3253	terpene	10	unknown terpene	no
stro4262-4271	type III PKS	10	aromatic polyketide	yes
stro4409-4429	NRPS	33	reductively offloaded tetrapeptide	yes
stro4437-4445	terpene	12	unknown carotenoid	yes

Salinispora arenicola

Gene Cluster Location	Type	Size (kb)	Actual** or Predicted Secondary Metabolite	In Both Genomes
sare0348-0368	NRPS	44	unknown pentapeptide	no
sare0474-0506	PKS/NRPS	34	unknown PKS/NRPS	no
sare0538-0565	enediyne PKS	34	9-membered enediyne	no
sare1036-1068	type I PKS	38	unknown PKS	no
sare1232-1278	type I PKS	99	rifamycin	no
sare1887-1896	NRPS	11	unknown NRPS	no
sare2017-2060	type I PKS	51	calicheamicin-like enediyne	no
sare2063-2107	PKS/NRPS	68	yersiniabactin-like siderophore	yes
sare2108-2126	enediyne PKS	21	calicheamicin-like enediyne	no
sare2254-2211	enediyne PKS	51	calicheamicin-like enediyne	no
sare2327-2340	indolocarb	18	staurosporine**	no
sare2399-2414	PKS/NRPS	25	unknown PKS/NRPS	no
sare2483-2491	aminocyclitol	25	unknown aminocyclitol	yes
sare2664-2694	type II PKS	30	spore pigment polyketide	yes
sare2728-2763	NRPS	18	desferrioxamine**	yes
sare2797-2838	type I PKS	47	unknown PKS	no
sare2939-2968	NRPS	47	unknown tetrapeptide	no
sare3054-3065	PKS/NRPS	13	unknown PKS/NRPS	no
sare3142-3164	type I PKS	78	unknown macrolide	no
sare3268-3293	PKS/NRPS	43	lymphostin	yes
sare4547-4579	NRPS	60	cyclomarin** and salinisporazine**	no
sare4693-4700	type III PKS	8	aromatic polyketide	yes
sare4879-4905	NRPS	40	reductively offloaded tetrapeptide	yes
sare4927-4935	terpene	12	unknown carotenoid	yes
sare4936-4963	type I PKS	33	aromatic polyketide	no

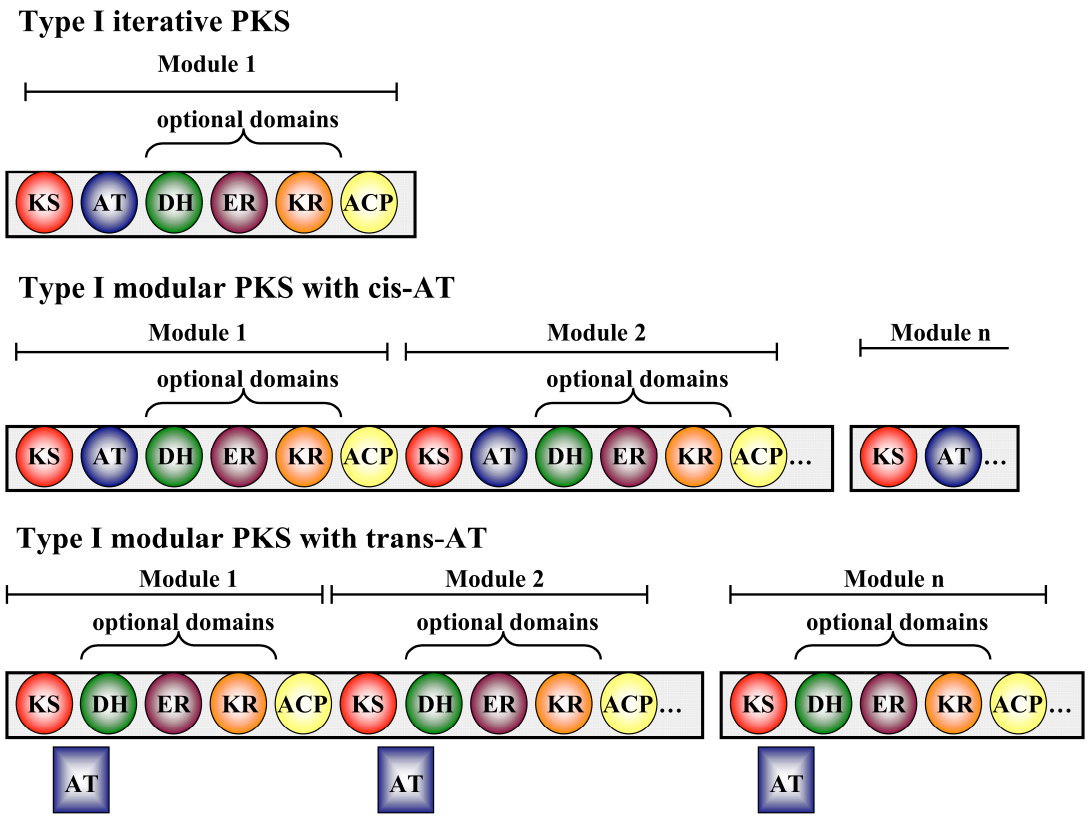

Figure 2.2: Schematic representation of iterative and modular type I polyketide biosynthesis. Distinct proteins are presented as rectangles whereas domains present within a single protein are depicted as circles. Optional domains of type I PKS pathways are indicated.

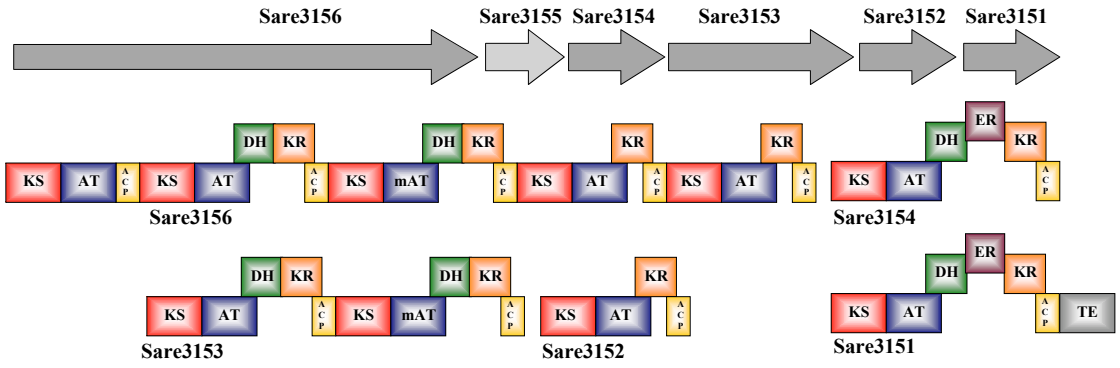

Figure 2.3: Genetic organization of six genes predicted to be involved with the biosynthesis of an as yet unidentified macrolide, including the five type I PKS genes (solid arrows). The five PKS genes and their domains are arranged below the arrow diagram to help illustrate a proposed evolutionary history involving several gene duplication events that may have led to the PKS pathway with ten modules. Sare3155 is proposed to function as a cytochrome P450.

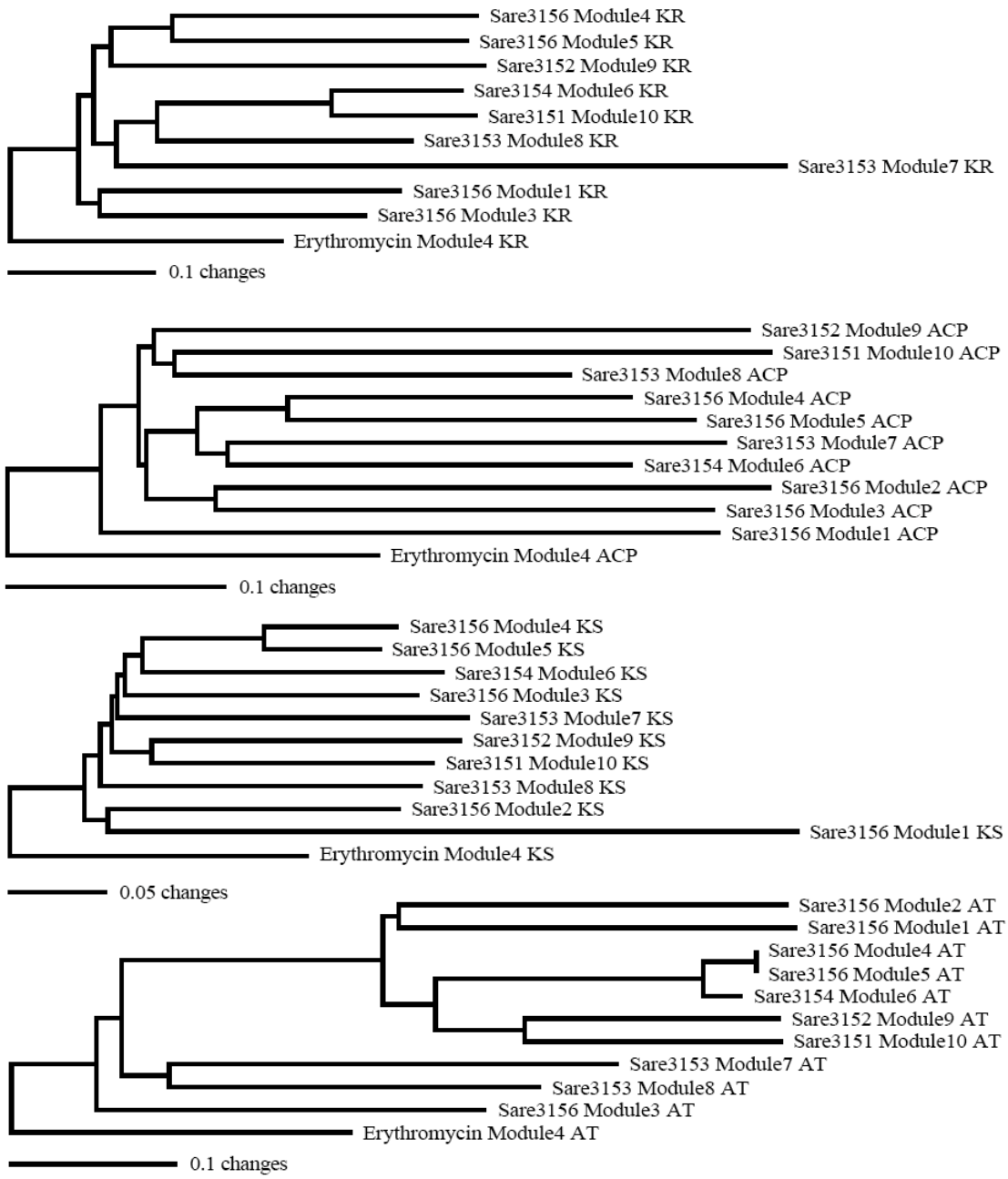

Figure 2.4: Neighbor-joining distance trees constructed in PAUP (43) using the aligned KS, AT, KR and ACP domain sequences from Sare3156, Sare3154, Sare3153, Sare3152 and Sare3151. For each phylogenetic tree, the appropriate domain from the fourth module of the type I PKS pathway involved in erythromycin biosynthesis was used to position the root.

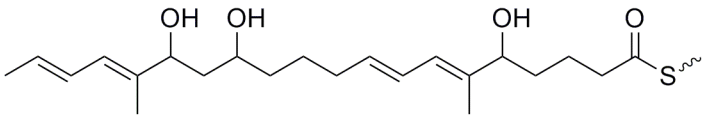

Figure 2.5: Predicted partial structure of an as yet unidentified macrolide from *Salinispora arenicola*. By taking into consideration the number of modules present, the domains present in each module and by using the colinearity rule, it is possible to predict a partial structure.

Table 2.2: List of the *Salinispora arenicola* genes predicted to be involved with the biosynthesis of a type I, rifamycin-like polyketide. For each gene, the proposed function and the amino acid identity with the homologous *Amycolatopsis mediterranei* rifamycin gene are given. Genes in bold are shared between the two pathways and when marked with an asterisk, the shared genes do not occupy the same position within their respective operon.

Proposed Function	*S. arenicola* Gene	*A. mediterranei* Homolog	Amino Acid Identity (%)
oxidoreductase domain protein	**Sare1242**	**rifS**	75%
oxidoreductase domain protein	**Sare1243**	**rifT**	56%
Hypothetical Protein	**Sare1244**	**orf35**	59%
Cytochrome P450 Monooxygenase	**Sare1245**	**orf0**	82%
Type I Modular PKS	**Sare1246**	**rifA**	72%
Type I Modular PKS	**Sare1247**	**rifB**	73%
Type I Modular PKS	**Sare1248**	**rifC**	71%
Type I Modular PKS	**Sare1249**	**rifD**	69%
Type I Modular PKS	**Sare1250**	**rifE**	72%
N-Acetyltransferase	**Sare1251**	**rifF**	69%
Hypothetical Protein	Sare1252	orf1	
3-Dehydroquinate Synthase	**Sare1253**	**rifG**	79%
3-Deoxy-7-Phosphoheptulonate Synthase	**Sare1254**	**rifH**	66%
Glutamine--Scyllo-Inositol Transaminase	**Sare1255**	**rifK**	86%
Oxidoreductase Domain Protein	**Sare1256**	**rifL**	79%
AHBA Synthesis Associated Protein	**Sare1257**	**rifM**	83%
Rieske [2Fe-2S] Domain Protein	Sare1258		
Putative Regulatory Gene		rifO	
Esterase		orf2	
Antibiotic Effux/Resistance		rifP	
Transcriptional Repressor		rifQ	
Protein of Unknown Function		orf3	
Cytochrome P450 Monooxygenase		orf4	
Cytochrome P450 Monooxygenase		orf5	
Cytochrome P450 Monooxygenase	**Sare1259***	**orf13***	79%
Cytochrome P450 Monooxygenase	**Sare1260***	**orf16***	73%
Protein of Unknown Function DUF1271	Sare1261		
Conserved Polyketide Synthase Associated Protein PapA5	**Sare1262***	**orf20***	59%
Protein of Unknown Function DUF1205	**Sare1263***	**orf7***	66%
DegT/DnrJ/EryC1/StrS Aminotransferase	**Sare1264***	**orf6***	86%
dNTP-Hexose 3,5 Epimerase		orf8	
Aminotransferase Class-III	**Sare1265**	**orf9**	78%
Oxidoreductase Domain Protein	**Sare1266**	**orf10**	70%
Reductase		orf11	
Putative Alpha-Chain Alkanal Monooxygenase		orf17	
NDP-Hexose 23-Dehydratase	**Sare1267**	**orf18**	79%
Monooxygenase FAD-Binding	**Sare1268**	**orf19**	74%
Oleoyl-[Acyl-Carrier-Protein] Hydrolase	**Sare1269**	**rifR**	70%
Regulatory Protein LuxR	**Sare1270***	**orf36***	64%
Methyltransferase Type 11	**Sare1271**	**orf14**	66%
Transketolase Domain Protein	**Sare1272**	**orf15a**	80%
Transketolase Central Region	**Sare1273**	**orf15b**	70%
Shikimate Dehydrogenase Substrate Binding Domain Protein	**Sare1274***	**rifI***	55%
ROK Family Protein	**Sare1275***	**rifN***	53%
DoxX Family Protein	Sare1276		
3-Dehydroquinate Dehydratase	**Sare1277**	**rifJ**	78%

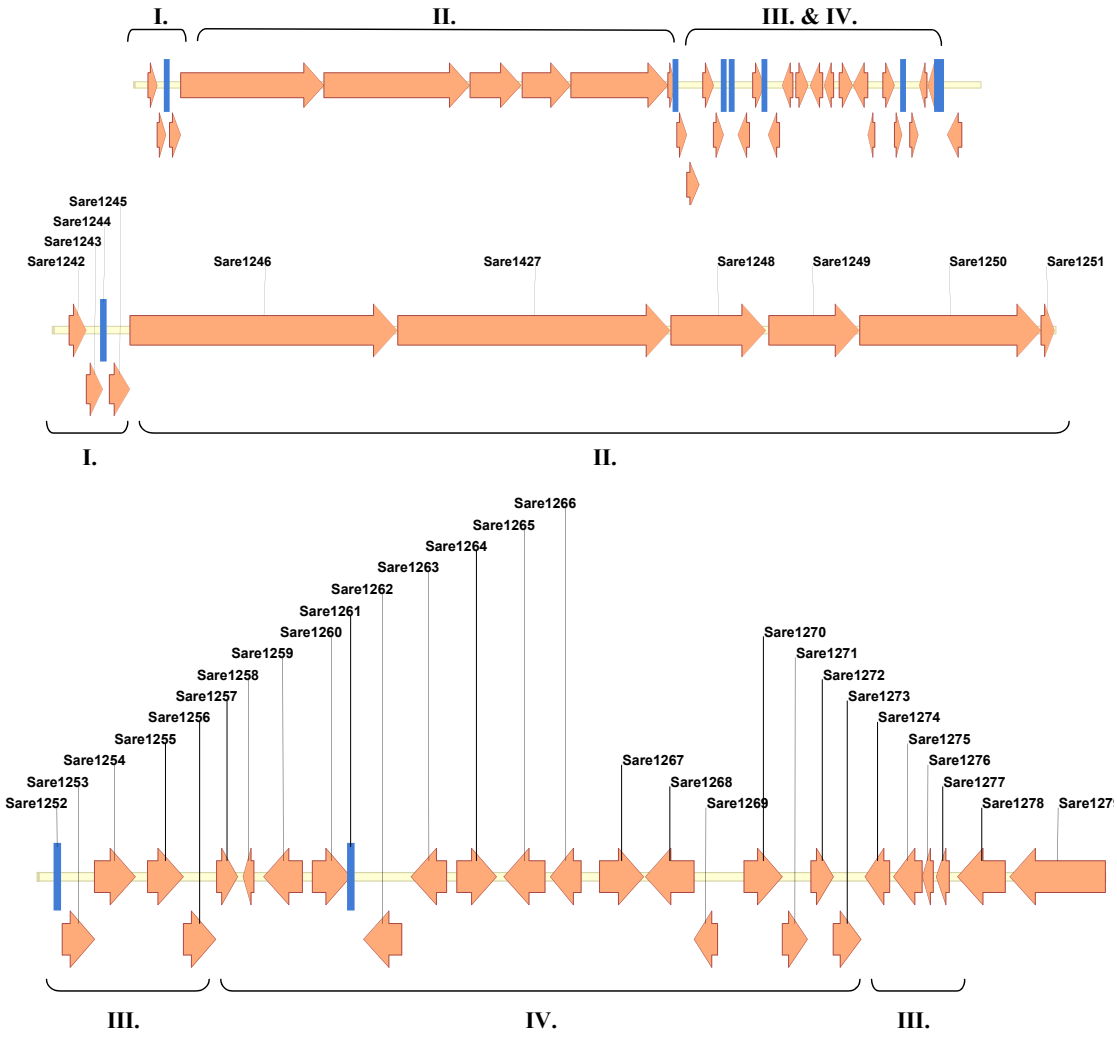

Figure 2.6: The biosynthetic gene cluster from *Salinispora arenicola* predicted to produce the rifamycin natural product isolated from this species. The gene cluster has been divided into four regions (I-IV) analogous to the four regions identified for the rifamycin biosynthetic gene cluster of *Amycolatopsis mediterranei*. Region I is predicted to encode modifying enzymes; region II encodes type I PKS genes including ten modules and a chain-terminating amide synthase; region III includes genes predicted to be responsible for the formation of the AHBA starter unit; and region IV contains genes that are predicted to be regulatory genes or genes that post-synthetically modify the polyketide.

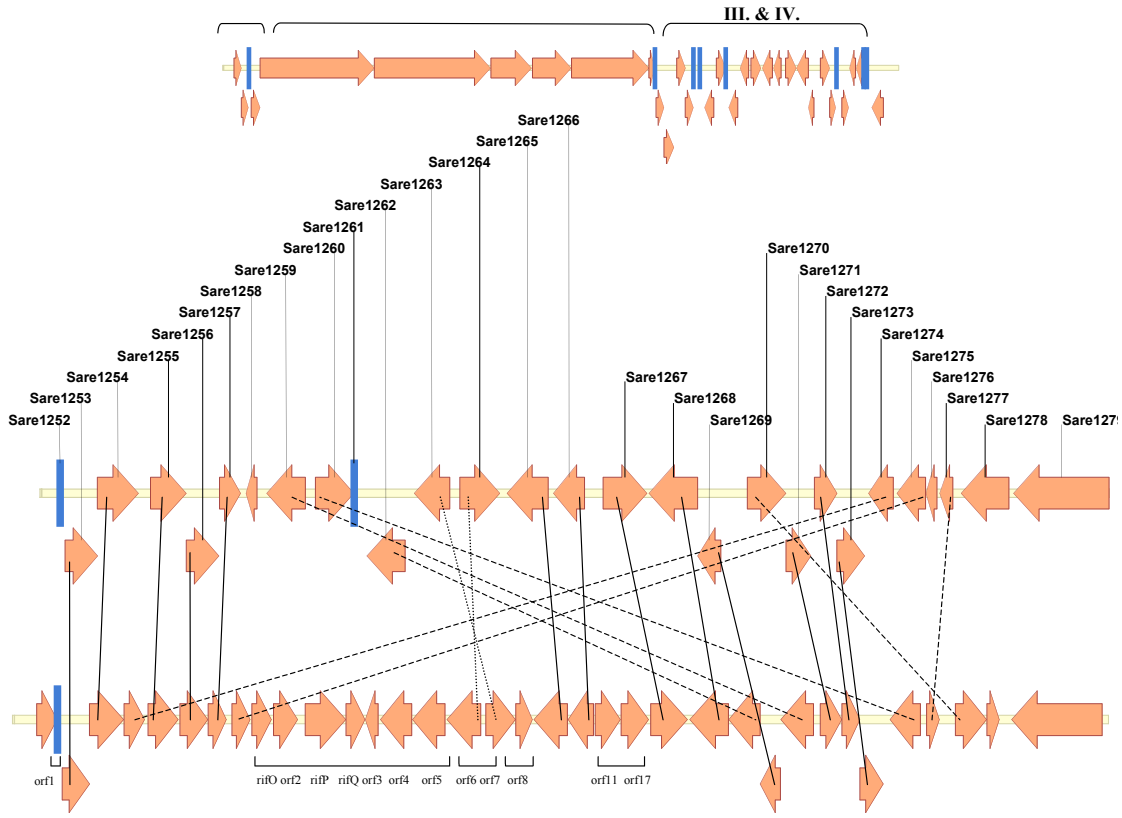

Figure 2.7: The genetic organization of the genes present in regions III and IV. Genes Sare1252 through Sare1277 are predicted to be involved in the formation of the AHBA starter unit and are predicted to be regulatory genes or genes that post-synthetically modify a rifamycin-like natural product. The *Salinispora arenicola* genes are aligned with the homologous region of the *Amycolatopsis mediterranei*. Solid lines are used to illustrate gene synteny, dotted lines are used to illustrate predicted gene inversions and dashed lines are used to depict genes that appear to have been rearranged and no longer share the same relative position on their respective operons. The eleven genes present within the *Amycolatopsis mediterranei* operon and not present within the *Salinispora arenicola* operon are labeled.

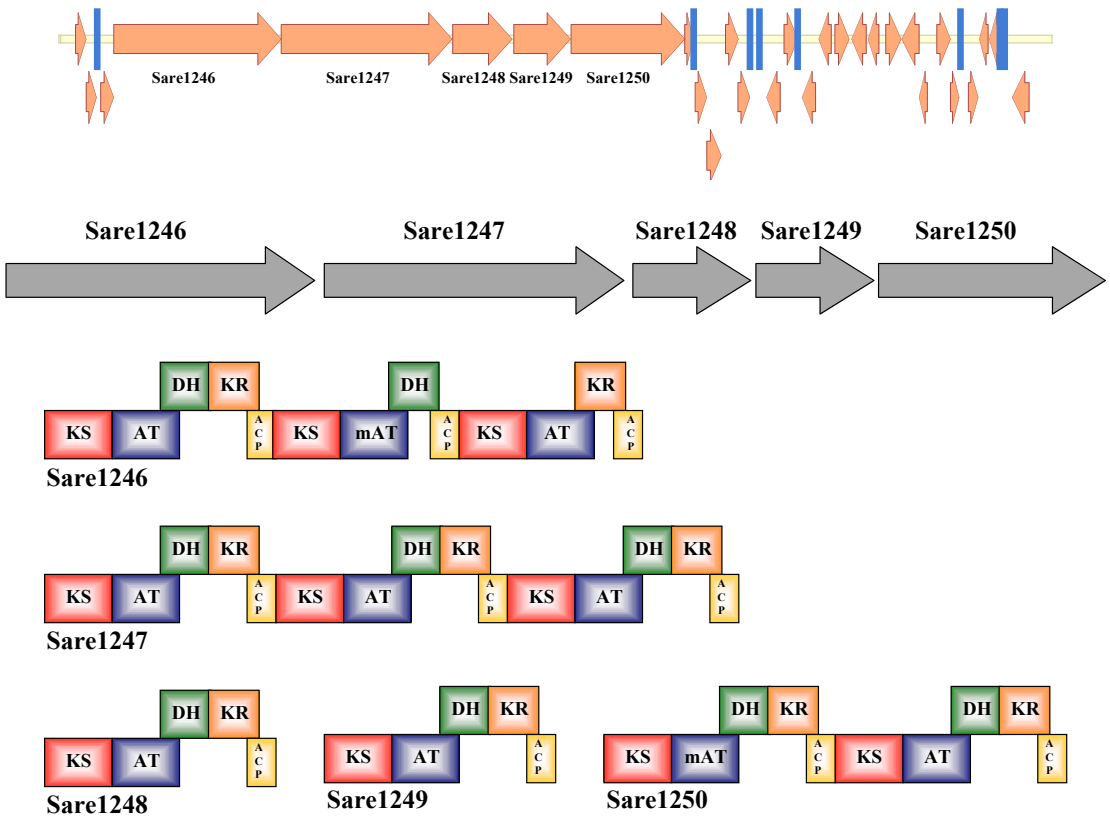

Figure 2.8: Genetic organization of the five type I PKS genes (Sare1246-Sare1250) predicted to be involved with the biosynthesis of a rifamycin-like natural product. The five PKS genes and their domain architecture are arranged below the arrow diagram.

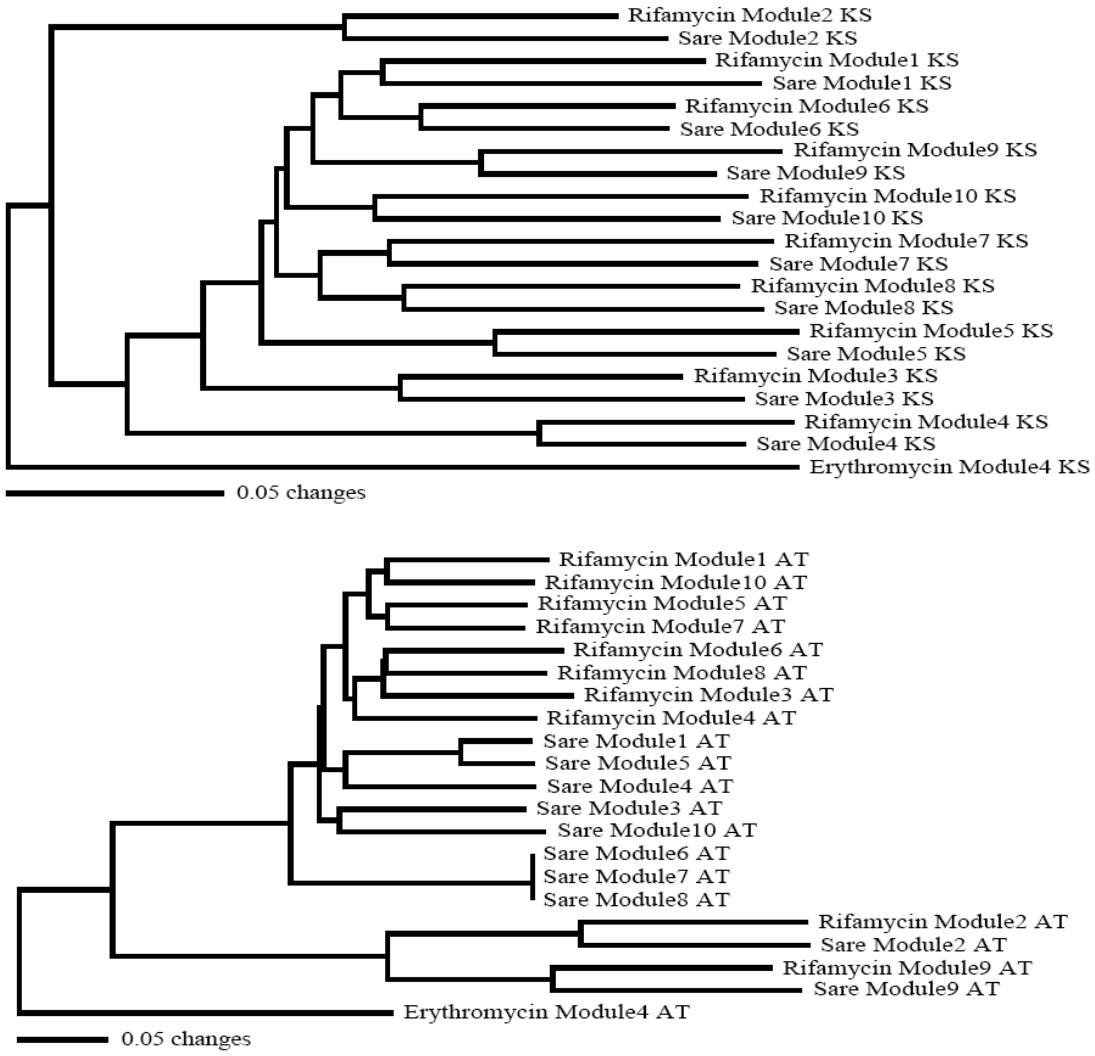

Figure 2.9: Neighbor-joining distance trees constructed in PAUP (43) using the aligned KS, AT, DH, KR and ACP domain sequences from the characterized rifamycin biosynthetic pathway of *Amycolatopsis mediterranei* and the type I modular PKS pathway of *Salinispora arenicola* predicted to biosynthesize a rifamycin-like natural product. For each phylogenetic tree, the appropriate domain from the fourth module of the type I PKS pathway involved in erythromycin biosynthesis was used to position the root.

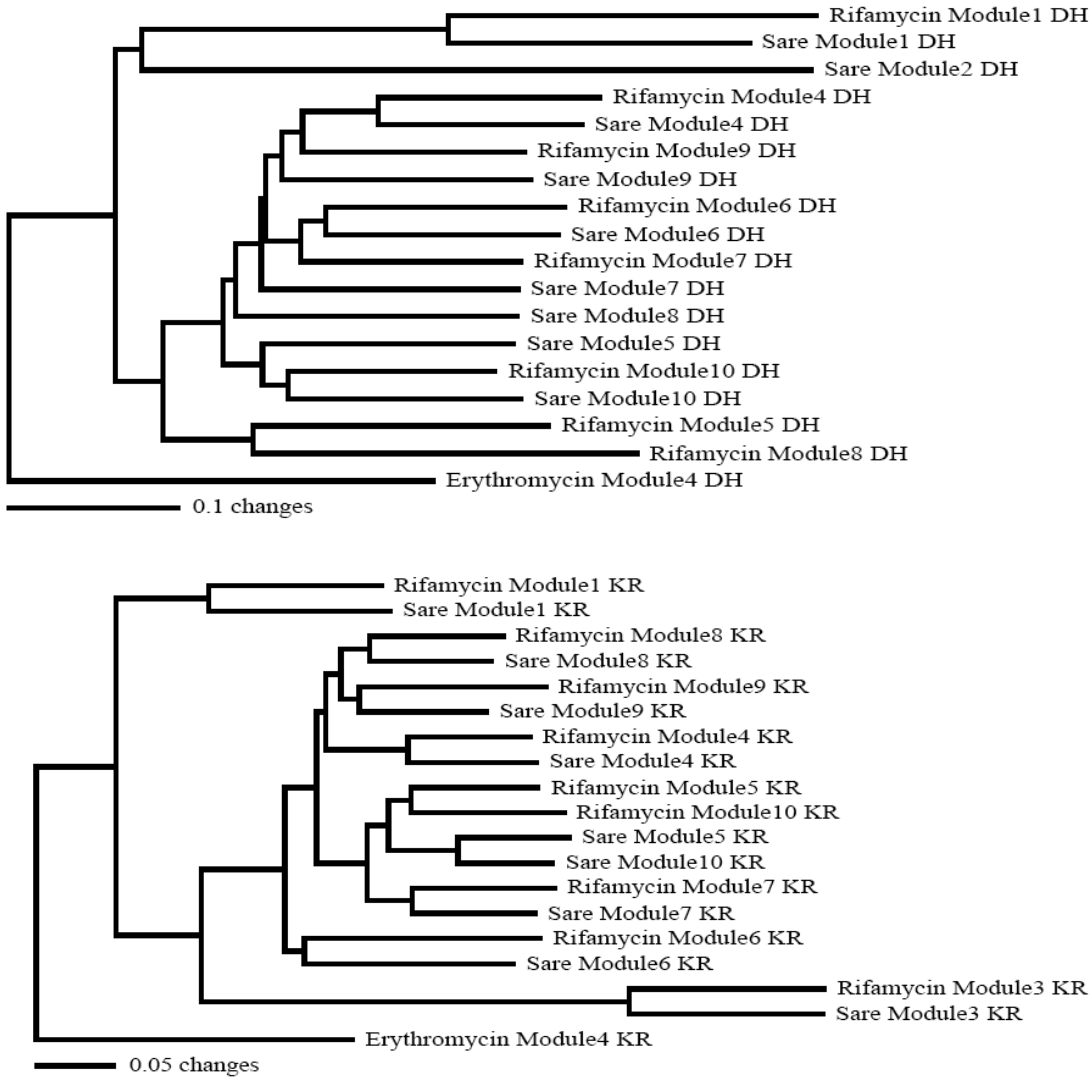

Figure 2.9 continued: Neighbor-joining distance trees constructed in PAUP (43) using the aligned KS, AT, DH, KR and ACP domain sequences from the characterized rifamycin biosynthetic pathway of *Amycolatopsis mediterranei* and the type I modular PKS pathway of *Salinispora arenicola* predicted to biosynthesize a rifamycin-like natural product. For each phylogenetic tree, the appropriate domain from the fourth module of the type I PKS pathway involved in erythromycin biosynthesis was used to position the root.

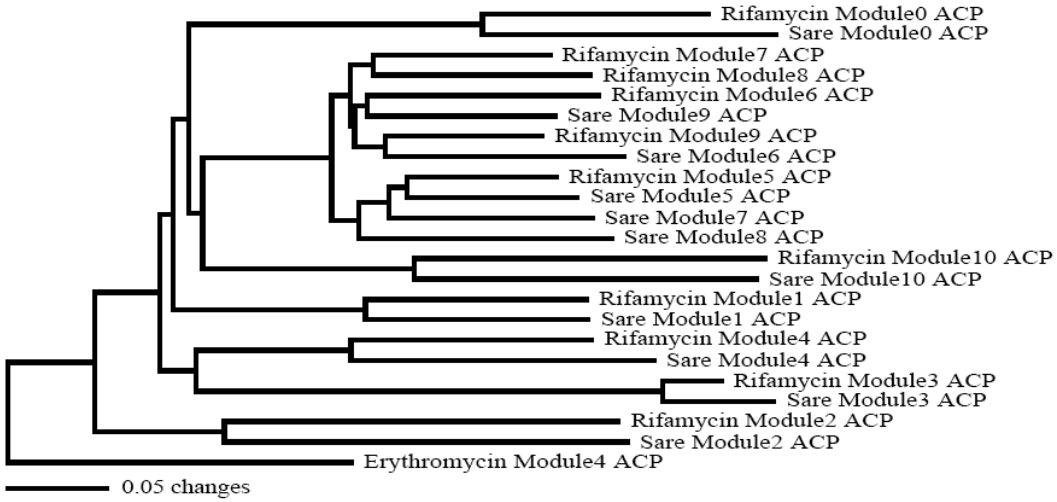

Figure 2.9 continued: Neighbor-joining distance trees constructed in PAUP (43) using the aligned KS, AT, DH, KR and ACP domain sequences from the characterized rifamycin biosynthetic pathway of *Amycolatopsis mediterranei* and the type I modular PKS pathway of *Salinispora arenicola* predicted to biosynthesize a rifamycin-like natural product. For each phylogenetic tree, the appropriate domain from the fourth module of the type I PKS pathway involved in erythromycin biosynthesis was used to position the root.

Chapter Three: Evaluating the Genetic Potential of Actinomycetes to Produce Diverse and Novel Secondary Metabolites

Abstract.

An era of exploration for biomedically relevant secondary metabolites from the marine environment is underway. Rare and novel microorganisms are being discovered and their ability to produce diverse and biologically active compounds is being recognized. Terrestrial actinomycetes are well known producers of secondary metabolites and in an effort to evaluate the genetic potential of actinomycetes cultured from the marine environment to produce secondary metabolites, molecular methods were employed. During this investigation, 60 actinomycete strains cultured from marine sediments, representatives of 52 operational taxonomic units (OTUs), were screened using a PCR-based approach for the presence of genes associated with type I polyketide synthase (type I PKS) pathways, enediyne polyketide synthase (PKSE) pathways and nonribosomal peptide synthetase (NRPS) pathways. Over half of the cultured *Actinomycetales* OTUs were found to possess genes associated with at least one PKS or NRPS biosynthetic pathway. Although some of these actinomycetes represent families well known to produce secondary metabolites, the results suggest that the search for novel secondary metabolites could be expanded to include other filamentous actinomycete families as well as unicellular actinomycete families, including the *Corynebacteriaceae*, *Gordoniaceae Intrasporangiaceae* and *Micrococcaceae*. Furthermore, phylogenetic analyses of the amplified KS domain sequences from type I PKS pathways allowed type I PKS pathway novelty and

richness to be evaluated. Results indicate that by using a phylogenetic approach, the diversity, novelty and identity of type I PKS metabolites can be predicted.

Introduction.

Terrestrial microorganisms, especially actinomycetes (defined here as bacteria within the order *Actinomycetales*), have long been recognized as prolific producers of natural products including polyketides, nonribosomal peptides and combinations thereof (e.g. polyketide/nonribosomal peptide hybrids) (19). Many of these natural products possess biomedically relevant properties, such as antibiotic, antitumor, antifungal, and immunosuppressive activities. Polyketides, nonribosomal peptides and polyketide/nonribosomal peptide hybrids are synthesized by the coordinated action of enzymatic assembly lines (PKS, NRPS and PKS/NRPS hybrid pathways respectively) which conduct the iterative chemical condensation of monomeric units, including carboxylic acid and/or amino acid monomers (10, 19, 42). The order and identity of each domain within a given assembly line specifies the sequence of monomer activation and incorporation, the chemistry that occurs at each step in the assembly line, and the length and functionality of the product released from the assembly line (19). The condensation of a carboxylic acid monomer to a growing acyl chain is accomplished via ketosynthase (KS), acyltransferase (AT) and acyl carrier protein (ACP) domains within PKSs. Following condensation, the presence and functionality of ketoreductase (KR), dehydratase (DH) and enoylreducase (ER) domains determines whether the oxidation state of the beta-carbonyl remains either a keto group or is reduced to a hydroxyl, methine or methylene (27). The incorporation

of amino acid monomers into non-ribosomal peptides is accomplished via the condensation (C), adenylation (A) and peptidyl carrier protein (PCP) domains associated with NRPSs. When present, epimerization (E), methyltransferase (MT) and oxidase (Ox) domains tailor the amino acid being added (19). The immense structural complexity and functional diversity of polyketide, nonribosomal peptide or hybrid compounds is generated both by the intricately coordinated enzyme organization of these assembly lines and specific tailoring enzymes responsible for post-assembly modifications (10, 19).

Over the past several decades, a dramatic drop in the rate of discovery of novel natural products has occurred due in large part to the frequent re-isolation of known metabolites from common terrestrial microorganisms. A rise in the number of drug-resistant pathogens coupled with the limited success of strategies such as combinatorial chemistry has led to a renewed focus on the isolation of rare and novel microorganisms in order to identify novel natural products (33) as well as a call by the National Institutes of Health for the development of new methods and approaches to promote natural product discovery.

In an effort to identify new sources of bioactive secondary metabolites, degenerate primer sets have been developed to screen for the presence of genes associated with PKS and NRPS pathways in genomic (24, 26) and metagenomic libraries (sponge 18, 37, soil 21). PCR primers have also been designed to screen a variety of cultured organisms, including cyanobacteria (9, 15, 34), dinoflagellates (34) and Gram-positive bacterial isolates (2, 3) for PKS and NRPS pathways. The PCR-based approaches have been used successfully to amplify genes associated with

secondary metabolite biosynthesis and therefore it is possible to predict whether secondary metabolite pathways are present within an organism. It becomes possible to predict or estimate the number and novelty of secondry metabolite pathways following phylogenetic analyses of the amplified and sequenced secondary metabolite genes (21). Phylogenetic analyses of the secondary metabolite genes allow predictions to be made because clustering is often observed among KS domains associated with the same type I PKS pathway (21, 22, 31), KS domains associated with PKS/NRPS hybrid pathways (22, 31) and some adenylation domains that activate the same amino acid (11). Cohesion amongst KS domains from the same type I PKS pathway has been proposed to have evolved via gene duplication followed by sequence divergence (25) whereas the clustering of KS domains involved in PKS/NRPS hybrid biosynthesis has been suggested to arise through convergent evolution driven by functional constraints (22, 31).

The world's oceans cover over 70% of the planet's surface and include some of the most diverse ecosystems on Earth. Although for over 30 years structurally diverse natural products with a remarkable range of bioactivities have been recovered from marine algae and invertebrates (7 and references therein, 14, 16 and references therein), the world's oceans are now also beginning to be recognized as a significant resource of microorganisms, including actinomycetes, with the capacity to produce novel natural products (6, 8, 17, 35). Culture-dependent investigations of marine sediments suggest that the marine environment is home to a diverse assemblage of microorganisms, many of which have never been detected in the terrestrial environment (23, 36, 38). These microorganisms provide new opportunities for

natural product research and provide inroads to learning more about the biosynthetic pathways by which natural products are produced.

Using a PCR-based approach, cultured strains from 52 *Actinomycetales* operational taxonomic units (OTUs) recovered during a research expedition to the Republic of Palau were screened for the presence of genes associated with secondary metabolite biosynthesis. The PCR-based approach was used to screen for the presence of domains associated with PKS pathways (including type I and enediyne pathways) and NRPS pathways. In addition to screening OTUs from families well known to be secondary metabolite producers, the presence of genes associated with secondary metabolite biosynthesis was evaluated in lesser-studied bacteria, including unicellular actinomycetes. In addition to surveying *Actinomycetales* OTUs for domains associated with a variety secondary metabolite biosynthetic pathways, a phylogenetic analysis of KS domain protein sequences from type I PKS pathways was employed to evaluate the diversity and novelty of type I PKS pathways present among the cultured strains. The results suggest that over half of the cultured *Actinomycetales* OTUs possess at least one secondary metabolite biosynthetic pathway and that by using a phylogentic approach, the diversity, novelty and identity of the secondary metabolites produced by a cultured microorganism can be predicted. The present study illustrates the power and utility of molecular tools and phylogenetic prediction, describing a method that will improve the process of secondary metabolite discovery.

Materials and Methods.

Nucleic acid extraction, PCR amplification, cloning and sequencing. A total of 60 actinomycete strains, representing 52 *Actinomycetales* OTUs cultured during a research expedition to the Republic of Palau (23) were examined. Genomic DNA was extracted from each isolate according to the DNeasy protocol (Qiagen Inc., Valencia, CA) with several modifications. After RNase A (2 mg/ml) was added to the enzymatic lysis buffer, the resuspended bacterial pellet was incubated for two hours at 37°C. Following the addition of proteinase K and buffer AL, the sample was held for one hour at 70°C. Genomic DNA was eluted from the spin column with 100 μl of buffer AE for immediate use or storage at –20°C.

Ketosynthase (KS) domains of type I polyketide synthase (PKS) pathways were PCR amplified from genomic DNA using the following four primers: PKSI-Fa (5'-CCSCAGSAGCGCSTSTTSCTSGA-3'), PKSI-Fb (5'-CCSCAGSAGCGCSTSCTSCTSGA-3'), PKSI-Ra (5'-GTSCCSGTSCCGTGSGTSTCSA-3') and PKSI-Rb (5'-GTSCCSGTSCCGTGSGCCTCSA-3') (modified from 21). The 50 μl PCR mixture contained 20 to 50 ng of DNA, 400 pmol of each primer, 10X PCR Buffer II (Applied Biosciences, Foster City, CA), 2.5 mM $MgCl_2$ (Applied Biosciences, Foster City, CA), 1.5 U of AmpliTaq Gold DNA polymerase (Applied Biosciences, Foster City, CA), 400 μM deoxynucleoside triphosphate mixture and 7% DMSO. The PCR protocol consisted of a 15 min denaturation at 95°C, 1 cycle of 1 min denaturation at 95°C, 1 min annealing at 65°C and 1 min extension at 72°C, 35 cycles of 1 min

denaturation at 95°C, 1 min annealing at 62°C and 1 min extension at 72°C, followed by 10 min extension at 72°C.

The N-terminal region of iterative enediyne PKS pathways, inclusive of the KS domain, was PCR amplified using the primers EdyA (5'-CCCCGCVCACATCACSGSCCTCGCSGTGAACATGCT-3') and EdyE (5'-GCAGGCKCCGTCSACSGTGTABCCGCCGCC-3') (30). The 50 μl PCR mixture contained 20 to 50 ng of DNA, 500 pmol of each primer, 10X PCR Buffer II (Applied Biosciences, Foster City, CA), 1.5 mM $MgCl_2$ (Applied Biosciences, Foster City, CA), 1 U of AmpliTaq Gold DNA polymerase (Applied Biosciences, Foster City, CA), 200 μM deoxynucleoside triphosphate mixture and 5% DMSO. PCR amplification was performed according to the following protocol: 5 min denaturation at 94°C, 30 cycles of 45 sec denaturation at 94°C, 90 sec annealing at 62°C and 2 min extension at 72°C, followed by 7 min extension at 72°C.

Adenylation domains of NRPS pathways were PCR amplified using the primers NRPS-A3 (5'-GCSTACSYSATSTACACSTCSGG-3') and NRPS-A7R (5'-SASGTCVCCSGTSCGGTAS-3') (3). The 50 μl PCR mixture contained 20 to 50 ng of DNA, 400 pmol of each primer, 10X PCR Buffer II (Applied Biosciences, Foster City, CA), 2.5 mM $MgCl_2$ (Applied Biosciences, Foster City, CA), 1.5 U of AmpliTaq Gold DNA polymerase (Applied Biosciences, Foster City, CA), 400 μM deoxynucleoside triphosphate mixture and 10% DMSO. PCR amplification of NRPS adenylation domains was performed according to the following PCR protocol: 5 min denaturation at 95°C, 35 cycles of 30 sec denaturation at 95°C, 90 sec annealing at 59°C and 1 min extension at 72°C, followed by 10 min extension at 72°C. All PCR

reactions were examined by gel electrophoresis. Bands corresponding to the KS domains (expected size ~670 bp), N-terminal regions of enediyne PKS pathways (expected size ~1.4 kb) or the adenylation domains (expected size ~700 bp) were excised using a sterile scalpel and purified using Qiagen's QIAquick Gel Extraction kit according to the manufacturer's suggested protocol (Qiagen Inc., Valencia, CA).

PCR products were immediately cloned using the TOPO TA cloning kit (Invitrogen Co., Carlsbad, CA) according to the manufacturer's suggested protocol with the following modifications. The 6 μl TOPO cloning reaction consisted of 1 μl fresh PCR product, 1 μl salt solution, 3 μl sterile water and 1 μl TOPO vector. The reactions were mixed gently and allowed to incubate at room temperature for 15 minutes. Once the ligation reaction mixture was added to the competent cells, the cells were incubated on ice for 30 min, heat shocked at 42^{o}C for 30 sec, and then allowed to recover during one hour of shaking at 37^{o}C with 250 μl of SOC. Following recovery, 25 μl and 50 μl of the transformed cells were plated onto LB agar plates containing X-gal and kanamycin [50 μg]. Following overnight incubation at 37^{o}C, the presence of an insert was verified by restriction digest. Vectors were isolated using Qiagen's QIAprep Spin Miniprep kit according to the manufacturer's suggested protocol (Qiagen Inc., Valencia, CA). Following isolation, 5 μl of the vector were added to 1 μl of the restriction enzyme EcoR I (New England Biolabs, Inc., Beverly, MA), 2 μl of 10X reaction buffer and 12 μl of sterile water. The 20 μl digest was incubated at 37^{o}C for one hour and vectors confirmed to contain correctly sized inserts were sequenced using the M13F primer on an ABI 3100 DNA sequencer at the DNA

Sequencing Shared Resource, UCSD Cancer Center (funded in part by the NCI Cancer Center Support Grant #2 P30CA23100-18).

Sequence analysis. For every PCR reaction yielding a correctly sized product, at least one insert-containing vector was sequenced to verify the amplification of the appropriate secondary metabolite pathway component. To perform a phyolgenetic analysis of KS domain protein sequences from type I PKS pathways, as many as 51 individual clones were sequenced following PCR amplification of the KS domain or domains from a single organism. Nucleotide sequences were analyzed and manually edited using the Sequencher software package (version 4.5; Gene Codes Co., Ann Arbor, MI) and translated using the Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/). Protein sequences were compared to sequences within the NCBI protein database (http://www.ncbi.nlm.nih.gov/) using the Basic Local Alignment Search Tool (BLAST). The most closely related KS domain for each sequence as determined by BLAST, as well as all other KS domains from that type I biosynthetic pathway, were used in part to create a reference data set.

Phylogenetic analysis. The protein sequences of cloned KS domains were aligned with a large reference data set of published sequences. This reference data set includes 332 KS domains from 34 type I PKS pathways, including 23 experimentally characterized pathways previously analyzed by Ginolhac *et al.* (21, 22). Of the 11 additional pathways used to construct the reference data set, eight had been experimentally characterized whereas three had been sequenced and annotated but not yet experimentally characterized as of the preparation of this manuscript (Appendix B). The reference data set included the best BLAST matches for each of the KS

domains sequenced during this study. The 390 KS domains sequenced during this study were aligned to the reference data set using ClustalX version 1.83 (40) and manually edited using MacClade version 4.07 (32). Phylogenetic trees were constructed using the distance neighbor-joining, distance unweighted-pair group method using average linkages (UPGMA), and maximum parsimony methods in PAUP (39).

Fermentation and chemical analysis. Strain CNR-925 was inoculated from a frozen stock into 25 ml of medium A1 (1.0% soluble starch, 0.4% yeast extract, 0.2% peptone, 75% seawater and 25% deionized (DI) water). This culture was incubated at 27ºC, stirring at 230 rpm for three days and then transferred to a 2.8 L Fernbach flask containing 1 L of the same medium. Following another three days of incubation at 27ºC, stirring at 230 rpm, 25 ml of the culture was added to 1 L each of the production media, A1BFe+C (1.0% starch, 0.4% yeast extract, 0.2% peptone, 0.1% $CaCO_3$, 0.5% $Fe_2(SO_4) \cdot 4H_2O$, 0.5% KBr, 1 L seawater, pH 8.0) and E1 (1.0% glucose, 1.0% soluble starch, 0.5% corn steep liquor, 0.3% NaCl, 0.1% $MgSO_4 \cdot 7H_2O$, 0.5% $CaCO_3$, 1 L of DI water, pH 7.0 (29)). On days three, five and seven, 25 ml of each culture were removed and extracted with 50 ml of ethyl acetate. On day nine, the remainder of the cultures was extracted with 2 L of ethyl acetate. The organic layers were separated, dried over sodium sulfate, decanted, and concentrated under vacuum. The resulting crude extracts were dissolved in methanol and analyzed using an Agilent 1100 LC/MS at 254 and 350 nm in the positive mode using a standard acetonitrile/water gradient, from 10% acetonitrile in 90% water to 100% acetonitrile in 20 minutes.

Method validation and development. The KS domain protein sequences of type I, type II and enediyne PKS pathways were obtained from the genome sequence of *Salinispora arenicola* strain CNS-205 (NC009953). The genome was queried for type I, type II and enediyne KS domain sequences by using the Joint Genome Institute (JGI) BLAST search engine (http://genome.ornl.gov/microbial/sare/), using KS domains from the rifamycin (CAA11035), actinorhodin (Q02059) and calicheamicin (AAM94794) biosynthetic pathways respectively. These KS domain sequences were aligned using ClustalX version 1.83 (40) and manually edited using MacClade version 4.07 (32). A bioinformatic evaluation of the pathways associated with the biosynthesis of type I polyketides identified two type I modular PKS pathways, two type I iterative PKS pathways and three KS domains associated with the biosynthesis of PKS/NRPS hybrid compounds. In order to determine the effectiveness of the PCR primers, the KS domains of CNS-205 were amplified, cloned and sequenced as described above. In an effort to assess the effect of PCR bias, the cloning experiments were performed following a single PCR and after pooling three separate PCR reactions. The cloned sequences were compared to the KS domains retrieved from the genome, allowing the number and type of KS domains amplified to be enumerated.

Results.

PCR screening. Using a PCR-based approach, 60 actinomycete strains representing 52 OTUs were screened for the presence of gene sequences associated with type I PKS, enediyne PKS and NRPS pathways. Of the 60 actinomycetes screened, 40 (66.7%) possessed a domain from at least one of the three pathways

(Table 3.1). The presence of type I KS domain sequences was verified in 26 (43.3%) isolates from 22 (42.3%) of the OTUs, whereas the presence of an enediyne KS domain sequence was verified in five (8.3%) isolates representing three (5.8%) *Actinomycetales* OTUs. Adenylation domains were present in 38 (63.3%) strains from 32 (61.5%) of the 52 OTUs.

The 60 actinomycete isolates screened are members of 18 different family level groupings and 25 genera. Twelve (66.7%) of the families and 13 (52.0%) of the genera possessed a domain from at least one of the three secondary metabolite biosynthetic pathways. Type I KS domains were identified in eight (44.4%) of the families and nine (36.0%) of the genera and adenylation domains were observed in ten (55.6%) of the families and a total of 11 (44.0%) of the genera. Only the two genera of the *Micromonosporaceae* family possessed KS domains associated with the production of an enediyne warhead via an enediyne PKS.

Cloning and sequence analysis. Clone libraries were generated from all 26 of the actinomycete strains that tested positive for the presence of type I PKS sequences, with as many as 51 clones being sequenced from any one library (Table 3.2). Cloning provided an opportunity to sequence verify the inserts and assess the richness and novelty of potential pathway products using phylogenetic analyses.

From CNJ-863, a strain most closely related to *Gordonia nitida*, a total of 26 KS domains were sequenced. All 26 sequences were almost identical and none shared greater than 70% sequence identity with any previously sequenced KS domain. However, BLAST analyses of the sequences indicated that they were most closely related to KS domains associated with hybrid PKS/NRPS biosynthetic pathways

including the microcystin biosynthetic pathway (41). During subsequent phylogenetic analyses, the 26 sequences claded tightly with KS domains from the reference data set associated with hybrid PKS/NRPS biosynthetic pathways including the epothilone, myxothiazol, microcystin and nostopeptolide pathways (data not shown).

A total of 41 clones from the KS domain library of CNR-885, a strain most closely related to *Streptomyces tendae*, were sequenced. Of the 41 clones sequenced, 14 unique KS domains were identified (Figure 3.1). Although the 14 KS domains did not share greater than 70% amino acid sequence identity with any previously sequenced KS domains, phylogenetic analyses identified five separate clades of KS domains among other type I modular PKS pathways. KS domains from the same type I PKS pathways generally cluster together and the sequence identity between those KS domains from the same type I PKS pathway is generally greater than 80%. KS domains from different type I PKS pathways almost always share significantly less than 70% sequence identity. KS domains among the five clusters shared no more than 70% sequence identity whereas when multiple domains were present, KS domains within a cluster shared no less than 80% sequence identity. Two of the five clades consisted of three KS domains whereas one cluster was made up of six KS domain sequences. The two remaining KS domains shared less than 70% sequence identity with one another and the other KS domain sequences leading to a prediction that CNR-885 has at least five type I PKS pathways.

From CNR-925, a strain most closely related to *Streptomyces thermocoprophilus*, 14 KS domains were sequenced. Six unique KS domains were identified and all six

KS domains shared high sequence identity (> 95%) with KS domains of known type I PKS pathways (Figure 3.2). Three of the KS domains shared high sequence identity (no less than 97% sequence identity) with three different domains from the type I PKS pathway responsible for tetronomycin biosynthesis. The other three KS domains shared high sequence identity (no less than 95% sequence identity) with three different KS domains of the recently described actinofuranone biosynthetic pathway.

Fermentation studies. To assess whether the phylogenetic approach could be used to predict the production of two compounds, fermentation experiments were designed to facilitate tetronomycin and actinofuranone production in CNR-925. LC/MS confirmed the production of tetronomycin on day five in medium E1. At an absorbance of 350 nm, a peak with the retention time of 14.858 min matched the UV profile and mass fragmentation pattern of tetronomycin (29). LC/MS also confirmed the production of actinofuranone on day seven in medium A1BFe+C. At an absorbance of 350 nm, a peak with the retention time of 16.336 min matched the UV profile and mass fragmentation pattern of actinofuranone (4, 12). Of all the KS domains sequenced during this study, only the six KS domains of CNR-925 shared high sequence identity with any known type I PKS pathway.

Method validation. A total of 25 type I KS domains, six type II KS domains and two enediyne KS domains were obtained following a bioinformatic analysis of the genome sequence of *Salinispora arenicola* strain CNS-205. Of the 25 type I KS domains aligned, 20 were associated with two type I modular PKS pathways, two were associated with two type I iterative PKS pathways, whereas the final three KS domains were associated with biosynthetic pathways predicted to produce PKS/NRPS

hybrid compounds. From the forty-seven KS domains sequenced following a single PCR amplification, seven unique type I KS domains were identified (Figure 3.3). Of the four KS domains associated with a modular type I PKS pathway, three were associated with the rifamycin biosynthetic pathway whereas the other KS domain was associated with the second modular type I PKS pathway for which the product is not yet known. Sequencing also revealed KS domains associated with both of the type I iterative PKS pathways and one of the three PKS/NRPS hybrid pathways. In total, five of the seven type I PKS or hybrid PKS/NRPS pathways were detected. Cloning the combined products of three individual PCR reactions did not increase the number of KS domains detected. KS domains associated with either type II or enediyne PKS pathways were not detected using the PCR primers designed to detect type I PKS pathways.

The rifamycin biosynthetic pathway was first described in *Amycolatopsis mediterraneii* (1). Rifamycin production was later identified in *Salinispora arenicola* and the biosynthetic pathway hypothesized to have been obtained via horizontal gene transfer (28). After sequencing the genome of *Salinispora arenicola* strain CNS-205, the KS domains of the two rifamycin pathways were compared. KS domains of most type I modular PKS pathways cluster tightly together and therefore, should the two rifamycin biosynthetic pathways have shared a common ancestor, it would follow that the KS domains of the two pathways would be interleaved. As expected, each of the 10 KS domains within the *Salinispora arenicola* rifamycin biosynthetic pathway was more closely related to the corresponding KS domain in the *Amycolatopsis mediterranei* pathway than to any of the other nine KS domains within the

Salinispora rifamycin pathway (Figure 3.1). Of the 223 amino acids compared between each KS domain, corresponding KS domains shared no more than 211 amino acids (94.6% sequence identity) and no less than 187 amino acids (83.9% sequence identity). The high sequence identity among KS domains of the rifamycin biosynthetic pathways encouraged predictions to be made regarding compound production in CNR-925. As alluded to above, when type I KS domains shared high amino acid sequence identity with previously sequenced KS domains from characterized biosynthetic pathways, the strain was predicted to have the capacity to produce the known compound.

Discussion.

The number of drug leads identified using discovery platforms such as combinatorial chemistry and high throughput screening has not met expectations (20). However, advancements in the areas of small molecule detection, isolation and elucidation, DNA sequencing and an improved understanding of the molecular genetics of secondary metabolite biosynthesis, have renewed interest in the investigation of novel and rare microorganisms as a source for novel natural products. The research described provides an assessment of the secondary metabolite biosynthetic potential of diverse actinomycetes isolated from marine sediments and illustrates how a phylogenetic approach can be used in conjunction to predict metabolite diversity and novelty prior to chemical isolation and characterization of cultured strains. The study used a PCR-based screen for type I and enediyne PKS pathways as well as NRPS pathways, and the phylogenetic analysis of sequenced type

I KS domains enabled predictions regarding the number and novelty of type I PKS pathways present to be made.

A culture-dependent assessment of Gram-positive bacterial diversity present in tropical marine sediments identified 52 *Actinomycetales* OTUs from 18 distinct actinomycete families (23). When screened using a PCR-based approach for genes associated with type I PKS, enediyne PKS and NRPS pathways, 34 (45.6%) OTUs representing 12 (66.7%) families harbored genes associated with at least one of the three biosynthetic pathways. Although the *Streptomycetaceae* and *Micromonosporaceae* families are the actinomycetes well known to produce secondary metabolites, the present study suggests that the search for novel secondary metabolites need not be confined only to well-studied families. The present study suggests that a search for novel secondary metabolites could be expanded to include other filamentous actinomycetes as well as unicellular actinomycete families such as the *Corynebacteriaceae*, *Gordoniaceae*, *Intrasporangiaceae* and *Micrococcaceae*.

The PCR-based approach offers a method by which to quickly screen a large collection of isolates in order to identify strains with the potential to produce specific structural classes of secondary metabolites. Not all strains within a family are likely to produce PKS or NRPS derived metabolites. Therefore the PCR-based method provides a rapid means of surveying cultured strains and genera to determine how to best manage and proceed with fermentation and isolation experiments aimed at identifying novel natural products. Although the screening method offers a foundation upon which to build a natural products drug discovery platform, the PCR-based approach described here does have limitations. Because members of the

Streptomycetaceae and *Micromonosporaceae* families have been identified as among the most prolific producers of secondary metabolites, it follows that the secondary metabolite biosynthetic pathways of these two actinomycete families are the best characterized. As a result, PCR primers designed to amplify genes associated with PKS and NRPS pathways are biased towards the gene sequences of these two families. Despite the noted bias, a large number of strains from families other than the *Streptomycetaceae* and *Micromonosporaceae* yielded PCR products that were later sequence verified as PKS or NRPS genes. Given that the PCR primers are likely to be biased towards known biosynthetic pathways, by identifying genes associated with PKS or NRPS pathways in phylogenetically novel isolates and families from which no secondary metabolite have so far been found, we demonstrate that the PCR-based approach can successfully suggest which isolates or genera to investigate more rigorously. Additionally, the PCR-based approach can only indicate when a gene is present. If a PCR reaction fails to yield a product, the conclusion cannot be that the strain lacks such a gene. While possible that the strain does not harbor any PKS or NRPS genes, it is equally plausible that the PCR primers are not specific enough to amplify those genes in strains less well known to produce secondary metabolites of biomedical relevance. Regardless of its limitations, the screening method quickly identified a large number of isolates to be chemically analyzed more rigorously. In the future, as the number of genome sequences available for study increases, so will the number of biosynthetic pathways available with which to design better primers to amplify diverse PKS and NRPS genes.

The PCR-based approach provides a qualitative answer as to whether a gene from a type I PKS, enediyne PKS or NRPS biosynthetic pathway is present. While the information is useful, a phylogenetic analysis of genes associated with secondary metabolism can offer additional insight into both the novelty and diversity of pathways present within a microorganism. A phylogenetic approach using the type I KS domain sequences of 26 actinomycete isolates illustrated how sequence information can be used to predict the number and novelty of type I PKS and mixed PKS/NRPS pathways present.

Twenty-six KS domains cloned from CNJ-863 were sequenced and determined to be almost identical. Among the sequences, the small nucleotide differences were scattered throughout and are most likely attributed to errors that occurred during cloning. The 26 KS domain sequences did not share high sequence identity with any previously sequenced KS domain and while possible that these KS domains are associated with the production of a known compound whose pathway has not yet been characterized, equally plausible is the conclusion that the KS domains are novel and associated with an as yet unknown pathway and product. During BLAST and phylogenetic analyses, the domain sequences grouped most closely with other KS domains associated with hybrid PKS/NRPS biosynthetic pathways. The clustering of KS domains involved in PKS/NRPS hybrid biosynthesis has been observed previously and is suggested to occur due to a convergent evolution of KS domains involved in such hybrid systems (22, 31). An investigation of KS domains clustered within a hybrid PKS/NRPS group discovered that they are preceded by an NRPS module leading to the hypothesis that the clustering of KS domains is due to a functional

constraint (22). KS domains positioned downstream of a NRPS modules have an unusual function; they mediate the condensation of an acyl extender unit onto the amino acid chain from the previous NRPS module (22). Coevolution to perform his common might explain why these KS domains of hybrid PKS/NRPS pathways are clustered together as opposed to with other type I PKS KS domains (22). While unable to determine whether an NRPS module precedes the KS domains from CNJ-863, the suggestion is that a single KS domain amplified from CNJ-863 is involved in the biosynthesis of a natural product through a mixed PKS/NRPS pathway. Although no compounds of mixed biosynthetic origin have been elucidated through fermentation experiments thus far (data not shown), it is possible the pathway is either nonfunctional or not expressed under the experimental conditions used.

A total of 41 KS domains from CNR-885 were sequenced. Of the 14 unique KS domains identified, not one shared a high sequence identity with any previously sequenced KS domain. While a lack of high sequence identity to any known KS domain present in the NCBI database could be attributed to the amplification of KS domains of known compounds produced by as yet uncharacterized biosynthetic pathways, alternatively the KS domains could be associated with novel biosynthetic pathways with as yet unelucidated natural products. The subsequent phylogenetic analysis suggested that at least five type I modular PKS clusters are present within CNR-885. The KS domain richness and novelty has invited reexamination of this strain's extracts.

Almost all of the KS domains sequenced as part of this study shared low amino acid sequence identity with any previously characterized pathway. Observations of

the *Amycolatopsis* and *Salinispora* rifamycin biosynthetic pathways suggested that KS domains performing the same function share at least 80% amino acid sequence identity. Therefore, when the six KS domains of CNR-925 shared no less than 95% amino acid sequence identity with either KS domains associated with actinofuranone biosynthesis or tetronomycin biosynthesis, CNR-925 was predicted to make both of the known compounds. Under the appropriate fermentation conditions, both of these predicted compounds were identified and as such, the results suggest that a phylogenetic approach provides a means to dereplicate known compounds without the need for chemical screening.

CNR-925 produced actinofuranone under one set of fermentation conditions and tetronomycin under a different set of fermentation conditions. The fact that different fermentation conditions were required to isolate each of the predicted compounds coupled with the fact that tetronomycin production was only observed on day five, underscores the importance of how strains are grown and when they are extracted. Genomic analyses of long studied strains such as *Streptomyces coelicolor* strain A3(2) reveal clusters of genes characteristic of biosynthetic pathways responsible for the production of never before isolated secondary metabolites (5). Secondary metabolite pathways are likely only to be active under specific physical, chemical and biological conditions. Therefore, as suggested by fermentation studies of CNR-925, the strategies by which novel natural products are elucidated should be varied. Novel actinomycete isolates as well as long-studied actinomycetes may have the capacity to produce a large number of secondary metabolites that have yet to be described and it is imperative that the means of coaxing production of novel natural products be better

understood. By challenging strains to grow in the presence of physical (desiccation, high/low temperature), chemical (low nutrient, low iron) and/or biological (competition) stresses, and by observing cultures at different stages of their development, novel natural products will be discovered. Patience, creativity and an enhanced understanding of the processes that control the production of secondary metabolites will allow an increasing number of natural products to be discovered.

Strong conservation of the enediyne PKS across several actinomycete genera suggested that the enediyne PCR primers used during this study might effectively identify diverse actinomycete strains in possession of enediyne warhead cassette genes. Of the 60 strains screened, five strains from the *Salinispora* and *Micromonospora* genera yielded PCR products consistent with enediyne PKS genes. Although the percentage of marine-derived actinomycetes in possession of enediyne PKS genes was lower than that reported for soil-derived actinomycetes (43), the fact that marine obligate *Salinispora* strains possess enediyne PKS genes suggests that the marine environment may be an important source of novel enediyne secondary metabolites. Upon further investigation of the genome of *Salinispora arenicola* strain CNS-205, the presence of pathways responsible for the putative production of both a nine-membered and a ten-membered enediyne was confirmed. Although the products of these two pathways have not yet been elucidated, other *Salinispora* strains have been hypothesized to produce metabolites derived from enediyne precursors (8, 35). Although all primers used during the present study may be biased towards the well-studied actinomycete families, members of the *Micromonosporaceae* cultured in Palau appear to contain a wealth of secondary metabolite biosynthetic genes. The PCR-

based approach only hints at the diversity of pathways present. The genome sequence of CNS-205, in addition to the two enediyne PKS pathways, harbors type I, type II and type III PKS pathways in addition to NRPS and PKS/NRPS biosynthetic pathways. Although genome sequences provide tremendous insight into the secondary metabolite biosynthetic potential of isolates, a PCR-based approach coupled with phylogenetic analyses currently provides a more cost effective and faster means of screening a large collection of isolates for strains with the capacity to produce secondary metabolites. The analysis allows both energy and effort to be focused on the most promising strains in the search for novel natural products.

The PCR-based screening approach coupled with a phylogenetic analysis of the secondary metabolite gene sequences offers a robust method by which to focus a search for novel natural products on cultured isolates with the greatest potential to produce natural products. As the number of genome sequences and secondary metabolite gene clusters available in the public domain increases, so will the effectiveness of this method. Over time, this method will become a robust way to predict the presence of pathways responsible for novel chemistry by dereplicating known compounds prior to chemical analyses. As the number of characterized secondary metabolite pathways increases, so will the number of sequences available for primer design. The type I PKS primers used during this study did not amplify 100% of the KS domains present in CNS-205, a *Salinispora arenicola* strain known to have 25 type I KS domains. The effect of degenerate primers as well as PCR bias cannot fully account for the discrepancy between the number of KS domains present and the numbers of KS domains amplified. Many of the KS domains not amplified

shared fewer mismatches with the PCR primers than KS domains successfully amplified and pooling PCR reactions did nothing to increase the number of domains amplified. Differences in the secondary structure of various regions of the genome may explain why some genes are more amenable to amplification while others are not. In silico analyses of the type I KS domain sequences of CNS-205 using Beacon Designer version 7.0 software (http://mac.softpedia.com/get/Math-Scientific/Beacon-Designer.shtml) suggest that DNA folding at or near primer annealing sites is likely to reduce effective PCR amplification (data not shown). Future development of this method should take DNA folding into consideration during the design of primers and the PCR amplification protocol.

For each cultured microorganism screened for the presence of type I PKS genes using the PCR-based approach, only a genomic DNA extraction followed by a single PCR, a single clone library and a single sequencing reaction is required. Sequencing multiple clones and performing a phylogenetic analysis on those sequences can provide additional information regarding the number and diversity of PKS pathways present. Although multiple clones must be sequenced for the purposes of assessing the novelty and diversity of pathways present, screening for the presence of a type I PKS pathway requires only a couple of quick steps and many isolates can be screened at a time. The PCR-based screen for secondary metabolite pathways cannot replace genome sequencing with respect to information gained, however the sequencing of genomes of individual isolates has yet to become cost effective.

The PCR-based approach also offers advantages over the scanning of large metagenomic libraries for genes associated with secondary metabolism. While a

metagenomic approach offers direct access to the genomes of many as-yet-uncultured microorganisms, the genome scanning approach requires environmental DNA extraction, construction of a metagenomic fosmid library, PCR screening of clones for type I PKS genes, localization of the type I PKS positive clones, and the sequencing of fosmids (21). Coupled with a phylogenetic analysis of the sequence information, the metagenomic approach offers an opportunity to decide which clones should be the focus of further study and yet there are many additional concerns to be addressed prior to the identification of the putative compound. A tremendous amount of work goes into identifying the 0.23% of clones from a metagenomic library expected to have type I PKS genes (21, 13), and yet even if the genes are present, the entire pathway may not have been captured on the fosmid. If the entire pathway was indeed captured there could be issues with respect to heterologous expression of the pathway, and unless the fosmid also captured signatures of the original organism, no information about how future efforts to culture similar isolates with a capacity to produce novel secondary metabolites will be available. While the PCR-based screen will not allow the product of a novel pathway to be predicted (unlike what may be possible for a sequenced fosmid), for cultured isolates the PCR-based method is a rapid and effective way to determine which strains need to be more rigorously studied. The method also offers insight into which other families and strains possess the capacity to produce secondary metabolites and should be more thoroughly investigated for their chemical compound production.

The future of natural product drug discovery will undoubtedly continue to be influenced by advances in our understanding of the molecular genetics of secondary

metabolite biosynthesis. Sequence-based approaches will likely be incorporated into discovery paradigms and provide perspective on the biosynthetic potential of an organism prior to fermentation, extraction and chemical analysis. As these methods are developed and refined, it is conceivable that the stochastic approach by which microorganisms have traditionally been cultured and screened will soon be abandoned in favor of genetic approaches that provide insight into the organism's genetic potential. The present study provides evidence that tools are now available to assess the potential utility of phylogenetic prediction as a method for improving the process of secondary metabolite discovery.

References.

1. **August, P. R., L. Tang, Y.J. Yoon, S. Ning, R. Müller, T.-W. Yu, M. Taylor, D. Hoffmann, C.-G. Kim, X. Zhang, C.R. Hutchinson, and H.G. Floss.** 1998. Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the rif biosynthetic gene cluster of *Amycolatopsis mediterranei*. Chem. Biol. 5(2):69-79

2. **Ayuso, A., D. Clark, I. Gonzalez, O. Salazar, A. Anderson, and O. Genilloud.** 2005. A novel actinomycete strain de-replication approach based on the diversity of polyketide synthase and nonribosomal peptide synthetase biosynthetic pathways. Appl. Microbiol. Biotechnol. 67:795-806.

3. **Ayuso-Sacido, A. and O. Genilloud.** 2005. New PCR primers for the screening of NRPS and PKS-I systems in actinomycetes: detection and distribution of these biosynthetic gene sequences in major taxonomic groups. Microb. Ecol. 49:10-24.

4. **Banskota, A. H., J. B. McAlpine, D. Sorensen, M. Aouidate, M. Piraee, A.-M. Alarco, S. Omura, K. Shiomi, C. M. Farnet, and E. Zazopoulos.** 2006. Isolation and identification of three new 5-alkenyl-3,3(2H)-furanones from two *Streptomyces* species using a genomic screening approach. J. Antibiot. 59(3):168-176.

5. **Bentley, S. D., K. F. Chater, A.-M. Cerdeño-Tárraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A.**

Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C.-H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M.-A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill, and D. A. Hopwood. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature. 417:141-147.

6. **Bernan, V. S., M. Greenstein and G. T. Carter**. 2004. Mining marine microorganisms as a source of new antimicrobials and antifungals. Curr. Med. Chem. Anti.-Infect. Agents. 3(3):181-195

7. **Blunt, J. W., B. R. Copp, W.-P. Hu, M. H. G. Munro, P. T. Northcote and M. R. Prinsep.** 2007. Marine natural products. Nat. Prod. Rep. 24(1):31-86.

8. **Buchanan, G. O., P. G. Williams, R. H. Feling, C. A. Kauffman, P. R. Jensen, and W. Fenical.** 2005. Sporolides A and B: structurally unprecedented halogenated macrolides from the marine actinomycete *Salinispora tropica*. Org. Lett. 7(13):2731-2734.

9. **Burns, B. P., A, Seifert, F. Goh, F. Pomati, A.-D. Jungblut, A. Serhat, and B. A. Neilan.** 2005. Genetic potential for secondary metabolite production in stromatolite communities. FEMS Microbiol. Lett. 243:293-301.

10. **Cane, D. E., C. T. Walsh, and C. Khosla.** 1998. Harnessing the biosynthetic code: combinations, permutations, and mutations. Science. 282:63-68.

11. **Challis, G. L., J. Ravel and C. A. Townsend.** 2000. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. Chem. Biol. 7:211-224.

12. **Cho, J. Y., H. C. Kwon, P. G. Williams, C. A. Kauffman, P. R. Jensen, and W. Fenical.** 2006. Actinofuranone A and B, new polyketides from a marine bacterium of the genus *Streptomyces* (Actinomycetes) J. Nat. Prod. 69(3):425-428.

13. **Courtois, S., C. S. Cappellano, M. Ball, F. X. Francou, P. Normand, G. Helynck, A. Martinez, S. J. Kolvek, J. Hopke, M. S. Osburne, P. R. August, R. Nalin, M. Guerineau, P. Jeannin, P. Simonet, and J. L. Pernodet.** 2003. Recombinant environmental libraries prove access to microbial diversity for drug discovery from natural products. Appl. Environ. Microbiol. 69:49-55.

14. **Davidson, S. K., S. W. Allen, G. E. Lim, C. M. Anderson, and M. G. Haygood.** 2001. Evidence for the biosynthesis of bryostatins by the bacterial symbiont "*Candidatus* Endobugula sertula" of the bryozoan *Bugula neritina* Appl. Environ. Microbiol. 67(10):4531-4537.

15. **Ehrenreich, I. M., J. B. Waterbury, and E. A. Webb.** 2005. Distribution and diversity of natural product genes in marine and freshwater cyanobacterial cultures and genomes. Appl. Environ. Microbiol. 71(11):7401-7413.

16. **Faulkner, J. D.** 2002. Marine natural products. Nat. Prod. Rep. 19(1):1-48.

17. **Feling, R. H., G. O. Buchanan, T. J. Mincer, C. A. Kauffman, P. R. Jensen, and W. Fenical.** 2003. Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora*. Angew. Chem. Int. Ed. 42:355-357.

18. **Fiesler, L., U. Hentschel, L. Grozdanov, A. Schirmer, G. Wen, M. Platzer, S. Hrvatin, D. Butzke, K. Zimmerman, and J. Piel.** 2007. Widespread occurrence and genomic context of unusually small polyketide synthase genes in microbial consortia associated with marine sponges. Appl. Environ. Microbiol. 73(7):2144-2155.

19. **Fischbach, M. A. and C. T. Walsh.** 2006. Assembly-Line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem. Rev. 106:3468-3496.

20. **Gershell, L. J. and J. H. Atkins.** 2003. A brief history of novel drug discovery technologies. Nature Reviews Drug Discovery. 2:321-327

21. **Ginolhac, A., C. Jarrin, B. Gillet, P. Robe, P. Pujic, K. Tuphile, H. Bertrand, T. M. Vogel, G. Perriere, P. Simonet, and R. Nalin.** 2004. Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. Appl. Environ. Microbiol. 70(9):5522-5527.

22. **Ginolhac, A., C. Jarrin, P. Robe, G. Perriere, T. M. Vogel, P. Simonet, and R. Nalin.** 2005. Type I polyketide synthases may have evolved through horizontal gene transfer. J. Mol. Evol. 60:716-725.

23. **Gontang, E. A., W. Fenical, and P. R. Jensen.** 2007. Phylogenetic diversity of gram-positive bacteria cultured from marine sediments. Appl. Environ. Microbiol. 73: 3272-3282.

24. **Haydock, S. F., T. Mironenko, H. I. Ghoorahoo, and P. F. Leadlay.** 2004. The putative elaiophylin biosynthetic gene cluster in *Streptomyces* sp. DSM4137 is adjacent to genes encoding adenosylcobalamin-dependent methylmalonyl CoA mutase and to genes for synthesis of cobalamin. J. Biotechnol. 113:55-68.

25. **Hopwood, D. A.** 1997. Genetic contributions to understanding polyketide synthases. Chem. Rev. 97:2465-2497.

26. **Izumikawa, M., M. Murata, K. Tachibana, Y. Ebizuka, and I. Fujii.** 2003. Cloning of modular type I polyketide synthase genes from the salinomycin producing strain of *Streptomyces albus*. Bioorg. Med. Chem. 11:3401-3405.

27. **Jenke-Kodama, H., A. Sandmann, R. Muller, and E. Dittmann.** 2005. Evolutionary implications of bacterial polyketide synthases. Mol. Biol. Evol. 22(10):2027-2039.

28. **Jensen, P. R., P. G. Williams, D.-C. Oh, L. Zeigler, and W. Fenical.** 2007. Species-specific secondary metabolite production in marine actinomycetes of the genus *Salinispora*. Appl. Envir. Microbiol. 2007 73:1146-1152.

29. **Keller-Juslén, C., H. D. King, M. Kuhn, H.-R. Loosli, W. Pache, T. J. Petcher, H. P. Weber, and A. von Wartburg.** 1982. Tetronomycin, a novel polyether of unusual structure. J. Antibiot. 35(2):142-150.

30. **Liu, W., J. Ahlert, Q. Gao, E. Wendt-Pienkowski, B. Shen, and J. S. Thorson.** 2003. Rapid PCR amplification of minimal enediyne polyketide synthase cassettes leads to a predictive familial classification model. Proc. Natl. Acad. Sci. U.S.A. 100(21):11959-11963.

31. **Lopez, J. L.** 2003. Naturally mosaic operons for secondary metabolite biosynthesis; variability and putative horizontal gene transfer of discrete catalytic domains of the epothilone polyketide synthase locus. Mol. Gen. Genomics. 270:420-431.

32. **Maddison, D. R. and W. P. Maddison.** 2001. Macclade 4. Sunderland, MA, USA: Sinauer Associates.

33. **Magarvey, N. A., J. M. Keller, V. Bernan, M. Dworkin, and D. H. Sherman.** 2004. Isolation and characterization of novel marine-derived actinomycete taxa rich in bioactive metabolites. Appl. Environ. Microbiol. 70(12):7520-7529.

34. **Moffitt, M. C. and B. A. Neilan.** 2003. Evolutionary affiliations within the superfamily of ketosyntahses reflect complex pathway association. J. Mol. Evol. 56:446-457

35. **Oh, D.-C., P. G. Williams, C. A. Kauffman, P. R. Jensen, and W. Fenical.** 2006. Cyanosporosides A and B, unprecedented cyclopenta[a]indene glycosides from the obligate marine actinomycete *Salinispora pacifica*. Org. Lett. 8(6):1021-1024.

36. **Ravenschlag, K., K. Sahm, J. Pernthaler, and R. Amann.** 1999. High bacterial diversity in permanently cold marine sediments. Appl. Environ. Microbiol. 65(9):3982-3989.

37. **Schirmer, A., R. Gadkari, C. D. Reeves, F. Ibrahim, E. F. DeLong, and C. R. Hutchinson.** 2005. Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta*. Appl. Environ. Microbiol. 71(8):4840-4849.

38. **Stach, J. E. M., L. A. Maldonado, D. G. Masson, A. C. Ward, M. Goodfellow, and A. T. Bull.** 2003. Statistical approaches for estimating actinobacterial diversity in marine sediments. Appl. Environ. Microbiol. 69(10):6189-6200.

39. **Swofford, D. L**. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Software version 4.0b 10. Sinauer Associates, Sunderland, MA.

40. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876-4882.

41. **Tillett, D., E. Dittmann, M. Erhard, H. von Dohren, T. Borner and B. A. Neilan.** 2000. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system. Chem. Biol. 7:753-764.

42. **Walsh, C. T.** 2004. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. Science. 303:1805-1810.

43. **Zazopoulos, E., K. Huang, A. Staffa, W. Liu, B. O. Bachmann, K. Nonaka, J. Ahlert, J. S. Thorson, B. Shen, and C. M. Farnet.** 2003. A Genomics-Guided Approach for Discovering and Expressing Cryptic Metabolic Pathways. Nature Biotechnol. 21:187-190.

Table 3.1: List of the 60 actinomycete isolates screened for genes associated with type I PKS, enediyne PKS or NRPS pathways. The presence of sequence verified PCR products is indicated for each isolate by a plus sign. When no PCR product was observed, the result is indicated by a minus sign.

Phylogenetic group (family)	Representative isolate (accession no.)	Type I PKS	Enediyne PKS	NRPS
Brevibacteriaceae	CNJ737 (DQ448693)	-	-	-
Corynebacteriaceae	CNJ954 (DQ448694)	-	-	+
Dermacoccaceae	CNJ855 (DQ448695)	-	-	-
Dietziaceae	CNJ898 (DQ448696)	-	-	-
Geodermatophilaceae	CNJ868 (DQ448697)	-	-	-
	CNJ793 (DQ448698)	-	-	-
	CNJ794 (DQ448774)	-	-	-
Gordoniaceae	CNJ756 (DQ448699)	-	-	+
	CNJ863 (DQ448700)	+	-	+
	CNJ754 (DQ448701)	-	-	+
	CNJ752 (DQ448702)	-	-	+
Intrasporangiaceae	CNJ824 (DQ448703)	-	-	-
	CNJ927 (DQ448704)	+	-	-
Microbacteriaceae	CNJ745 (DQ448705)	-	-	-
	CNJ930 (DQ448706)	-	-	-
	CNJ743 (DQ448707)	-	-	-
	CNJ797 (DQ448708)	-	-	-
Micrococcaceae	CNJ723 (DQ448709)	-	-	-
	CNJ928 (DQ448783)	-	-	-
	CNJ900 (DQ448710)	-	-	-
	CNJ770 (DQ448711)	-	-	+
	CNJ787 (DQ448773)	-	-	+
	CNJ719 (DQ448712)	-	-	-
Micromonosporaceae	CNS326 (DQ448713)	-	-	+
	CNJ878 (DQ448714)	+	+	+
	CNS051 (DQ448715)	+	+	+
	CNS055 (DQ224159)	+	-	+
	CNS143 (DQ092624)	+	+	+
	CNS205 (CP000850)	+	+	+
	CNS237 (DQ318246)	+	+	+
Mycobacteriaceae	CNJ859 (DQ448716)	+	-	+
	CNJ823 (DQ448717)	+	-	+
Nocardiaceae	CNS044 (DQ448718)	+	-	+
Nocardioidaceae	CNJ889 (DQ448719)	-	-	-
	CNJ780 (DQ448720)	-	-	-
	CNJ872 (DQ448721)	-	-	+
	CNJ892 (DQ448722)	-	-	-
Nocardiopsaceae	CNR923 (DQ448723)	+	-	+
Promicromonosporaceae	CNJ734 (DQ448724)	-	-	-
Pseudonocardiaceae	CNJ888 (DQ448725)	-	-	+
	CNS139 (DQ448726)	-	-	+
	CNS004 (DQ448727)	-	-	+
Streptomycetaceae	CNJ962 (DQ448737)	+	-	+
	CNR872 (DQ448734)	+	-	+
	CNR879 (DQ448785)	+	-	+
	CNR877 (DQ448740)	-	-	+
	CNR880 (DQ448735)	-	-	-
	CNR881 (DQ448730)	+	-	+
	CNR884 (DQ448728)	-	-	+
	CNR885 (DQ448739)	+	-	+
	CNR887 (DQ448738)	+	-	+
	CNR918 (DQ448731)	+	-	+
	CNR924 (DQ448732)	+	-	+
	CNR925 (DQ448742)	+	-	+
	CNR926 (DQ448729)	-	-	+
	CNR940 (DQ448741)	+	-	+
	CNS177 (DQ448736)	+	-	+
	CNR927 (DQ448786)	+	-	+
	CNR876 (DQ448784)	+	-	+
Thermomonosporaceae	CNU125 (DQ448743)	+	-	-

Table 3.2: List of the 26 actinomycete isolates from which a type I PKS KS domain was amplified. The culture number and identity of the nearest type strain for each isolate are given. Type I PKS KS domain clone libraries were constructed for each isolate and the number of KS domains sequenced and analyzed is listed.

CN#	Nearest type strain	KSs seq'd
CNU-125	*Actinomadura cremea*	2
CNJ-863	*Gordonia nitida*	26
CNJ-878	*Micromonospora endolithica*	2
CNJ-859	*Mycobacterium brisbanense*	1
CNJ-823	*Mycobacterium porifera*	2
CNS-044	*Nocardia arthritidis*	13
CNR-923	*Nocardiopsis lucentensis*	3
CNS-051	*Salinispora arenicola*	2
CNS-205	*Salinispora arenicola*	47/51*
CNS-143	*Salinispora pacifica*	47
CNS-055	*Salinispora pacifica 'A'*	37
CNS-237	*Salinispora pacifica 'B'*	40
CNJ-927	*Serinicoccus marinus*	1
CNR-876	*Streptomyces aureofaciens*	1
CNR-881	*Streptomyces bikiniensis*	2
CNR-918	*Streptomyces caviscabies*	3
CNR-924	*Streptomyces chartreusis*	3
CNR-872	*Streptomyces hebeiensis*	2
CNS-177	*Streptomyces lydicus*	3
CNJ-962	*Streptomyces sampsonii*	1
CNR-887	*Streptomyces sampsonii*	1
CNR-879	*Streptomyces tendae*	2
CNR-885	*Streptomyces tendae*	41
CNR-925	*Streptomyces thermocoprophilus*	14
CNR-927	*Streptomyces thermocoprophilus*	8
CNR-940	*Streptomyces thermocoprophilus*	35

*: number of KS domain clones sequenced following a single PCR run/three pooled PCR runs

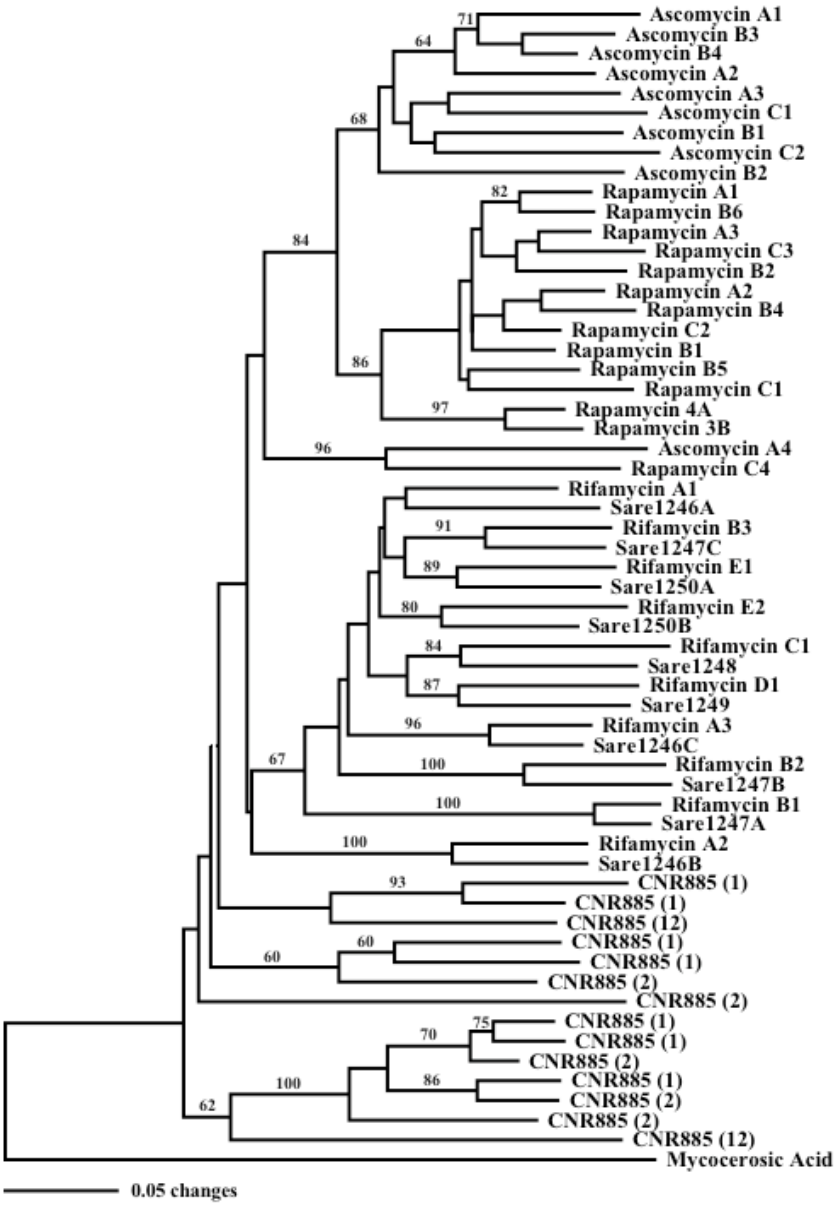

Figure 3.1: Neighbor-joining distance tree constructed in PAUP (39) using aligned KS domain sequences (223 amino acid positions) from CNR-885 and four type I PKS pathways from the reference data set. All KS domain sequences from the ascomycin, rapamycin and two rifamycin biosynthetic pathways (from *Amycolatopsis mediterranei* and *Salinispora arenicola*) are presented with a gene and module designation corresponding to Appendix B. From CNR-885, 14 unique KS domain sequences were identified; the total number of clones with that specific sequence is given in parentheses. A KS domain from the type I fatty acid synthase (FAS) involved in mycocerosic acid biosynthesis was used to position the root. Bootstrap values (in percent) calculated from 1000 re-samplings using the parsimony full heuristic search are shown at their respective nodes for values ≥ 60%. Although some within pathway branching patterns changed when distance neighbor-joining, distance UPGMA and maximum parsimony treeing methods were applied, the overall tree topology was maintained.

Figure 3.2: Neighbor-joining distance tree constructed in PAUP (39) using aligned KS domain sequences (223 amino acid positions) from CNR-925 and five type I PKS pathways from the reference data set. All KS domain sequences from the actinofuranone, ascomycin, rapamycin, rifamycin and tetronomycin pathways are presented with a gene and module designation corresponding to Appendix B. From CNR-925, six unique KS domain sequences were identified; the total number of clones with that specific sequence is given in parentheses. A KS domain from the type I fatty acid synthase (FAS) involved in mycocerosic acid biosynthesis was used to position the root. Bootstrap values (in percent) calculated from 1000 re-samplings using the parsimony full heuristic search are shown at their respective nodes for values ≥ 60%. Although some within pathway branching patterns changed when distance neighbor joining, distance UPGMA and maximum parsimony treeing methods were applied, the overall tree topology was maintained.

Figure 3.3: Neighbor-joining distance tree constructed in PAUP (39) using aligned KS domain sequences of type I PKS, type II PKS and enediyne PKS pathways from the genome sequence of *Salinispora arenicola* strain CNS-205. A total of 25 KS domains from type I PKS pathways are included, including ten type I modular KS domains from the rifamycin biosynthetic pathway, ten type I modular KS domains from an as yet unidentified macrolide, two type I iterative KS domains, and three KS domains that appear to be associated with the biosynthesis of hybrid PKS/NRPS compounds. For only one pathway associated with hybrid biosynthesis has a product, lymphostin, been identified. An asterisk denotes the seven KS domains cloned and sequenced during this study. The number of KS domains cloned is given in parentheses; the number of clones sequenced following a single PCR is on the left and the number of KS domains cloned after pooling three reactions is on the right. The two KS domains associated with enediyne PKS pathways were used to position the root.

Chapter Four: New PCR Primer Set For the Taxonomically Novel and Chemically Rich MAR2 Actinomycetes

Introduction.

The antibiotic streptothricin, discovered in 1942, was the first antibiotic to be elucidated from the actinomycete genus *Streptomyces*. However, it was the discovery of streptomycin by Selman Waksman in 1944 that triggered the systematic screening of the *Streptomyces* for antibiotics (16). For almost two decades, the number of antimicrobial compounds elucidated from the genus increased almost exponentially (16). Then in the 1970s, the discovery of novel secondary metabolites with antibiotic properties declined sharply, a decline linked to the frequent recovery of known metabolites from what had become common terrestrial *Streptomyces* isolates. The frequent re-isolation of known metabolites coupled with burgeoning techniques such as combinatorial biosynthesis (2, 7) and combinatorial catalysis (11), drew effort and enthusiasm away from drug discovery efforts that relied heavily on the isolation and elucidation of secondary metabolites from cultured *Streptomyces*.

Recently, several factors have encouraged a renewed focus on the isolation of rare and novel actinomycetes, especially *Streptomyces*, in the search for novel natural products of biomedical relevance. First and foremost, infectious diseases continue to pose a significant health threat. A rise in the prevalence of drug-resistant pathogens and immunodeficiency diseases, coupled with the use of immunocompromising treatments, has contributed to the growing need for new antimicrobial drug therapies (3, 4, 8, 15, 17). Another reason for a renewed focus on the isolation of antibiotic producing actinomycetes stems from the limited success of alternative discovery

platforms. Strategies such as combinatorial chemistry and high-throughput screening have simply not met expectations.

The advent of microbial genome sequencing has also encouraged a new cycle of natural products research by providing significant evidence that a plethora of secondary metabolites have yet to be recovered, particularly from well-studied bacterial isolates. For example, the recent genomic analysis of *Streptomyces coelicolor* strain A3(2) revealed clusters of genes characteristic of biosynthetic pathways responsible for the production of never before isolated secondary metabolites (1). Although *Streptomyces coelicolor* had been studied for over forty years, the strain was only known to produce three secondary metabolites. The genome sequence of *Streptomyces coelicolor* strain A3(2) suggested that the strain has the capacity to produce at least 18 additional compounds. Discoveries such as these are encouraging researchers to look more closely at other well-studied and newly identified *Streptomyces* in the search for novel secondary metabolites of biomedical relevance (1). Genomic analyses suggest that many natural products have yet to be discovered and these findings corroborate results from mathematical models that suggest researchers have only unearthed a tiny fraction of the antimicrobial compounds and other secondary metabolites that could be produced by actinomycetes. One statistical model estimated that the genus *Streptomyces* alone is capable of producing on the order of 100,000 antimicrobial compounds, an estimate which suggests we have only just scratched the surface in our discovery of secondary metabolites (16).

The isolation of novel actinomycete diversity, especially the isolation of novel *Streptomyces* and other closely related strains, provides an important means to discover novel secondary metabolites. Ocean sediments are a rich source of novel actinomycetes and many of these isolates produce novel secondary metabolites of biomedical relevance. Actinomycetes of the MAR2 group are an example of one such phylogenetically unique group of isolates with the ability to produce new small molecules with both antibiotic and anticancer activities.

Actinomycetes of the MAR2 group are most closely related to members of the genus *Streptomyces* and cluster together during phylogenetic analyses regardless of their geographical distribution. MAR2 strains isolated from sediments collected in San Diego, Catalina, the Sea of Cortez, the Bahamas, Florida, Guam and Palau cluster tightly together on phylogenetic trees constructed using nearly complete 16S rRNA gene sequences. At the time this bacterial group was discovered no terrestrial relative had ever been sequenced and therefore they were designated a MAR, or marine, group. As the group found after the MAR1, later described as the genus *Salinispora*, the group was designated the MAR2. Since the first MAR2 strains were isolated, closely related terrestrial strains have been identified through 16S rRNA gene sequencing. Although the MAR2 group now includes non-marine strains, all MAR2 strains share low 16S rRNA gene sequence identity with their most closely related type strains suggesting that these isolates may be a new *Streptomyces* species or represent several species of a new genus within the family *Streptomycetaceae*. Our recent collaborative efforts with researchers at the German collection of microorganisms and cell cultures (also known as the Deutsch Sammlung von

Mikroorganismen und Zellkulturen or DSMZ) and at Newcastle University suggest that these isolates may represent multiple species of a new genus within the family *Streptomycetaceae*.

In addition to generating interest among microbiologists and bacterial taxonomists, MAR2 strains have been rigorously studied by chemists, yielding an array of secondary metabolites from these bacteria, many with antibiotic and anticancer properties. Three strains, one isolated from sediments collected off San Diego (CNQ-140) and two isolated from samples collected off Palau (CNJ-923 and CNJ-962), produce marinomycins A-G and marinosporolides A-E. Marinomycins A-D exhibited significant antibiotic activity, with minimum inhibitory concentration (MIC) values of 0.1-0.6 μM, against methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *Enterococcus faecium* (VREF). When tested against the NCI's 60 cancer cell line panel, the marinomycins inhibited cancer cell proliferation with an average LC_{50} of 0.2-2.7 μM and also demonstrated potent and selective cytotoxicity against six of eight melanoma cell lines (9). Other MAR2 strains, such as the San Diego isolate CNQ-259, have been found to produce known antifungal and anticancer compounds, including strevertene D (personal communication H. C. Kwon).

The sheer number and diversity of the natural products produced by this group of strains has encouraged the search for additional MAR2 isolates. Although MAR2 isolates have been recovered from geographically diverse locations, our research team had only identified 19 MAR2 strains prior to the current study. In an effort to search a collection of cultured strains and quickly identify MAR2 isolates and also to prepare

to quickly identify sediments harboring MAR2 strains such that they could be targeted for rigorous processing, a set of PCR primers specific for MAR2 strains was developed. To date, the primer set has been used successfully to identify additional MAR2 isolates from a collection of marine isolates and could be used in the future to identify sediments harboring MAR2 strains, suggesting which sediment samples to target in an attempt to maximize the number of MAR2 isolates cultured.

Materials and Methods.

MAR2 bacterial strains used in this study were cultured from marine sediments collected in San Diego, Catalina, the Sea of Cortez, Florida, Guam and Palau as previously described (5, 6, 12). All strains were grown on A1 agar medium (18 g agar, 10 g starch, 4 g yeast extract, 2 g peptone, 1 liter natural seawater) amended with 0.2-μm filtered cycloheximide (100 μg/ml) after autoclaving. Strains were grown for three weeks and then transferred into A1 liquid medium for an additional 2 weeks. DNA was extracted according to the DNeasy protocol (Qiagen Inc., Valencia, Calif.) with the following modifications. After RNase A (2 mg/ml) was added to the enzymatic lysis buffer, the resuspended bacterial pellet was incubated for two hours at 37^{o}C. Following the addition of proteinase K and buffer AL, the sample was held for one hour at 70^{o}C. Genomic DNA was eluted from the spin column with 100 μl of buffer AE for immediate use or storage at –20^{o}C.

The 16S rRNA gene sequences of 19 MAR2 isolates were PCR amplified using the primers FC27 (5'-AGAGTTTGATCCTGGCTCAG-3') and RC1492 (5'-TACGGCTACCTTGTTACGACTT-3'). The 50 μl PCR reaction mixture contained

20 to 50 ng of DNA, 250 pmol of each primer, ThermoPol Buffer (New England BioLabs Inc., Beverly, Mass.), 2.5 U of TaqDNA polymerase (New England BioLabs Inc., Beverly, Mass.), and 100 μM deoxynucleoside triphosphate mixture. The PCR program consisted of 30 cycles of 1 min denaturation at 94°C, 1 min annealing at 60°C, and 1 min extension at 72°C, followed by a final extension step at 72°C for 7 minutes. Amplification products were examined by agarose gel electrophoresis and purified using Qiagen's QIAquick PCR cleanup kit according to the manufacturer's suggested protocol (Qiagen Inc., Valencia, Calif.). Nearly complete sequences were obtained for the 16S rDNA amplicons (*E. coli* numbering 20-1392) using six primers: FC27, RC1492, R530 (5'-CCGCGGCTGCTGGCACGTA-3'), R936 (5'-GTGCGGGCCCCCGTCAATT-3'), F514 (5'-GTGCCAGCAGCCGCGGTAA-3'), and F1114 (5'-GCAACGAGCGCAACCC-3'). Sequencing reactions were carried out with an ABI 3100 DNA sequencer at the DNA Sequencing Shared Resource, UCSD Cancer Center (funded in part by the NCI Cancer Center Support Grant #2 P30CA23100-18).

All nucleotide sequences were assembled, analyzed and manually edited using the Sequencher software package (version 4.5, Gene Codes Co., Ann Arbor, Mich.) and compared to sequences within the NCBI database (http://www.ncbi.nlm.nih.gov/) using the Basic Local Alignment Search Tool (BLAST). The 16S rRNA gene sequences and their most closely related type strains were imported into ARB (10) and aligned. Using ARB, the aligned 16S rRNA gene sequences of MAR2 strains, the most closely related type strains, a large number of characterized actinomycetes, other Gram-positive bacteria and a handful of Gram-negative bacteria were used to design a

set of PCR primers specific to MAR2 actinomycetes. The nearly complete, aligned 16S rRNA gene sequences of the MAR2 strains were also used to construct a phylogenetic tree using the distance neighbor-joining, distance unweighted-pair group method using average linkages (UPGMA), and maximum parsimony methods in PAUP (14). Nearly complete, aligned 16S rRNA gene sequences were also analyzed using the Clusterer program (http://www.bugaco.com/bioinf) and the number of operational taxonomic units (OTUs) calculated using ≥ 98% and ≥ 99% sequence identity values.

The MAR2 specific PCR primers were used to screen 24 isolates stored in a culture collection. These strains were Gram-positive, filamentous actinomycete-like isolates cultured from marine sediments collected during a 1996 research expedition off San Diego and a 2002 research expedition to the Bahamas. These isolates were removed from storage at –80°C and grown on A1 agar medium (18 g agar, 10 g starch, 4 g yeast extract, 2 g peptone, 1 liter natural seawater) amended with 0.2-μm filtered cycloheximide (100 μg/ml) after autoclaving. Strains were grown for three weeks and then transferred into A1 liquid medium for an additional 2 weeks. Genomic DNA was extracted and the 16S rRNA gene sequences determined as described above. Prior to 16S rRNA gene sequencing, the MAR2 specific primers were used to predict the number of MAR2 strains present among the strains from the culture collection being sequenced.

Results and Discussion.

The 16S rRNA genes of 19 MAR2 bacterial strains, cultured from marine sediments collected in San Diego, Catalina, the Sea of Cortez, Florida, Guam and Palau, were sequenced and used to design a PCR primer set specific for the MAR2 bacterial group. The set consisted of a MAR2 specific forward PCR primer, 811F (5'-TGTAAACGGTGGGAACTAGGTGTGGGCAG-3'), and the Gram-positive bacterial specific reverse PCR primer RC1492 (5'-TACGGCTACCTTGTTACGACTT-3'). Identifying MAR2 specific regions for primer design proved to be quite difficult. The PCR protocol and conditions were ultimately optimized for primers 811F and RC1492 and after testing the PCR primers on MAR2 strains and on both closely related and distantly related bacterial strains, a single fragment approximately 680 bp in length was amplified from MAR2 strains and not from other closely related or distantly related isolates. The optimized 50 μl PCR reaction mixture contained 20 to 50 ng of DNA, 250 pmol of each primer, ThermoPol Buffer (New England BioLabs Inc., Beverly, Mass.), 1.5 U of TaqDNA polymerase (New England BioLabs Inc., Beverly, Mass.), 100 μM deoxynucleoside triphosphate mixture and 5% DMSO. The PCR program consisted of a 5 min denaturation at 94^{o}C, 35 cycles of 1 min denaturation at 94^{o}C, 1 min annealing at 66.5^{o}C and 1 min extension at 72^{o}C, followed by a final extension step of 7 min at 72^{o}C.

Using the MAR2 specific PCR primers, 24 Gram-positive, filamentous actinomycete-like isolates from a culture collection were screened. Of the 24 isolates screened, 3 yielded an approximately 680 bp PCR product. Nearly complete 16S rRNA gene sequencing confirmed that the three strains, CNB-984, CNR-954 and CNR-955, are MAR2 strains and were the only MAR2 strains among the isolates

screened from the culture collection. The positive results suggest that a PCR based screen for MAR2 bacteria is an effective way to search a culture collection for additional members of this chemically prolific group of microorganisms. The method also encourages future testing and use of this method as an effective way to quickly identify sediments harboring MAR2 strains. By extracting environmental DNA from the sediments and using the MAR2 specific primers to detect the presence of MAR2 bacteria within a sediment sample, it may be possible to suggest which sediments should become the focus of more rigorous processing. While possible that the PCR signal could be too weak to be detected if the number of MAR2 bacteria within the sediment sample is low, the technique might be effective in combination with the use of enrichment cultures. With ways to suppress the Gram-negative bacterial population and promote the growth of Gram-positive bacterial populations, it might be possible to enrich for MAR2 bacteria and detect such successful enrichments using the PCR based technique. Another potential drawback of using this technique on environmental samples could be the amplification of non-MAR2 sequences that share sequence identity with MAR2 in the forward primer region. While the forward primer does not currently match any sequenced bacterial 16S rRNA gene sequences not already predicted to be MAR2 bacteria, there is a possibility that the PCR primers could yield false positive results when used with environmental samples. False positives could be viewed as a highly desirable result however, as more rigorous processing of those samples could lead to the isolation of many as yet uncultured bacteria from marine sediments, adding to a growing understanding of the bacterial diversity present within marine sediments and the world's oceans.

The MAR2 specific PCR primers were successfully used to identify three additional MAR2 strains from a culture collection. To date a total of 22 MAR2 strains have been identified (Table 4.1). The primer set could be tested and used in the future to identify sediments that harbor MAR2 strains, directing and focusing sediment processing efforts following research expeditions in an effort to culture additional strains from this chemically rich group of bacteria.

The diversity of the 22 cultured MAR2 bacteria was estimated by performing cluster analyses using the nearly complete 16S rRNA gene sequences. Six OTUs were identified using a sequence identity value of ≥ 98% whereas eight OTUs were identified using a sequence identity value of ≥ 99%. MAR2 bacteria are only distantly related to their nearest sequenced type strains (Figure 4.1) and at the time this thesis was written, in collaboration with scientists at Newcastle-upon-Tyne we are investigating the possibility that the 22 MAR2 strains represent a new genus within the family *Streptomycetaceae* consisting of at least six species. Although Stach *et al.* (13) suggested that a 16S rRNA gene sequence identity value of ≥ 99% could be used to define actinomycete OTUs, in which case the MAR2 strains could represent eight different species, the use of a sequence identity value of ≥ 98% provides a more conservative estimate.

The chemistry and phylogenetic novelty of the 22 MAR2 strains has attracted the attention of bacterial taxonomists in both Germany and Newcastle-upon-Tyne. The identification of a new genus within this family would be exceptionally exciting and the work so far has suggested that the MAR2 may warrant the description of a new genus. To date, researchers at Newcastle-upon-Tyne have completed scanning

electron microscopy as well as morphology and phenotypic tests in preparation for a formal taxonomic description of the MAR2 bacteria. In the near future, additional tests including an assessment of the fatty acid and menaquinone composition of the MAR2 strains will be determined in order to complete the tests necessary to write the taxonomic description. The MAR2 bacteria continue to be the focus of chemical investigations and with the possibility of representing a new genus, identifying additional new MAR2 strains will continue to be important.

References.

1. **Bentley, S. D., K. F. Chater, A.-M. Cerdeño-Tárraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C.-H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M.-A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill, and D. A. Hopwood.** 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature. 417:141-147.

2. **Cane, D. E., C. T. Walsh and C. Khosla.** 1998. Harnessing the biosynthetic code: combinations, permutations and mutations. Science. 282:63-68.

3. **Cassell, H.** 1997. Emergent antibiotic resistance: health risks and economic impact. FEMS Immunol. Med. Microbiol. 18:271-274.

4. **Culotta, E.** 1994. Funding crunches hobbles antibiotic resistance research. Science. 264:362-363.

5. **Gontang, E. A., W. Fenical, and P. R. Jensen.** 2007. Phylogenetic diversity of gram-positive bacteria cultured from marine sediments. Appl. Environ. Microbiol. 73: 3272-3282.

6. **Jensen, P. R., E. Gontang, C. Mafnas, T. J. Mincer and W. Fenical**. 2005. Culturable marine actinomycete diversity from tropical Pacific Ocean sediments. Environ. Microbiol. 7:1039-1048.

7. **Hutchinson, C. R.** 1998. Combinatorial biosynthesis for new drug discovery. Curr. Opin. Microbiol. 1:319-329.

8. **Kerr, K. G. and R. W. Lacey.** 1995. Why do we still get epidemics? p.179-203. *In*: Hunter P. A., G. K. Darby, and N. J. Russell (ed.), Fifty years of antimicrobials: past perspective and future trends. SGM symposium 53. Cambridge University Press, Cambridge.

9. **Kwon, H. C., C. A. Kauffman, P. R. Jensen and W. Fenical.** 2006. Marinomycins A-D, antitumor-antibiotics of a new structure class from a marine actinomycete of the recently discovered genus "*Marinispora*". J. Am. Chem. Soc. 128:1622-1632.

10. **Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Förster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lüßmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode and K.-H. Schleifer**. 2004. ARB: a software environment for sequence data. Nucleic Acids Res. 32:1363-1371.

11. **Michels, P. C., Y. L. Khmelnitsky, J. S. Dordick and J. S. Clark**. 1998. Combinatorial biocatalysis: a natural approach to drug discovery. Trends Biotechnol. 16:210-215.

12. **Prieto-Davó, A., W. Fenical and P. R. Jensen.** 2008. Comparative actinomycete diversity in marine sediments. Aquatic Microbial Ecology. 52(1):1-11.

13. **Stach, J. E. M., L. A. Maldonado, D. G. Masson, A. C. Ward, M. Goodfellow and A. T. Bull**. 2003. Statistical approaches for estimating actinobacterial diversity in marine sediments. Appl. Environ. Microbiol. 69:6189-9200.

14. **Swofford, D. L**. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Software version 4.0b 10. Sinauer Associates, Sunderland, MA.

15. **Travis, J.** 1994. Reviving the antibiotic miracle? Science. 264:360-362.

16. **Watve, M. G., R. Tickoo, M. M. Jog and B. D. Bhole.** 2001. How many antibiotics are produced by the genus Streptomyces? Arch. Microbiol. 176:386-390.

17. **Zaehner, H. and H. Fiedler.** 1995. The need for new antibiotics: possible ways forward. p.67-85. *In*: Hunter P. A., G. K. Darby, and N. J. Russell (ed.), Fifty years of antimicrobials: past perspective and future trends. SGM symposium 53. Cambridge University Press, Cambridge.

Table 4.1: List of the 22 MAR2 strains sequenced, their collection location, nearest type strain and their percent identity with the nearest type strain as determined by BLAST analysis.

CN #	Collection Location	Nearest Type Strain	Percent Identity
CNB-984*	San Diego	*Streptomyces armeniacus* (AB018093)	95%
CNJ-923	Palau	*Streptomyces endus* (AY999911)	94%
CNJ-962	Palau	*Streptomyces endus* (AY999911)	94%
CNP-027	Sea of Cortez	*Streptomyces sodiiphilus* (AY236339)	94%
CNQ-140	San Diego	*Streptomyces endus* (AY999911)	94%
CNQ-189	San Diego	*Streptomyces sodiiphilus* (AY236339)	94%
CNQ-233	San Diego	*Streptomyces endus* (AY999911)	94%
CNQ-259	San Diego	*Streptomyces endus* (AY999911)	96%
CNQ-695	Guam	*Streptomyces endus* (AY999911)	95%
CNQ-702	Guam	*Streptomyces sampsonii* (D63871)	95%
CNQ-703	Guam	*Streptomyces endus* (AY999911)	95%
CNQ-732	Guam	*Streptomyces endus* (AY999911)	95%
CNR-252	Guam	*Streptomyces sampsonii* (D63871)	95%
CNR-872	Palau	*Streptomyces sodiiphilus* (AY236339)	94%
CNR-954*	Bahamas	*Streptomyces endus* (AY999911)	95%
CNR-955*	Bahamas	*Streptomyces endus* (AY999911)	95%
CNS-560	San Diego	*Streptomyces armeniacus* (AB018093)	95%
CNS-562	Florida	*Streptomyces armeniacus* (AB018093)	95%
CNS-563	Florida	*Streptomyces armeniacus* (AB018093)	95%
CNS-565	Guam	*Streptomyces sodiiphilus* (AY236339)	95%
CNS-566	San Diego	*Streptomyces armeniacus* (AB018093)	95%
CNS-567	Catalina Island	*Streptomyces armeniacus* (AB018093)	95%

*isolated as a part of this study

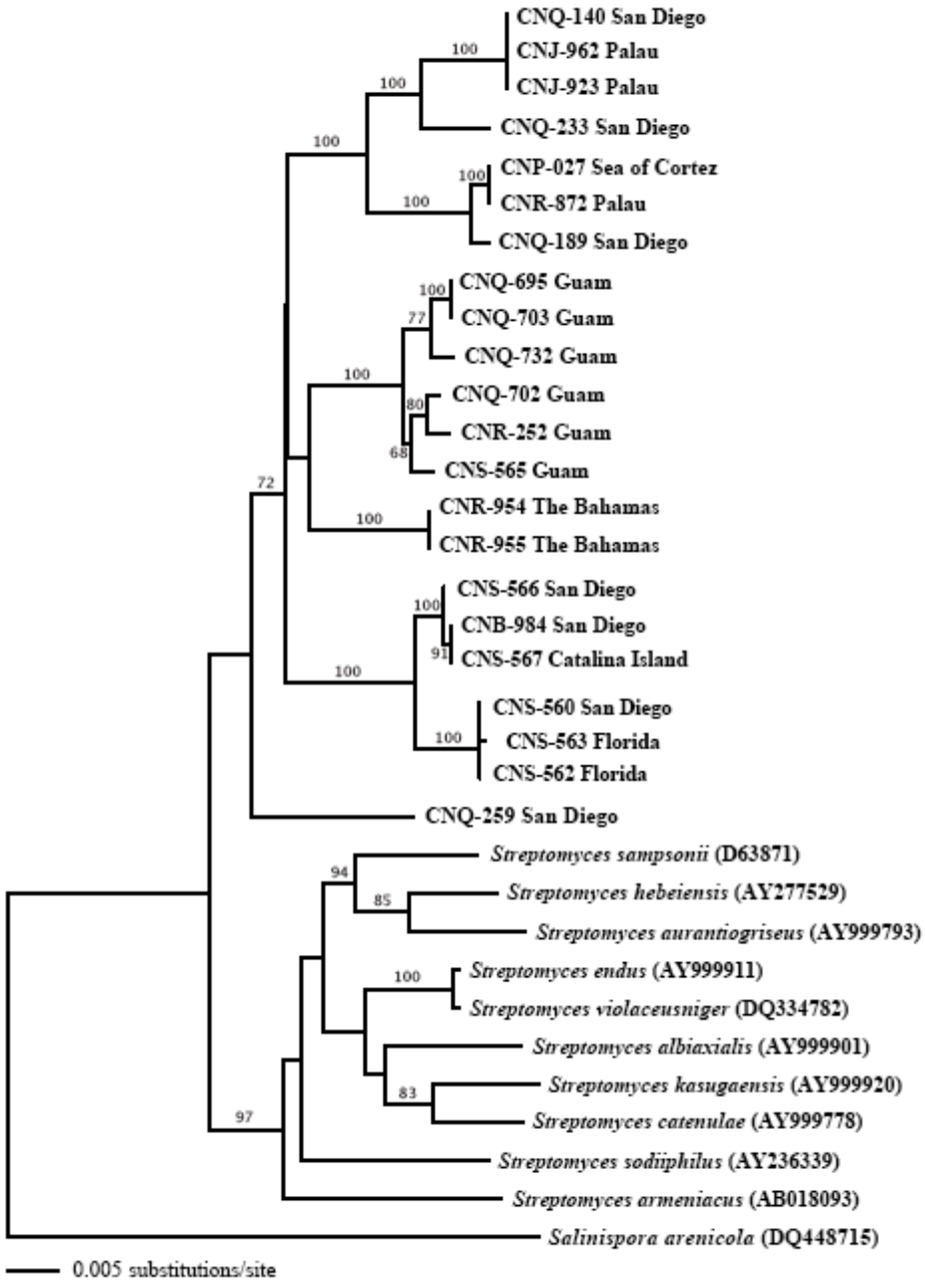

Figure 4.1: Neighbor-joining distance tree constructed in PAUP (14) using the aligned 16S rRNA gene sequences of 22 MAR2 strains and the 10 nearest type strains as determined by BLAST. The culture number of each MAR2 strain as well as the location from which it was collected is provided. Bootstrap values (in percent) calculated from 1,000 resamplings using the neighbor-joining method are shown at the nodes from values greater than or equal to 60%. The 16S rRNA gene sequence of *Salinispora arenicola* strain CNS-205 was used to position the root. The branching pattern did not change when distance UPGMA and maximum-parsimony treeing methods were used.

Conclusion.

A culture-dependent approach was successfully used to assess the diversity of Gram-positive bacteria in marine sediments collected around the islands of Palau. The survey of the total, aerobic Gram-positive bacterial diversity present in tropical marine sediments led to the isolation of 78 operational taxonomic units cultured (OTUs based on ≥ 98% 16S rRNA gene sequence identity), 52 of which were determined to be members of the order *Actinomycetales* while the other 26 were determined to be members of the order *Bacillales*. There is currently much to learn about Gram-positive bacteria in marine sediments. Like their terrestrial relatives, marine Gram-positive bacteria may play a significant role in the breakdown of recalcitrant organic matter and may therefore play an integral role in the ocean's biogeochemical cycle. These marine microorganisms may also act as new sources of bioactive secondary metabolites. It is clear from this single survey that considerable new Gram-positive bacterial diversity can be readily cultured from marine sediments. The continued use of cultivation-dependent techniques will undoubtedly lead to the discovery of additional Gram-positive diversity, and provide a direct means to learn more about their ecophysiology and applications in biotechnology.

Bacterial genome sequencing has provided opportunities to assess the secondary metabolite biosynthetic potential of individual strains. Like the genome of *Streptomyces coelicolor* strain A3(2), the genome sequence of *Salinispora arenicola* strain CNS-205 suggested that the secondary metabolite biosynthetic potential of this species has not yet been fully recognized. *S. arenicola* harbors a wealth of diverse and

novel secondary metabolite gene clusters including PKS, NRPS and hybrid PKS/NRPSbiosynthetic gene clusters. By elucidating entire biosynthetic pathways in the context of all other genes and pathways present, it is also possible to address questions related to the evolution of gene clusters or individual genes within a cluster. The complete genome sequence of *Salinispora arenicola* strain CNS-205 provided an opportunity to investigate the evolution of type I modular PKS gene clusters. The putative evolutionary history of an as yet unidentified macrolide was discussed as were the evolutionary forces affecting the putative rifamycin biosynthetic gene cluster. By analyzing the two type I modular PKS pathways of *S. arenicola*, evidence of gene duplication, gene loss, rearrangement, inversion, recombination, domain acquisition and horizontal gene transfer (HGT) were identified.

While a tremendous amount can be learned from analyzing the entire genetic blueprint of a microorganism, the sequencing of entire genomes remains costly and time consuming. In an effort to gain insight into the biosynthetic potential of unsequenced microorganisms prior to fermentation, extraction and chemical analyses, additional natural product discovery approaches are necessary. A PCR based approach was used successfully to suggest which representative isolates from the 52 *Actinomycetales* OTUs harbored type I PKS, enediyne PKS and NRPS pathways. Over half of the cultured *Actinomycetales* OTUs were found to possess genes associated with at least one PKS or NRPS biosynthetic pathway and although some of these actinomycetes represent families well known to produce secondary metabolites, the results suggest that novel secondary metabolites can be isolated from both filamentous and unicellular actinomycete families. In addition to determining if these

pathways are present, a phylogenetic approach was used successfully to suggest the number and novelty of type I PKS pathways present. By using the phylogenetic approach, the diversity, novelty and identity of type I PKS metabolites can be predicted.

The future of natural product drug discovery will undoubtedly continue to be influenced by advances in our understanding of the molecular genetics of secondary metabolite biosynthesis. Sequence-based approaches will likely be incorporated into discovery paradigms and provide perspective on the biosynthetic potential of an organism prior to fermentation, extraction and chemical analysis. As these methods are developed and refined, it is conceivable that the stochastic approach by which microorganisms have traditionally been cultured and screened will soon be abandoned in favor of genetic approaches that provide insight into the organism's genetic potential. Chapter three provided evidence that tools are now available to assess the potential utility of phylogenetic prediction as a method for improving the process of secondary metabolite discovery.

In addition to assessing the capacity of actinomycetes to produce secondary metabolites, there has also been interest in isolating additional strains from actinomycete groups known to produce secondary metabolites of biomedical relevance. The MAR2 bacteria are one such group generating interest among microbiologists and chemists as they may represent a new genus within the family *Streptomycetaceae* and are well known to produce an exciting array of secondary metabolites, many with antibiotic and anticancer properties. The sheer number and diversity of natural products produced by MAR2 bacteria, in addition to their

taxonomic novelty, has encouraged the search for additional isolates. In an effort to increase the number of MAR2 strains available for study, a set of PCR primers specific for MAR2 strains was developed to screen strain libraries as well as environmental DNA extracted from collected sediments. The MAR2 specific PCR primers were successfully used to identify additional MAR2 strains from a culture collection and to date, a total of 22 MAR2 strains have been identified. The primer set is now available to be tested and used to identify sediments samples that harbor MAR2 strains. By determining which sediments harbor MAR2 bacteria prior to processing, those sediment samples can be more rigorously processed in an attempt to maximize the number of MAR2 isolates cultured. MAR2 bacteria continue to be the focus of chemical investigations and with the possibility of representing a new genus, identifying additional strains continues to be important.

The isolation of rare and novel microorganisms from marine sediments has increased our understanding of the abundance and diversity of microorganisms present in our oceans. The cultivation of rare and novel Gram-positive bacteria is also of interest to researchers searching for new resources of natural products with biomedical relevance. Using the actinomycetes cultured during a research expedition to Palau in 2004, the evolutionary history of type I PKS genes and gene clusters was investigated, a PCR-based approach coupled with phylogenetic analyses was used to assess the capacity strains have to produce certain classes of secondary metabolites without the luxury of a large sequencing effort, and a new PCR primer set was designed to help future research expeditions increase their ability to isolate MAR2 bacteria, a group of bacteria known to make diverse compounds of biomedical relevance. Future culture-

dependent assessments of bacteria and future drug discovery efforts will continue to use and design new molecular tools to study and explore diverse microorganisms and their chemistry. The hope is that the results of studies described within this dissertation provide a portion of the foundation upon which future research efforts will be built.

Appendix A: Information on the 189 Gram-positive isolates, including collection depth, isolation method and medium, seawater requirement, and nearest type strain.

Media	Constituents
A1	18 g agar, 10 g starch, 4 g yeast extract, 2 g peptone, 1 liter natural seawater and rifampicin (5 ug/ml)
A2	18 g agar, 1 g starch, 0.4 g yeast extract, 0.2 g peptone, 1 liter natural seawater
A3	18 g agar, 2.5 g starch, 1 g yeast extract, 0.5 g peptone, 0.2 g glycerophosphate, 750 ml natural seawater and 250 ml DI water
A4	18 g agar and 1 liter natural seawater
A5	18 g agar, 750 ml natural seawater and 250 ml DI water
A6	18 g agar, 1 liter natural seawater and polymixin B sulfate (5 ug/ml)
A7	18 g agar, 1 liter natural seawater and kanamycin (5 ug/ml)
A8	18 g agar, 1 liter natural seawater and novobiocin (25 ug/ml)
A9	18 g agar, 1 liter natural seawater and rifampicin (5 ug/ml)
A10	8 g noble agar, 0.5 g mannitol, 0.1 g peptone, 1 liter natural seawater and rifampicin (5 ug/ml)
A11	18 g agar, 1 g KNO3, 0.1 g MgSO4-7H2O, 2 g Na2HPO4-7H2O, 0.14 g KH2PO4, 1 g NaCl, 1 liter DI water and 5 ml light, liquid paraffin*
	*paraffin added to molten agar after autoclaving
Method	Procedure
1	Ten ml of wet sediment were dried overnight and stamped onto agar media using a polyester fiber tipped sterile swab; swab was pressed to the dried sediment and then to the agar surface 35 to 40 times to create a serial dilution effect.
2	Sediments were diluted (1:4) in autoclaved, filter sterilized seawater and then heat shocked for 6 min at 55°C. Following 30 s of vigorous shaking, 50 ul of the resulting suspension was plated directly.
3	Sediments were diluted (1:4) in autoclaved, filter sterilized seawater and then heat shocked for 6 min at 55°C. Following 30 s of vigorous shaking, the resulting suspension was diluted (1:4) before 50 ul were used to inoculate agar media.
4	Pour plates were prepared by adding 0.5 g of wet sediment to 25 ml of autoclaved, molten (~42°C), 100% seawater agar amended with cyclohexamide (100 ug/ml) and rifampicin (5 ug/ml).
5	0.5 g dried sediment was added to 4 ml of autoclaved, filter-sterilized seawater without kanamycin (5 ug/ml). After vigorous shaking for 30 s, the sediment was allowed to settle for five minutes and then 50 ul were inoculated onto agar media and spread with an alcohol-sterilized glass rod.
6	0.5 g dried sediment was added to 4 ml of autoclaved, filter-sterilized seawater with kanamycin (5 ug/ml). After vigorous shaking for 30 s, the sediment was allowed to settle for five minutes and then 50 ul were inoculated onto agar media and spread with an alcohol-sterilized glass rod.

CN#	Depth (m)	Method	Media	SW Tested	SW Req.	Class	Family	Nearest Type Strain
U125	7	1	A9	Yes	No	*Actinobacteria*	*Thermomonosporaceae*	*Actinomadura cremea* (AF134067)
J889	2	1	A5	Yes	No	*Actinobacteria*	*Nocardioidaceae*	*Aeromicrobium erythreum* (AF005021)
J745	18	1	A1	No	No	*Actinobacteria*	*Microbacteriaceae*	*Agromyces aurantiacus* (AF389342)
J802	23	1	A6	No	No	*Actinobacteria*	*Microbacteriaceae*	*Agromyces aurantiacus* (AF389342)
J868	200	1	A5	Yes	No	*Actinobacteria*	*Geodermatophilaceae*	*Blastococcus saxobsidens* (AJ316570)
J737	1	2	A3	Yes	No	*Actinobacteria*	*Micrococcineae*	*Brevibacterium epidermidis* (X76565)
J954	37	1	A7	Yes	No	*Actinobacteria*	*Corynebacteriaceae*	*Corynebacterium variabile* (AJ222815)
J784	10	1	A2	Yes	Yes	*Actinobacteria*	*Dietziaceae*	*Dietzia maris* (X79290)
J790	6	1	A6	Yes	No	*Actinobacteria*	*Dietziaceae*	*Dietzia maris* (X79290)
J807	32	1	A6	Yes	No	*Actinobacteria*	*Dietziaceae*	*Dietzia maris* (X79290)
J820	23	1	A6	Yes	No	*Actinobacteria*	*Dietziaceae*	*Dietzia maris* (X79290)
J865	14	1	A7	Yes	Yes	*Actinobacteria*	*Dietziaceae*	*Dietzia maris* (X79290)
J871	14	1	A7	Yes	No	*Actinobacteria*	*Dietziaceae*	*Dietzia maris* (X79290)
J898	25	1	A6	Yes	Yes	*Actinobacteria*	*Dietziaceae*	*Dietzia maris* (X79290)
U124	6	1	A6	Yes	No	*Actinobacteria*	*Dietziaceae*	*Dietzia maris* (X79290)
J756	1	1	A1	Yes	No	*Actinobacteria*	*Gordoniaceae*	*Gordonia bronchialis* (X79287)
J786	5	1	A8	Yes	No	*Actinobacteria*	*Gordoniaceae*	*Gordonia bronchialis* (X79287)
J788	5	1	A10	Yes	No	*Actinobacteria*	*Gordoniaceae*	*Gordonia bronchialis* (X79287)
J798	1	1	A5	Yes	No	*Actinobacteria*	*Gordoniaceae*	*Gordonia bronchialis* (X79287)
J810	2	1	A6	Yes	No	*Actinobacteria*	*Gordoniaceae*	*Gordonia bronchialis* (X79287)
J863	33	1	A9	Yes	No	*Actinobacteria*	*Gordoniaceae*	*Gordonia nitida* (AF148947)
J753	1	1	A1	Yes	No	*Actinobacteria*	*Gordoniaceae*	*Gordonia polyisoprenivorans* (Y18310)
J754	1	1	A1	Yes	No	*Actinobacteria*	*Gordoniaceae*	*Gordonia polyisoprenivorans* (Y18310)
J755	1	1	A1	Yes	No	*Actinobacteria*	*Gordoniaceae*	*Gordonia polyisoprenivorans* (Y18310)
J752	5	1	A8	Yes	No	*Actinobacteria*	*Gordoniaceae*	*Gordonia terrae* (X81922)
J827	14	1	A6	Yes	No	*Actinobacteria*	*Gordoniaceae*	*Gordonia terrae* (X81922)
J832	14	1	A6	Yes	No	*Actinobacteria*	*Gordoniaceae*	*Gordonia terrae* (X81922)
J913	14	1	A6	Yes	No	*Actinobacteria*	*Gordoniaceae*	*Gordonia terrae* (X81922)
J723	3	3	A3	Yes	No	*Actinobacteria*	*Micrococcaceae*	*Kocuria marina* (AY211385)
J928	500	1	A6	Yes	No	*Actinobacteria*	*Micrococcaceae*	*Kocuria marina* (AY211385)
J900	8	1	A6	Yes	No	*Actinobacteria*	*Micrococcaceae*	*Kocuria palustris* (Y16263)
J770	20	1	A7	Yes	Yes	*Actinobacteria*	*Micrococcaceae*	*Kocuria rosea* (X87756)
J787	5	1	A10	Yes	No	*Actinobacteria*	*Micrococcaceae*	*Kocuria rosea* (X87756)
J855	45	1	A6	Yes	No	*Actinobacteria*	*Dermabacteraceae*	*Kytococcus sedentarius* (X87755)
J861	14	1	A7	Yes	No	*Actinobacteria*	*Dermabacteraceae*	*Kytococcus sedentarius* (X87755)
J867	15	1	A9	Yes	Yes	*Actinobacteria*	*Dermabacteraceae*	*Kytococcus sedentarius* (X87755)
J780	1	1	A1	Yes	Yes	*Actinobacteria*	*Nocardioidaceae*	*Marmoricola auranticus* (Y18629)
J872	14	1	A6	Yes	No	*Actinobacteria*	*Nocardioidaceae*	*Marmoricola auranticus* (Y18629)

CN#	Depth (m)	Method	Media	SW Tested	SW Req.	Class	Family	Nearest Type Strain
J805	2	1	A6	No	No	*Actinobacteria*	*Microbacteriaceae*	*Microbacterium flavescens* (Y17232)
J930	15	1	A2	Yes	No	*Actinobacteria*	*Microbacteriaceae*	*Microbacterium flavescens* (Y17232)
J743	2	1	A7	Yes	No	*Actinobacteria*	*Microbacteriaceae*	*Microbacterium imperiale* (X77442)
J797	2	1	A5	Yes	Yes	*Actinobacteria*	*Microbacteriaceae*	*Microbacterium schleiferi* (Y17237)
J719	200	1	A5	Yes	No	*Actinobacteria*	*Micrococcaceae*	*Micrococcus luteus* (AF542073)
J720	300	1	A5	Yes	No	*Actinobacteria*	*Micrococcaceae*	*Micrococcus luteus* (AF542073)
J736	1	2	A3	Yes	No	*Actinobacteria*	*Micrococcaceae*	*Micrococcus luteus* (AF542073)
J741	11	1	A6	Yes	No	*Actinobacteria*	*Micrococcaceae*	*Micrococcus luteus* (AF542073)
J746	18	1	A5	Yes	No	*Actinobacteria*	*Micrococcaceae*	*Micrococcus luteus* (AF542073)
J801	33	1	A9	Yes	No	*Actinobacteria*	*Micrococcaceae*	*Micrococcus luteus* (AF542073)
J811	14	1	A6	Yes	No	*Actinobacteria*	*Micrococcaceae*	*Micrococcus luteus* (AF542073)
J834	333	1	A9	Yes	No	*Actinobacteria*	*Micrococcaceae*	*Micrococcus luteus* (AF542073)
J835	333	1	A9	Yes	No	*Actinobacteria*	*Micrococcaceae*	*Micrococcus luteus* (AF542073)
J847	200	1	A10	Yes	No	*Actinobacteria*	*Micrococcaceae*	*Micrococcus luteus* (AF542073)
J848	200	1	A10	Yes	No	*Actinobacteria*	*Micrococcaceae*	*Micrococcus luteus* (AF542073)
J916	500	2	A4	Yes	No	*Actinobacteria*	*Micrococcaceae*	*Micrococcus luteus* (AF542073)
J925	500	1	A6	Yes	No	*Actinobacteria*	*Micrococcaceae*	*Micrococcus luteus* (AF542073)
J878	0	1	A6	Yes	No	*Actinobacteria*	*Micromonosporaceae*	*Micromonospora endolithica* (AJ560635)
S326	11	1	A6	No	No	*Actinobacteria*	*Micromonosporaceae*	*Micromonospora endolithica* (AJ560635)
J793	200	1	A5	Yes	No	*Actinobacteria*	*Geodermatophilaceae*	*Modestobacter multiseptatus* (Y18646)
J794	200	1	A5	Yes	No	*Actinobacteria*	*Geodermatophilaceae*	*Modestobacter multiseptatus* (Y18646)
J836	10	1	A5	Yes	No	*Actinobacteria*	*Mycobacteriaceae*	*Mycobacterium brisbanense* (AY012577)
J859	2	1	A5	Yes	Yes	*Actinobacteria*	*Mycobacteriaceae*	*Mycobacterium brisbanense* (AY012577)
J862	10	1	A5	Yes	No	*Actinobacteria*	*Mycobacteriaceae*	*Mycobacterium brisbanense* (AY012577)
J809	2	1	A5	Yes	Yes	*Actinobacteria*	*Mycobacteriaceae*	*Mycobacterium porifera* (AF480589)
J823	5	1	A10	Yes	No	*Actinobacteria*	*Mycobacteriaceae*	*Mycobacterium porifera* (AF480589)
J864	14	1	A5	Yes	No	*Actinobacteria*	*Mycobacteriaceae*	*Mycobacterium porifera* (AF480589)
J879	25	1	A8	Yes	No	*Actinobacteria*	*Mycobacteriaceae*	*Mycobacterium porifera* (AF480589)
J881	2	1	A7	Yes	No	*Actinobacteria*	*Mycobacteriaceae*	*Mycobacterium porifera* (AF480589)
J910	27	1	A5	Yes	No	*Actinobacteria*	*Mycobacteriaceae*	*Mycobacterium porifera* (AF480589)
U171	27	1	A5	Yes	No	*Actinobacteria*	*Mycobacteriaceae*	*Mycobacterium porifera* (AF480589)
R950	5	1	A9	No	No	*Actinobacteria*	*Nocardiaceae*	*Nocardia arthritidis* (AB108781)
S044	5	1	A10	No	No	*Actinobacteria*	*Nocardiaceae*	*Nocardia arthritidis* (AB108781)
J892	200	1	A5	No	No	*Actinobacteria*	*Nocardioidaceae*	*Nocardioides ganghwensis* (AY423718)
R923	1	1	A5	No	No	*Actinobacteria*	*Nocardiopsaceae*	*Nocardiopsis lucentensis* (X97888)
J721	15	1	A5	Yes	No	*Actinobacteria*	*Intrasporangiaceae*	*Ornithinimicrobium humiphilum* (AJ277650)
J824	39	1	A7	Yes	No	*Actinobacteria*	*Intrasporangiaceae*	*Ornithinimicrobium humiphilum* (AJ277650)
J870	8	1	A5	No	No	*Actinobacteria*	*Intrasporangiaceae*	*Ornithinimicrobium humiphilum* (AJ277650)

CN#	Depth (m)	Method	Media	SW Tested	SW Req.	Class	Family	Nearest Type Strain
J734	0	4	A5	Yes	No	*Actinobacteria*	*Promicromonosporaceae*	*Promicromonospora sukumoe* (AJ272024)
J888	2	1	A7	Yes	Yes	*Actinobacteria*	*Pseudonocardiaceae*	*Pseudonocardia antarctica* (AJ576010)
S139	8	1	A5	No	No	*Actinobacteria*	*Pseudonocardiaceae*	*Pseudonocardia yunnanensis* (AJ252822)
S004	19	1	A6	No	No	*Actinobacteria*	*Pseudonocardiaceae*	*Pseudonocardia zijingensis* (AF325725)
R921	13	1	A2	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora arenicola* (AY040619)
S051	10	1	A10	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora arenicola* (AY040619)
S069	20	2	A6	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora arenicola* (AY040619)
S070	20	1	A11	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora arenicola* (AY040619)
S073	14	1	A7	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora arenicola* (AY040619)
S078	20	2	A4	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora arenicola* (AY040619)
S082	15	1	A11	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora arenicola* (AY040619)
S084	5	1	A5	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora arenicola* (AY040619)
S121	36	1	A6	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora arenicola* (AY040619)
S123	33	1	A9	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora arenicola* (AY040619)
S127	28	1	A6	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora arenicola* (AY040619)
S130	25	1	A8	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora arenicola* (AY040619)
S133	20	1	A1	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora arenicola* (AY040619)
S205	20	1	A6	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora arenicola* (AY040619)
S243	400	1	A5	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora arenicola* (AY040619)
S296	40	1	A8	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora arenicola* (AY040619)
S054	20	1	A4	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora tropica* (AY040617)
S055	500	2	A8	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora tropica* (AY040617)
S056	20	2	A8	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora tropica* (AY040617)
S077	8	2	A8	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora tropica* (AY040617)
S103	500	2	A4	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora tropica* (AY040617)
S143	500	2	A6	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora tropica* (AY040617)
S251	8	2	A6	Yes	Yes	*Actinobacteria*	*Micromonosporaceae*	*Salinispora tropica* (AY040617)
J927	15	1	A6	Yes	No	*Actinobacteria*	*Intrasporangiaceae*	*Serinicoccus marinus* (AY382898)
R884	1	1	A3	No	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces arenae* (AJ399485)
R876	0	1	A6	No	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces aureofaciens* (AY289116)
R926	1	3	A3	No	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces aureofaciens* (AY289116)
R881	5	1	A10	No	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces bikiniensis* (X79851)
R918	20	1	A6	No	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces caviscabies* (AF112160)
R924	24	1	A6	No	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces chartreusis* (AJ399468)
R875	21	2	A3	No	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces galilaeus* (AB045878)
R872	15	1	A2	No	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces hebeiensis* (AY277529)
R880	0	4	A5	No	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces koyangensis* (AY079156)
S177	20	2	A6	No	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces lydicus* (Y15507)

CN#	Depth (m)	Method	Media	SW Tested	SW Req.	Class	Family	Nearest Type Strain
J962	15	2	A4	No	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces sampsonii* (D63871)
R887	1	1	A1	No	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces sampsonii* (D63871)
R879	0	4	A5	No	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces tendae* (D63873)
R885	1	1	A1	No	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces tendae* (D63873)
R888	1	1	A1	No	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces tendae* (D63873)
R877	0	1	A6	No	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces thermocarboxydovoran* (U94489)
J799	1	2	A3	Yes	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces thermocoprophilus* (AJ007402)
R925	2	6	A9	No	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces thermocoprophilus* (AJ007402)
R927	1	1	A5	No	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces thermocoprophilus* (AJ007402)
R940	1	3	A3	No	No	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces thermocoprophilus* (AJ007402)
J803	0	1	A6	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus algicola* (AY228462)
J901	400	1	A6	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus algicola* (AY228462)
J903	25	1	A6	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus algicola* (AY228462)
J929	250	1	A6	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus algicola* (AY228462)
J941	0	1	A6	Yes	Yes	*Bacilli*	*Bacillaceae*	*Bacillus algicola* (AY228462)
J943	11	1	A6	Yes	Yes	*Bacilli*	*Bacillaceae*	*Bacillus algicola* (AY228462)
J947	0	1	A6	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus algicola* (AY228462)
J733	23	1	A2	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus aquimaris* (AF483625)
J796	14	1	A6	No	No	*Bacilli*	*Bacillaceae*	*Bacillus aquimaris* (AF483625)
J817	30	1	A2	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus aquimaris* (AF483625)
J842	24	1	A5	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus aquimaris* (AF483625)
J843	14	1	A6	Yes	Yes	*Bacilli*	*Bacillaceae*	*Bacillus aquimaris* (AF483625)
J869	33	2	A4	Yes	Yes	*Bacilli*	*Bacillaceae*	*Bacillus aquimaris* (AF483625)
J804	6	1	A7	No	No	*Bacilli*	*Bacillaceae*	*Bacillus barbaricus* (AJ422145)
J815	24	1	A4	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus barbaricus* (AJ422145)
J904	35	1	A5	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus barbaricus* (AJ422145)
J826	14	1	A6	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus bataviensis* (AJ542508)
J732	23	1	A2	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus cereus* (AE017013)
J845	14	1	A7	No	No	*Bacilli*	*Bacillaceae*	*Bacillus cereus* (AE017013)
J816	30	1	A2	No	No	*Bacilli*	*Bacillaceae*	*Bacillus cohnii* (X76437)
J728	31	1	A2	No	No	*Bacilli*	*Bacillaceae*	*Bacillus decoloationis* (AJ315075)
J828	14	1	A5	Yes	Yes	*Bacilli*	*Bacillaceae*	*Bacillus decoloationis* (AJ315075)
J849	200	1	A10	Yes	Yes	*Bacilli*	*Bacillaceae*	*Bacillus decoloationis* (AJ315075)
J902	0	2	A4	Yes	Yes	*Bacilli*	*Bacillaceae*	*Bacillus decoloationis* (AJ315075)
J937	0	1	A6	Yes	Yes	*Bacilli*	*Bacillaceae*	*Bacillus decoloationis* (AJ315075)
J958	19	1	A6	No	No	*Bacilli*	*Bacillaceae*	*Bacillus endophyticus* (AF295302)
J905	15	1	A5	Yes	Yes	*Bacilli*	*Bacillaceae*	*Bacillus firmus* (AJ717384)
J933	36	1	A7	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus firmus* (AJ717384)

CN#	Depth (m)	Method	Media	SW Tested	SW Req.	Class	Family	Nearest Type Strain
J940	19	1	A6	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus firmus* (AJ717384)
J759	31	1	A2	No	No	*Bacilli*	*Bacillaceae*	*Bacillus horikoshii* (X76443)
J775	15	1	A4	Yes	Yes	*Bacilli*	*Bacillaceae*	*Bacillus humi* (AJ627210)
J808	44	1	A6	Yes	Yes	*Bacilli*	*Bacillaceae*	*Bacillus humi* (AJ627210)
J911	14	1	A6	Yes	Yes	*Bacilli*	*Bacillaceae*	*Bacillus humi* (AJ627210)
J782	1	1	A2	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus indicus* (AJ583158)
J909	15	1	A5	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus indicus* (AJ583158)
J763	21	3	A3	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus megaterium* (X60629)
J764	30	2	A3	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus megaterium* (X60629)
J767	19	5	A10	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus megaterium* (X60629)
J768	19	5	A10	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus megaterium* (X60629)
J769	19	5	A10	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus megaterium* (X60629)
J778	23	1	A2	Yes	Yes	*Bacilli*	*Bacillaceae*	*Bacillus megaterium* (X60629)
J748	10	1	A2	No	No	*Bacilli*	*Bacillaceae*	*Bacillus methanolicus* (AB112727)
J731	21	3	A3	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus pumilus* (AY876289)
J742	22	3	A3	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus pumilus* (AY876289)
J762	21	3	A3	Yes	No	*Bacilli*	*Bacillaceae*	*Bacillus pumilus* (AY876289)
J771	50	1	A2	No	No	*Bacilli*	*Bacillaceae*	*Exiguobacterium aestuarii* (AY594264)
J772	50	1	A2	No	No	*Bacilli*	*Bacillaceae*	*Exiguobacterium aestuarii* (AY594264)
J781	50	1	A2	No	No	*Bacilli*	*Bacillaceae*	*Exiguobacterium aestuarii* (AY594264)
J791	6	1	A6	Yes	No	*Bacilli*	*Bacillaceae*	*Exiguobacterium aestuarii* (AY594264)
J915	0	1	A6	Yes	Yes	*Bacilli*	*Bacillaceae*	*Halobacillus litoralis* (X94558)
J936	0	1	A6	Yes	Yes	*Bacilli*	*Bacillaceae*	*Halobacillus litoralis* (X94558)
J938	0	1	A6	Yes	Yes	*Bacilli*	*Bacillaceae*	*Halobacillus litoralis* (X94558)
J895	20	2	A6	Yes	Yes	*Bacilli*	*Bacillaceae*	*Halobacillus salinus* (AF500003)
J939	19	1	A6	Yes	Yes	*Bacilli*	*Bacillaceae*	*Halobacillus salinus* (AF500003)
J812	14	1	A5	Yes	Yes	*Bacilli*	*Bacillaceae*	*Halobacillus trueperi* (AJ310149)
J931	250	1	A6	Yes	Yes	*Bacilli*	*Bacillaceae*	*Halobacillus trueperi* (AJ310149)
J955	1	3	A3	No	No	*Bacilli*	*Bacillaceae*	*Halobacillus trueperi* (AJ310149)
J912	14	1	A6	Yes	Yes	*Bacilli*	*Bacillaceae*	*Pontibacillus chungwhensis* (AY553296)
J934	250	1	A6	Yes	Yes	*Bacilli*	*Paenibacillaceae*	*Paenibacillus turicensis* (AF378697)
J795	15	2	A3	Yes	Yes	*Bacilli*	*Thermoactinomycetaceae*	*Laceyella sacchari* (AF138737)
R873	23	2	A4	No	No	*Bacilli*	*Thermoactinomycetaceae*	*Laceyella sacchari* (AF138737)
R922	0	1	A8	No	No	*Bacilli*	*Thermoactinomycetaceae*	*Laceyella sacchari* (AF138737)
R949	11	1	A6	No	No	*Bacilli*	*Thermoactinomycetaceae*	*Laceyella sacchari* (AF138737)
S015	15	3	A3	No	No	*Bacilli*	*Thermoactinomycetaceae*	*Laceyella sacchari* (AF138737)
J924	50	1	A5	Yes	Yes	*Bacilli*	*Staphylococcaceae*	*Staphylococcus capitis* (AY688040)

Appendix B: Information regarding the type I PKS pathways used to construct the reference data set. Gene names, accession numbers and the number of modules present are listed.

Organism (Order, genus, species)	GenBank Accession Number	Molecule	Protein Name (no. of modules)	Reference
Actinomycetales				
Streptomyces aculeolatus	ABB88519	Actinofuranone	FurA (3)	Banskota et al. 2006
NRRL 18422	ABB88520		FurB (2)	
	ABB88521		FurC (1)	
	ABB88522		FurD (2)	
	ABB88523		FurE (1)	
Streptomyces sp.	AAQ82561	Candicidin	FcsA (1)	Chen et al. 2003
FR-008	AAQ82564		FcsC (6)	
	AAQ82565		FcsB (3)	
	AAQ82566		FcsF (1)	
	AAQ82567		FcsE (4)	
	AAQ82568		FcsD (6)	
Streptomyces antibioticus	AAZ77673	Chlorothricin	ChlB1 (1)	Jia et al. 2006
	AAZ77693		ChlA1 (3)	
	AAZ77694		ChlA2 (2)	
	AAZ776936		ChlA3 (3)	
	AAZ776937		ChlA4 (1)	
	AAZ776938		ChlA5 (2)	
	AAZ776939		ChlA6 (1)	
Streptomyces neyagawaensis	AAZ94388	Concanamycin	ConA (3)	Haydock et al. 2005
	AAZ94389		ConB (2)	
	AAZ94390		ConC (3)	
Streptomyces aizunensis	AAX98184	ECO-02301	EcoA (4)	McAlpine et al. 2005
	AAX98185		EcoB (2)	
	AAX98186		EcoC (4)	
	AAX98187		EcoD (1)	
	AAX98188		EcoE (3)	
	AAX98189		EcoF (3)	
	AAX98190		EcoG (2)	
	AAX98191		EcoH (4)	
Streptomyces halstedii	BAD38978	Halstoctacosanolides	HalA (3)	Tohyama et al. 2004
	BAD38979		HalB (4)	
	BAD38980		HalC (2)	
Streptomyces aureofaciens	ABB05102	Lipomycin	LipA (1)	Bihlmaier et al.*
	ABB05103		LipB (2)	
	ABB05104		LipC (2)	
	ABB05105		LipD (2)	
Streptomyces violaceusniger	ABJ97437	Meridamycin	MerA (3)	Sun et al. 2006
	ABJ97438		MerB (4)	
	ABJ97439		MerC (7)	
Streptomyces albus	ABG02263	Salinomycin	SalA (1)	Knirschova et al. *
	ABG02264		SalB (2)	
Streptomyces sp.	BAE93722	Tetronomycin	TetA (2)	Demydchuk & Leadlay *
NRRL 11266	BAE93725		TetB (1)	
	BAE93728		TetC (1)	
	BAE93729		TetD (2)	
	BAE93730		TetE (1)	
	BAE93731		TetF (3)	
	BAE93739		TetG (1)	
	BAE93740		TetH (2)	
Streptomyces halstedii	BAD08373	Vicenistatin	VicA (3)	Ogasawara et al. 2004
	BAD08358		VicB (1)	
	BAD08359		VicC (2)	
	BAD08360		VicD (2)	
Mycobacterium avium	YP_881572	Mycoserosic Acid	Myca (1)	Fleischmann et al. *

* KS domain(s) sequenced and annotated but not yet experimentally characterized