**OREGON HEALTH & SCIENCE UNIVERSITY**
**SCHOOL OF MEDICINE-GRADUATE STUDIES**

# ISOLATION AND BIOSYNTHESIS OF BIOACTIVE NATURAL PRODUCTS PRODUCED BY MARINE SYMBIONTS

By

Sherif Ismail Elshahawi

A DISSERTATION

Presented to the Division of Environmental & Biomolecular Systems and
the Oregon Health & Science University

In partial fulfillment for the degree of
Doctor of Philosophy
in
Biochemistry and Molecular Biology

April 2012

DIVISION OF ENVIRONMENTAL AND BIOMOLECULAR SYSTEMS

INSTITUTE OF ENVIRONMENTAL HEALTH

OREGON HEALTH & SCIENCE UNIVERSITY

---

CERTIFICATE OF APPROVAL

---

This is to certify that the Ph.D. dissertation of

Sherif Ismail Elshahawi

has been approved

_____
Dr. Peter Zuber, Professor
Examination Committee Chair

_____
Dr. Margo G. Haygood, Professor
Research Advisor

_____
Dr. James W. Whittaker, Associate Professor

_____
Dr. Eric Schmidt, Associate Professor
University of Utah
External Examiner

## Abstract

The marine environment is a rich source for chemical and biological diversity. Many marine animals are closely associated with symbiotic bacteria. Marine animals have been shown to contain natural products with complex structure and interesting pharmacological activities. These compounds are usually produced within the animal to perform a certain function relevant to the holobiont. These natural products are often synthesized by the bacteria using dedicated biosynthetic enzymes. Thus, genome analysis of microbes can predict the potential for the production of natural products.

I analyzed the genomes of four symbionts of the shipworms *Bankia gouldi* and *B. setacea,* in addition to the metagenome of *B. setacea*. They all showed high potential for the production of active secondary metabolites, especially polyketides and nonribosomal peptides. One of these bacterial symbionts, *Teredinibacter turnerae* T7901, was subjected to an antibacterial-guided fractionation to yield two compounds. Analysis of these compounds using 1 and 2D nuclear magnetic resonance spectrometry, in addition to mass spectrometry, led to their structure elucidation. These two compounds belonged to the macrodiolide polyketides, tartrolons. The biosynthetic gene cluster of this group was identified using gene disruption mutation and analysis and a biosynthetic scheme was proposed. Tartrolons are prevalent in other *T. turnerae* strains and are found to be expressed in the shipworm host suggesting an important function. I propose that tartrolons might be playing a role in the regulation of the boron levels in the animal, in addition to the inhibition of specific bacteria within the shipworm. Tartrolon E was found to be active against pathogenic bacteria. Another system that involves a potential symbiotic relationship is associated with the ascidian *Ecteinascidia* and is based on the

anticancer compound Et743. A survey on different *Ecteinascidia* animals from different geographical locations was performed, using high resolution mass spectrometry and polymerase chain reaction to establish the presence of Et743, in addition to potential gene fragments that could be involved in the biosynthesis of the compound in these populations.

The study of natural products produced by symbiotic bacteria within marine invertebrates could provide a rich source for biologically active compounds. Moreover, the study of their biosynthetic gene clusters could lead to the heterologous expression of these pathways to produce higher amounts or to the formation of combinatorial biosynthetic analogs of higher activity and/or less toxicity.

## Acknowledgements

Many people have impacted my research and personality during the time I have spent at OHSU and helped my development as a scientist. I would like to start with my PhD advisor and mentor Margo Haygood who gave me the chance to join her lab. Her way of mentoring students challenges them to work as independent researchers and prepares them for the next level. I benefited from her valuable suggestions during our one-on-one weekly meetings. I also hope my future lab would be as organized as hers. She encouraged me to go to different conferences that increased my experience. Margo also taught me to be less skeptical about results and encouraged me to enjoy life beside work. She pushed me to snorkel for the first time in my life during our field trip to Palau then to our following field trip to Egypt, which turned out to be one of the most magnificent experiences in my life. Margo understands the significance of collaboration with other scientists and gave me the chance to work with another smart scientist, Eric Schmidt. Eric who was also my co-advisor is a great resource for information when it comes to natural products. No matter whether it is related to chemistry or biology, he will always have a suggestion and an idea. It was great to meet with him and his group remotely every other week during the first two years of my research and then every week during our ICBG elluminate meetings to discuss progress and problems in our research. Thank you both for the great knowledge I have gained.

My other committee members were also helpful. I have learned a lot from Peter Zuber and James Whittaker from participation in their classes, in addition to helpful discussions. I have learned many molecular biology techniques from Peter and a lot about biochemistry from James. Their comments on this dissertation were invaluable.

I appreciate the Haygood-Tebo-Simon weekly meetings where I had the chance to present my work from time to time. These meetings included people from many diverse areas and thus, have increased my knowledge about other fields of research. Everyone and especially Bradley Tebo and Holly Simon provided me with many helpful comments on my work. These weekly meetings also taught me how to explain my research to people outside the field in an easy way. I hope I have succeeded in that. The round tables every three weeks where everyone shares his research challenges during the previous period was very helpful. It taught me that everyone has a cycle of success and failure and that I am not alone suffering in research.

I am grateful to many past and present members of the Haygood lab. Carolyn Sheehan was responsible for managing the Haygood-Tebo group and tried to accommodate everyone's needs. She showed me how to take care of the shipworms and the *Ecteinascidia* samples in the aquaria and to keep them happy. Carolyn was specialized in organizing great graduation parties and I appreciate the awesome work she did in mine. Amaro Trindade-Silva, who was a visiting student in our lab, helped me in my primary attempts in growing *Teredinibacter turnerae*. Amro Hanora was a visiting scientist from my home country, Egypt. He was a hard working person full of enthusiasm and loves to learn as much techniques as possible. Amro was always eager to teach others anything he knows. Mohamed ElBaz, another person from my home country who was a member in our lab and was also a hard working person. Baz was always optimistic and a fun person to work with. I also enjoyed playing table tennis with him. It was good to work with Andrew Han in the lab during the years. Andrew was the best person to perform fluorescence in situ hybridization in our lab. Brian Fishman was a research-focused

person. It was fun watching Meghan Betcher extracting shipworms from wood and dissecting them. Giang Ong was a good person to chat with. I am also grateful to many members in the Tebo group. Rick Davis answered my silly questions about phylogeny during my first year. Other Tebo members include Wendy Smythe, Cristina Butterfield, Roberto Anitori, Sung-Woo Lee, Nicholas Caputo, John Buzzo, Kati Geszvain, Craig Anderson, and Suzanna Brauer all helped my stay at OHSU to be enjoyable.

I am grateful to the Philippine Mollusk Symbiont, International Cooperative Biodiversity Group (PMS-ICBG) project (grant number 1U01 TW008163) which gave me the opportunity to interact with many people from similar and different research areas. This interaction, through our weekly virtual meetings gave me the chance to present my work several times and hear helpful comments. In addition, the helpful discussions from everyone and especially Daniel Distel and Zhenjian Lin besides Margo and Eric, enriched my experience. The ICBG also gave me the chance to share my chemistry experience with Gwen Limbaco and Imelda Forteza during their training in our lab. I hope I didn't make them hate chemistry.

I am particularly grateful to the analytical facility at Portland State University. I am thankful to Cheryl Hudson and David Peyton for training me on the NMR instrument. David is brilliant when it comes to NMR techniques and parameters. I also appreciate Andrea DeBarber who trained me to use the MS and was extremely patient to answer my calls at odd times to solve machine problems.

I am also grateful to our previous collaborator in the Suez Canal University; Ali Gab-Alla. Ali is an amazing ecologist who helped us during our collection of *Ecteinascidia*

samples from the Red Sea. I am thankful to Nancy Christie for solving all my non-research problems during the years. She is the person to go to for any administrative-related issues. I also want to thank James Mohans for solving my IT problems during the years.

I cannot thank enough my father and mother, in addition to my sister for their endless support and encouragement during my graduate studies even if I had to be thousands of miles away to achieve my dream.

Last but not least, I want to express my gratitude to my wife Shaimaa and my two children; Janna and Nour-Eldin. Shaimaa supported me and listened to my endless complaints about experiments not working, without boredom. She also provided great suggestions and support throughout the years.

## Dedication

This work is dedicated to my wonderful parents. Without the tireless support, encouragement, and prayers of my great father, Ismail Elshahawi and mother throughout the years, I would have given up long time ago.

## Statement of Contribution

Chapter 2. Sequencing of all shipworm bacteria strains and metagenome was done by the Department of Energy, Joint Genome Institute under a community sequencing program award to Dr. Daniel Distel (Ocean Genome Legacy).

Chapter 3. Dr. Amro Hanora (Suez Canal University, Egypt) tested the prevalence of region 2 in different *T. turnerae* strains using polymerase chain reaction.

Chapter 3. Dr. Andrew Han (Oregon Health & Science University) constructed the gene insertion mutant AH02 in region 2 of *T. turnerae* T7901.

Chapter 3. Ms. Rowena Antemano (University of the Philippines) tested the activity of tartrolon E in the dorsal root ganglion assay.

Chapter 3. Ms. Malem Flores (University of the Philippines) tested the activity of tartrolon E against the pathogenic bacteria *P. aeruginosa*, and methicillin sensitive and methicillin resistant *S. aureus*. In addition, she tested the cytotoxicity of the compound against the MCF-7 breast cancer cell line.

Chapter 4. Dr Eric Schmidt (The University of Utah) designed the degenerate primers for the reductase domain.

## List of Abbreviations

| | |
|---|---|
| A | adenylation |
| ABC | ATP-binding cassette |
| ACP | acyl carrier protein |
| AIDS | acquired immune deficiency syndrome |
| antiSMASH | antibiotics and secondary metabolite analysis shell |
| AT | acyltransferase |
| BLAST | Basic Local Alignment Search Tool |
| bp | base pair |
| C | condensation |
| $CDCl_3$ | deuterated chloroform |
| $CHCl_3$ | chloroform |
| COSY | correlation spectroscopy |
| CTAB | cetyltrimethylammonium bromide |
| DH | dehydratase |
| E | epimerase |
| EDTA | ethylenediamine tetraacetic acid |

| | |
|---|---|
| ER | enoylreductase |
| FT-IR | Fourier transform infrared spectroscopy |
| HMBC | heteronuclear multiple bond correlation |
| HPLC | high performance liquid chromatography |
| HR-MS | high resolution mass spectrometry |
| HSQC | heteronuclear single-quantum correlation spectroscopy |
| IAA | isoamyl alcohol |
| $IC_{50}$ | inhibition concentration of 50% |
| IR | infrared spectroscopy |
| kbp | kilobase pair |
| KR | ketoreductase |
| KS | ketosynthase |
| LC/MS | high performance liquid chromatography linked to mass spectrometry |
| Mbp | megabase pair |
| MIC | minimum inhibitory concentration |
| MRSA | methicillin resistant *Staphylococcus aureus* |
| MS | mass spectrometry |

| | |
|---|---|
| MT | methyltransferase |
| NBD | nucleotide binding domain |
| NMR | nuclear magnetic resonance |
| NOESY | nuclear Overhauser enhancement spectroscopy |
| NRP | nonribosomal peptide |
| NRPS | nonribosomal peptide synthetase |
| ORF | open reading frame |
| PCP | peptidyl carrier protein |
| PCR | polymerase chain reaction |
| PDA | photodiode array detector |
| PK | polyketide |
| PKS | polyketide synthase |
| PP | phosphopantetheine |
| RP | ribosomal peptide |
| T | thiolation |
| TE | thioesterase |
| $T_{10}E_1$ | tris-EDTA in the ratio of 10 mM to 1 mM |

| | |
|---|---|
| TMS | tetramethylsilane |
| UV | ultraviolet spectrum |

Contents

## List of Figures

## List of Tables

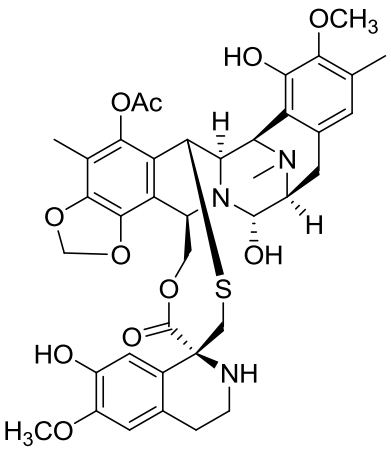
# 1. Introduction

## 1.1. Secondary metabolites and therapeutic activities

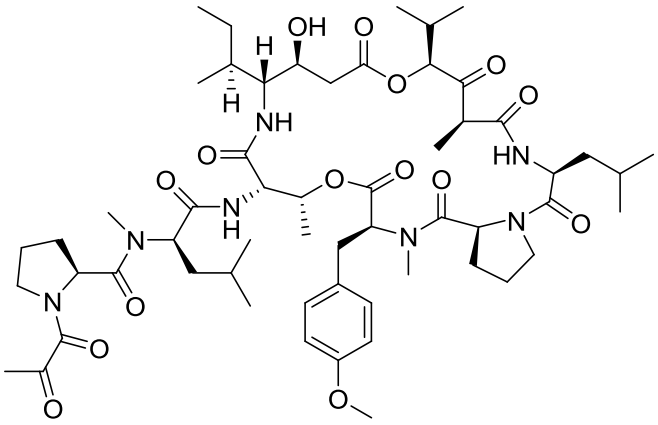
Secondary metabolites are small organic molecules produced by plants, fungi, insects, or microorganisms (Davies and Ryan, 2011). They are not utilized for primary metabolism and considered not important for the vegetative growth of the producing organisms making them different from primary metabolites. Nevertheless, secondary metabolites play different essential roles in the ecology of the producing and the surrounding organisms. They usually target specific cell components and may function as chemical defenses, pheromones, interspecies communication signals, siderophores, transcription regulators, in addition to other roles. Some of the interest in secondary metabolites is due to their reported pharmacological activities. Many secondary metabolites have been reported to act as anti-inflammatory, antibacterial, antitumor, antiviral, or antifungal agents.

Most of the biodiversity of the planet resides in the oceans among animals; much of it lies in invertebrates. Marine invertebrates have proved to be rich sources of bioactive compounds (Figure 1-1). For example, the Et743; trabectedin **1**, marketed under the name Yondelis® is approved for the treatment of soft tissue sarcoma and ovarian carcinoma. Et743 was originally isolated from the marine sea squirt *Ecteinascidia turbinata* (Rinehart et al., 1990). The cyclodepsipeptide plitidepsin **2** or Aplidin®, reported from the tunicate *Aplidium albicans* was also found to have an anticancer effect (Rinehart and Lithgow-Bertelloni, 1998) and is currently in phase III clinical trials for the treatment of relapsed and refractory multiple myeloma. The ω-conotoxin ziconotide was originally

isolated from the cone snail *Conus magus* (Olivera, 2000). It was found to possess a strong analgesic activity. It is marketed under the name Prialt® and prescribed for patients with severe chronic pain in those suffering AIDS or cancer.




ET743 (trabectedin); Yondelis® (**1**)



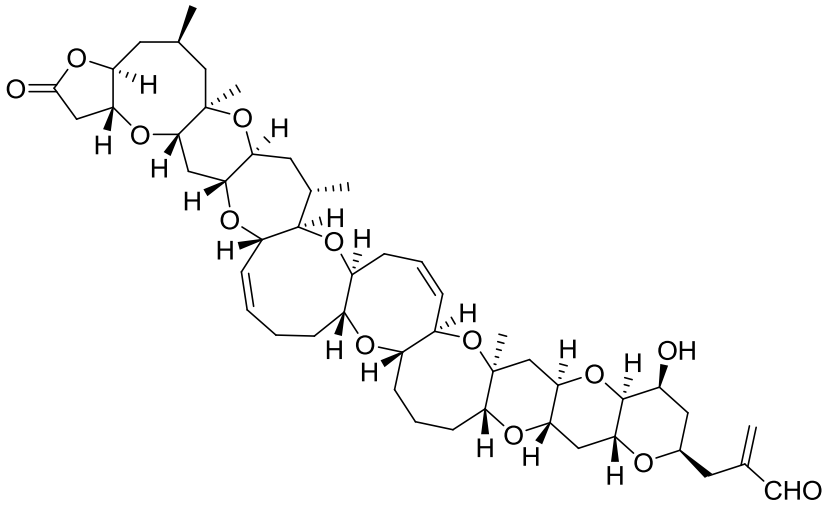

plitidepsin (dehydrodidemnin B); Aplidin® (**2**)

**Figure 1-1. Examples of marine compounds of significant pharmacological activities.**

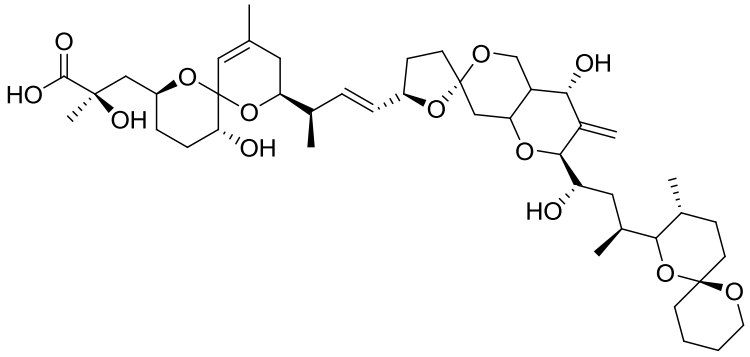
## 1.2. Ecological functions of marine secondary metabolites

Small molecules in marine organisms often play an important role in the chemical defense of many sessile animals. Tissues of soft sessile marine animals have been shown to be a rich source of unusual secondary metabolites and thus most of the marine natural products known are assumed to play a defensive role (Pawlik, 1993). Most of these compounds are thought to act as feeding deterrents against other predators or as antifouling agents although other compounds act as chemical signals for biological processes such as larval settlement or reproduction while others inhibit the overgrowth of neighboring organisms (Figure 1-2). For example, the cyclic polyether brevetoxins **3** found in the dinoflagellate *Karenia brevis* are extremely toxic to other marine organisms (Baden et al., 2005). These compounds were found to possess a neurotoxic effect and cause neurotoxic shellfish poisoning by binding voltage gated sodium channels. Okadaic acid **4** was identified in sponges (Tachibana et al., 1981) and dinoflagellates (An et al., 2010). It serves as a defense molecule in sponges against parasitic annelids (Schröder et al., 2006) and possesses negative allelopathic activity against other marine algae in dinoflagellates (Windust, 1996). The indole alkaloid lyngbyatoxin A **5** isolated from the cyanobacteria *Lyngbya majuscula* was shown to be toxic to marine organisms (Cardellinae, 1979). Caulerpenyne **6** is a sesquiterpene isolated from different species of the green alga *Caulerpa* (Guerriero et al., 1992) and have shown to possess allelopathic and feeding deterrence activities (Smyrniotopoulos et al., 2002). The two terpenoids; pukalide **7** and 11β-acetoxypukalide **8** have been shown to act as signals for larval settlement and gametogenesis (Slattery et al., 1999). Mycosporine amino acids derivatives **9** are present in shallow marine organisms such as sea anemones and

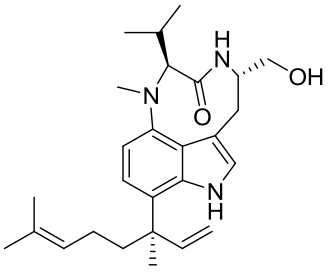
cyanobacteria that are exposed to solar irradiation, serve as protectants against the harmful effects of this irradiation (Schmidt, 2011). Many of these molecules are occasionally proposed to be produced by a symbiotic microorganism.



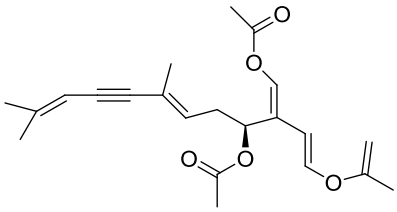
brevetoxin A –Pb Tx-1 (**3**)



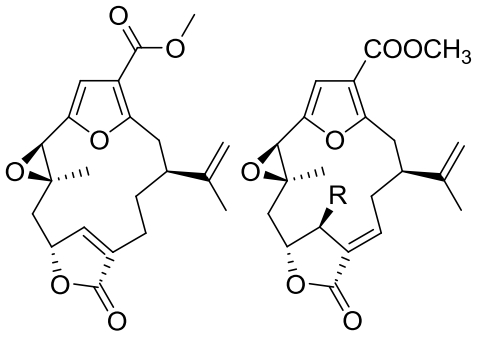
okadaic acid (**4**)



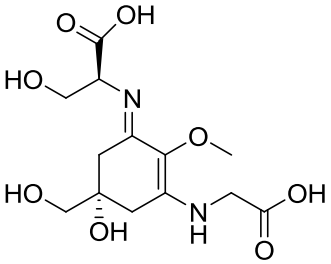

lyngbyatoxin A (**5**)




caulerpenyne (**6**)




pukalide (**7**, left) and 11β-acetoxypukalide (**8**, right) R=$OCOCH_3$




shinorine (**9**), an example of mycosporine amino acids (MAA)

**Figure 1-2. Examples of compounds of ecological relevance to marine organisms.**

## 1.3. Symbiosis

Most, if not all, eukaryotes live in close association and continuous interaction with microorganisms. Some of these microorganisms contribute slightly to the total holobiont. They retain all major essential genes allowing them to live independently. However, other microorganisms, usually called symbionts, rely partially or completely on the host. Those that have become obligate symbionts might have lost essential metabolic genes and thus cannot be cultured except under specific conditions (Schmidt, 2008). Some of these symbionts have been shown to synthesize essential secondary metabolites for the benefit of the host. These molecules are important to the symbiosis and their biosynthetic genes are usually retained. This kind of symbiosis that is based on bioactive compounds is called bioactive metabolite symbiosis (Hildebrand et al., 2004). The symbiont produces beneficial compounds for the host and in return acquires a safe habitat in the host cells feeding on the host's metabolites or wastes. Symbiosis exists in both terrestrial and marine organisms.

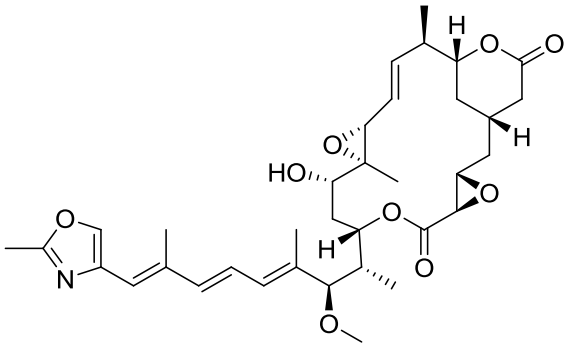
### 1.3.1. Symbiosis in terrestrial systems

There are several examples of bioactive metabolite symbiosis (Figure 1-3). One is the symbiotic relationship between the plant pathogen *Rhizopus microsporus* fungus and its intracellular proteobacterial symbiont *Burkholderia rhizoxinicia*. The fungus attacks crops and feed on the dying plant (Partida-Martinez et al., 2007). The pathogenic factor in this attack is rhizoxin **10** and is produced by the symbiont and its biosynthesis is specified by a hybrid gene cluster encoding polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) enzymes. Bioactive metabolite symbiosis also exists in two symbiotic relationships; in the leaf cutting ants *Acromymex* and the Southern pine beetles

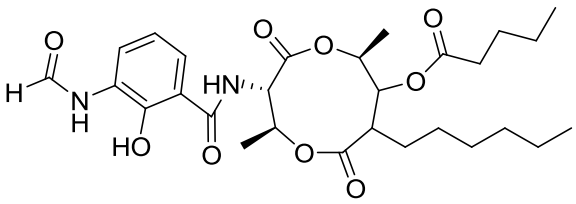
*Dendroctonus frontalis* which both engage in symbiosis with actinobacteria to protect their food fungus. The leaf cutting ants grow the fungus *Leucoagaricus gongylophorous* by supplying it with organic matter and use it as their main source for nutrition. In order to protect this fungus from other pathogenic fungi, the ants harbor actinobacteria: *Pseudonocardia* and *Streptomyces* species. These symbionts produce antifungal compounds such as antimycins **11** that are active against the pathogenic but not the food fungus (Schoenian et al., 2011). The southern pine beetles also harbor different actinobacterial symbionts to protect their fungus *Entomocorticium* sp. from pathogenic ones. One of these symbionts produces a polyene peroxide; mycangimycin **12** that was found to act as a selective antifungal agent against pathogenic fungi (Scott et al., 2008). Another example of an actinobacteria-insect symbiosis is the association of a group of *Streptomyces* symbionts with the beewolf digger wasps, *Philanthus* spp. (Kroiss et al., 2010). The symbionts transferred from the antennal gland of the female beewolf to the larval cocoon, produce piericidin $A_1$ **13**, in addition to eight other antibiotics that protect the larva from pathogenic bacteria and fungi.

Species of the gammaproteobacterial genera *Xenorhabdus* and *Photorhabdus* are also involved in symbiotic relationships. These bacteria are known to produce a wide array of small molecules as evidenced from the isolated compounds or the secondary metabolites genes encoded in their genomes (Bode, 2009). The species of *Xenorhabdus* and *Photorhabdus* engage in a symbiotic relationship with nematodes of the genera *Steinernema* and *Heterorhabditis*, respectively. The symbionts produce stilbene **14** that attacks the immune system of insects by inhibiting their phenoloxidase activity.

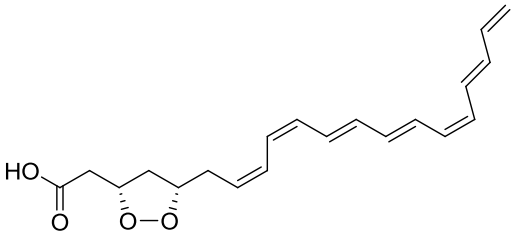
Moreover, stilbene wards off other microbial organisms trying to compete for nutrients in the decaying insect.



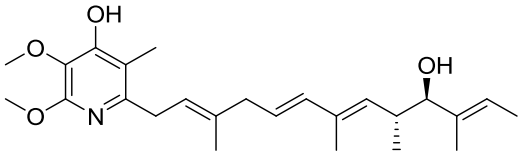
rhizoxin (**10**)



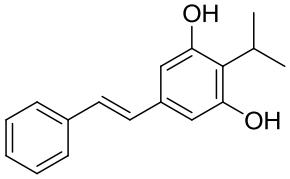
antimycin $A_1$ (**11**)



mycangimycin (**12**)



piericidin $A_1$ (**13**)



stilbene (**14**)

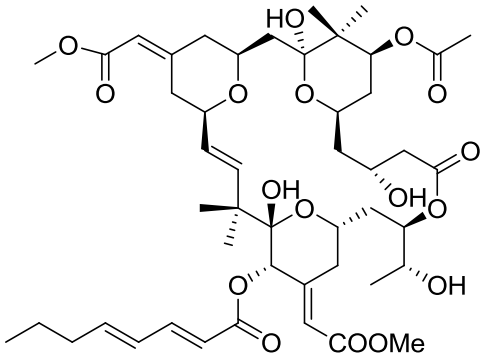
**Figure 1-3. Structures of compounds engaged in symbiosis in the terrestrial system.**

### 1.3.2. Symbiosis in marine systems

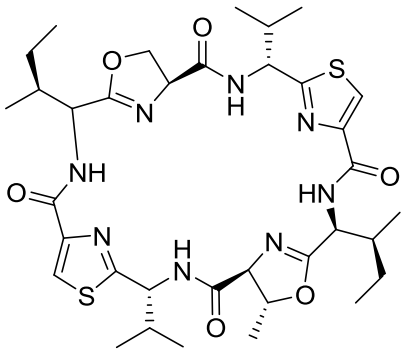
Symbiosis in terrestrial organisms is relatively better understood than that in marine organisms. Nevertheless, a few examples have been reported lately (Figure 1-4). One of the best known examples is the symbiotic relationship between the bryozoan *Bugula neritinia* and its gammaproteobacterium symbiont *Candidatus* Endobugula sertula. Bryostatins **15** once thought to be produced by the bryozoans are actually produced by the gammaproteobacterium symbiont (Davidson and Haygood, 1999; Davidson et al., 2001). Bryostatins are macrolides that are biosynthesized through a polyketide synthase mechanism (Sudek et al., 2007). It was shown to act as feeding deterrent to protect the highly vulnerable larvae from predation by other organisms (Lopanik et al., 2004a; Lopanik et al., 2004b).

*Lissoclinum* ascidians are known to contain a group of cyclic ribosomal peptides (cyanobactins) such as patellamides exemplified by patellamide A **16**, and other structurally related derivatives. These compounds are actually produced by an as yet uncultivated cyanobacterium; *Prochloron* spp. as determined by two different methods; metagenome analysis and heterologous expression (Schmidt et al., 2005; Long et al., 2005).

The ascidians *Ecteinascidia turbinata* and *E. thurstoni* are known to contain a group of tetrahydroisoquinoline compounds called ecteinascidins **1**. They are thought to protect these sessile animals from predation by other marine organisms. Several reports have shown that these compounds are actually produced by a gammaproteobacterium symbiont *Candidatus* Endoecteinascidia frumentensis that contains the gene cluster for these metabolites (Perez-Matos et al., 2007; Elbaz, 2009; Rath et al., 2011).




bryostatin 1 (**15**)




patellamide A (**16**)

**Figure 1-4. Examples of compounds involved in symbiosis in the marine system.**

## 1.4. Supply problem

Some of the active compounds isolated from marine invertebrates have made it to the market while others are in clinical trials. However, a major problem of these compounds is the supply problem. Most of these compounds are produced in minute amounts; tons of animals may be required to obtain milligram quantities. Thus, extracting the compounds directly from the animal provides insufficient quantities for clinical trials and market requirements and is detrimental for both the animal and the environment. Chemical synthesis is hindered by the numbers of stereocenters, complexity of the molecule, and low yield. In addition, aquaculture requires huge investments and does not produce the expected results possibly due to the complex conditions required. Many of these

compounds originally extracted from the invertebrate have shown to be actually produced by a microorganism while others share structural similarity with different microbial metabolites raising the possibility that a microbial symbiont within the animal is the true producer of these bioactive metabolites. Many of these symbionts are extremely difficult to cultivate. However, identifying the gene cluster responsible for the biosynthesis of the compound and expressing this cluster in a heterologous, easily manipulated host can lead to overproduction of the desired compound in higher amounts, as shown in the case of patellamides (Schmidt et al., 2005). In addition, changes in the gene cluster either by adding or deleting specific genes can potentially lead to the production of combinatorial biosynthetic libraries (Fisch et al., 2011) that contain analogs with higher activities or less toxicity.

## 1.5. Biosynthesis of secondary metabolites

Many secondary metabolites are produced by specific multifunctional proteins that assemble specific substrates into the required metabolite (Figure 1-5). These proteins are synthesized from transcripts produced from the DNA sequence information in the genome. Analysis of the possibility of secondary metabolite biosynthesis could either be based on a proteomic study where specific enzymes for secondary metabolite biosynthesis are analyzed using mass spectrometry (Bumpus et al., 2009) or a study of the total transcriptome which could also provide evidence about such compounds (Hoover et al., 2007). However, proteomic and transcriptomic analysis provide only information about enzymes capable of producing compounds that are actually expressed under a specific growth condition. Microbial genome analysis has the advantage that it can detect all potential enzymes that have homologs in the database and capable of

producing compounds. However, enzymes involved in the synthesis of novel compounds using a yet unknown mechanism may be missed. Targeting a compound based on a predicted structure from analysis of the potential gene cluster could be approached by different strategies including different growth conditions, mutation analysis, expression in a heterologous organism, feeding studies, or induction or deletion of transcription regulators. The benefit of studying microbial versus eukaryotic genomes besides the lack of introns is that genes involved in a specific function are usually clustered together in microorganisms.

Secondary metabolites can be classified according to the kind of genes that biosynthesize them. There are two major types that have attracted attention because they form complex bioactive compounds and are produced by large enzyme complexes. These are polyketides (PKs) and nonribosomal peptides (NRPs) synthesized by polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs), respectively. PKSs and NRPSs are megaenzymes that are multimodular and are encoded by large open reading frames (ORFs) in gene clusters that can reach more than 150 kbp (Laureti et al., 2011). Type I PKSs catalyze the decarboxylative condensation reaction of multiple carboxylic acid units leading to a chain of multiple ketide units. The diversity of the functional groups arises from the size of the ORFs and the presence or absence of accessory domains within each module that can convert the ketide group to an alcohol, an olefin, or a completely saturated bond. This is different from fatty acid synthases (FAS) where all ketide units are completely saturated forming a long saturated fatty acid chain. NRPs on the other hand are made by the amide bond formation between different proteinogenic and nonproteinogenic amino acids. This is different from ribosomal peptides (RP) such as

bacteriocins, which are derived from short precursor peptides and made from proteinogenic amino acids. Thorough analysis of the domains in each module in the NRPS of the gene cluster could lead to prediction of the exact compound or at least the class of compound that is produced by that specific cluster.

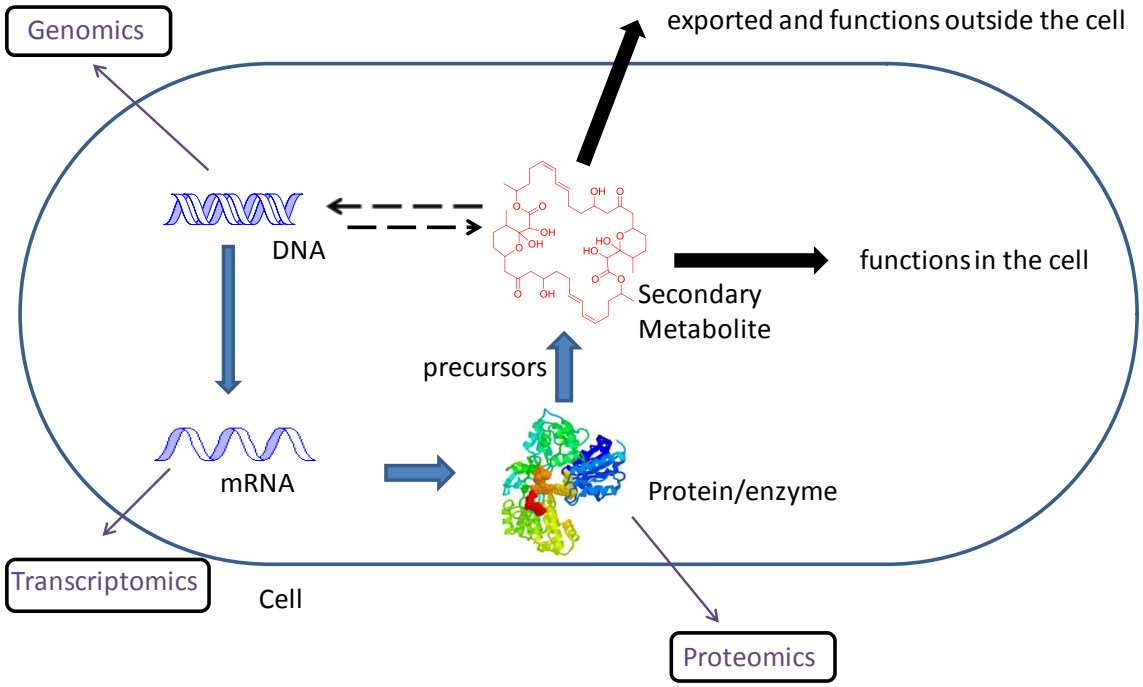



**Figure 1-5. General steps for the biosynthesis of secondary metabolites in bacteria and methods of predicting production.**

## 1.6. Shipworms

Shipworms are marine wood boring bivalve mollusks of the family Teredinidae (Turner, 1966; Distel, 2003). They utilize wood as a source for shelter and nutrition (Figure 1-6). Members of the family Teredinidae differ from other mollusks in that they have an elongated body and reduced shell that have adapted to the wood boring process. The shell, with its teeth-like projections, is used as a drill creating a burrow where the shipworm resides. The shipworm uses siphons, located on the posterior side of the body,

for respiration and filter feeding. When the siphons are withdrawn, a pair of specialized pallets is used to plug the burrow, protecting the shipworm against adverse environmental conditions. Shipworms spawn under conditions of suitable temperature and salinity based on the species, leading to the release of free swimming larvae. These larvae attach themselves to wood structures and develop into adult shipworms.

Shipworms harbor a group of closely related gammaproteobacteria in their gills. It is thought that these bacterial symbionts play a major role in the cellulose degradation of wood particles through the production of cellulases (Ekborg et al., 2007). In addition, it was shown that these bacteria undergo nitrogen fixation and probably help to compensate for the poor nitrogen diet of the host (Lechene et al., 2007). It is possible that these symbionts contribute to the symbiosis by providing secondary metabolites that could fulfill essential functions for the host or the microbial community.

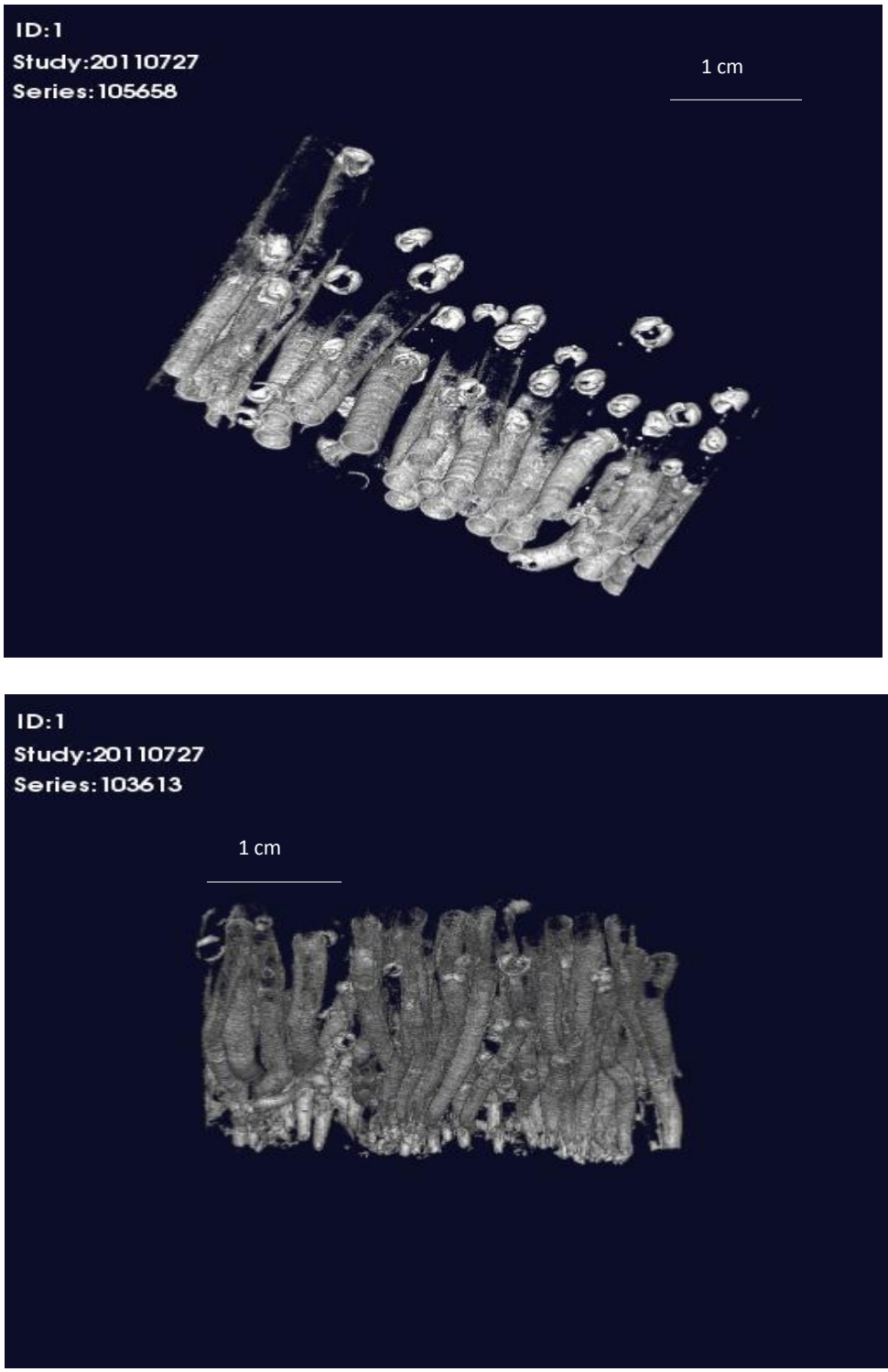



**Figure 1-6. A computed tomography (CT) scan of a wood piece infested with shipworm species.** A picture of upper (top) and lower (bottom) body part of the shipworms is shown. Spherical objects are the wood-drilling shells while the tubular objects are the calcareous lining of the burrow.

## 1.7. Summary

Marine invertebrates and associated symbionts represent an attractive source for the study of secondary metabolites due to their essential ecological function, in addition to their pharmacological bioactivities. The ability of microorganisms to produce secondary metabolites could be predicted by the bioinformatic analysis of microbial genomes. In chapter 2 of this dissertation, I present analysis of the genomes of four shipworm isolates in addition to the metagenome of the microbial community of one shipworm. Evidence for the ability of these strains to produce secondary metabolites is presented. I also show possible intermediates of biosynthesis that are produced. In chapter 3, I describe the isolation and identification of bioactive macrolides from one of the shipworm strains, *Teredinibacter turnerae* T7901, and identify the gene cluster responsible for their biosynthesis. Analysis of the gene cluster is presented and a biosynthetic scheme is predicted. I also show bioactivities of the main compound. In chapter 4, I show a potential gene fragment from the Et743 biosynthetic gene cluster. In addition, I screen different collections of ascidians for the presence of the compound using mass spectrometry and polymerase chain reaction.

# 2. Bioinformatic Analysis of the Secondary Metabolite Genes in Shipworm Symbionts

## 2.1. Introduction

Organisms produce different classes of natural products such as polyketides, peptides, and terpenoids to perform certain functions including but not limited to chemical defense, quorum signalling, or interspecies communication. Polyketides (PKs) and nonribosomal peptides (NRPs) have attracted wide attention because of their bioactivities including antitumor, antibacterial, antifungal, anti-inflammatory, and antiviral activities.

Polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs) are megaenzymes organized into multifunctional modules that produce PKs and NRPs, respectively. The structure of active-site modules in these enzymes can be used in some cases to predict the metabolite produced. Both PKSs and NRPSs contain acyl carrier protein (ACP) and peptidyl carrier protein (PCP; or T for thiolation) domains, respectively that carry a 4'-phosphopantetheine (4'-PP) group. The chemistry and flexibility of the 4'-PP group acts as a tether and allows the growing chain and the monomer to covalently bind to the enzyme as thioesters. Both also carry a thioesterase (TE) domain that releases the growing chain through hydrolysis or cyclization (Walsh and Fischbach, 2010).

PKs are formed through the oligomerization of small carboxylic acids such as malonyl or methylmalonyl CoA that introduces two and three carbon units, respectively. A loading module loads the substrate and the chain grows through the action of multiple modules (Figure 2-1). A basic PKS module consists of an acyltransferase (AT) domain which selects the monomer building unit, an ACP domain, and a β-ketosynthase (KS) domain

which condenses the monomer unit on the ACP of the same module with the growing chain on the ACP of the preceding module. In iterative KSs, ATs are only present in the loading modules and are often located outside the PKS entirely. Each loaded monomer could be derivatized by accessory domains such as a β-ketoreductase (KR) domain which reduces a carbonyl (-CH-C=O) to an alcoholic group (-CH-CH-OH), a β-hydroxyacyl dehydratase (DH) domain which dehydrates the alcoholic group to form an olefin group (-CH=CH-), an α,β-enoyl reductase (ER) domain that saturates a double bond ($-CH_2-CH_2-$) (Figure 2-1), and a methyltransferase (MT) domain, which catalyzes methylation. These accessory domains are the main reason for PK diversity (Walsh and Fischbach, 2010). Examples of PKs include the antibacterial erythromycin, the antifungal amphotericin B, and the immunosuppressant FK-506.

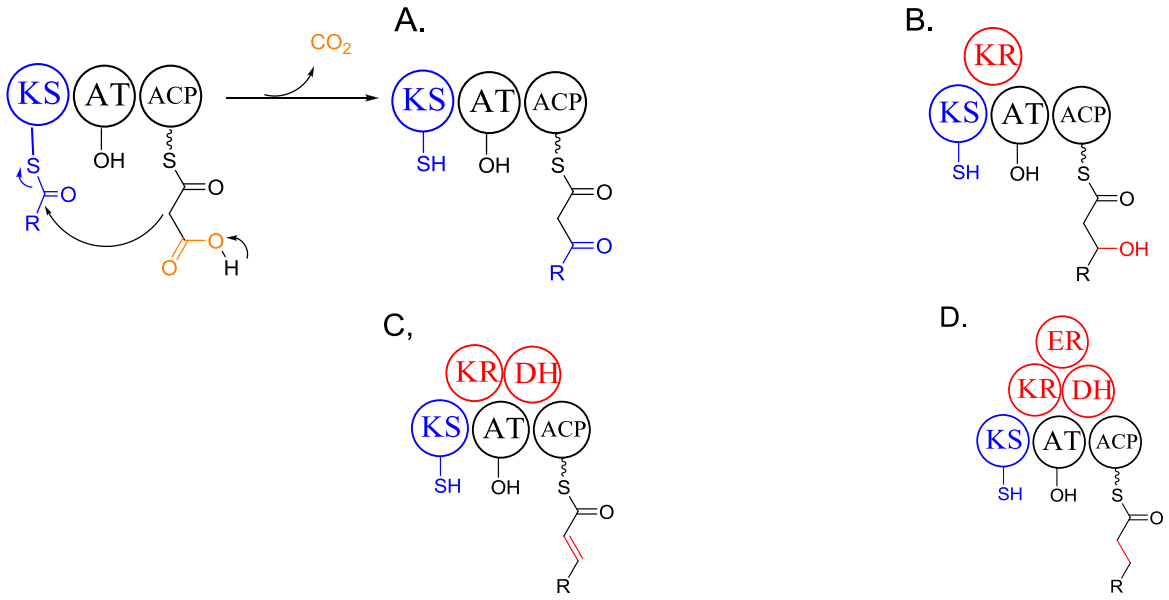



**Figure 2-1. Mechanism of chain elongation in polyketide synthases.**
A) The basic PKS is formed of KS, AT (only present in the loading module in trans AT where it is often located outside the PKS entirely), and ACP domains and lead to the formation of a carbonyl derivative. The presence of B) KR causes reduction to alcohol, C) KR and DH forms olefin, D) KR, DH, and ER forms a saturated methylene. Other accessory domains include MT (not shown). Blue color indicates fragments incorporated by KS domains. Red color highlights the change in organization caused by the accessory domains indicated.

NRPs, on the other hand, are formed by the oligomerization of amino acids which could be proteinogenic or nonproteinogenic. A basic NRPS module is formed from three domains. In addition to the PCP domain mentioned earlier, an adenylation (A) domain selects a specific amino acid and activates it as an amino acid adenylate through an ATP-dependent reaction and transfers it to the 4'-PP group of the PCP domain. The substrate selectivity of the A domain can in principle be predicted by analyzing 8-10 critical residues in the A domain (Stachelhaus et al., 1999; Challis et al., 2000). A condensation (C) domain located at the beginning of the module catalyzes the amide bond formation between intermediates tethered on the PCP present in the same module and that bound to the preceding one (Figure 2-2). Accessory domains may include an epimerization (E) domain causing the conversion of L-amino acid to the D form and a methylation domain (MT). In contrast to the diversity of PKs, the diversity of NRPs comes from the wide selection of proteinogenic and non-proteinogenic amino acids that can be loaded by the A domain (Walsh and Fischbach, 2010). In fact, 526 different amino acids have been identified in NRPs (Norine database: http://bioinfo.lifl.fr/norine/; accessed 01/2012). Examples of NRPs include the antibacterial daptomycin and the immunosuppressant cyclosporin A.

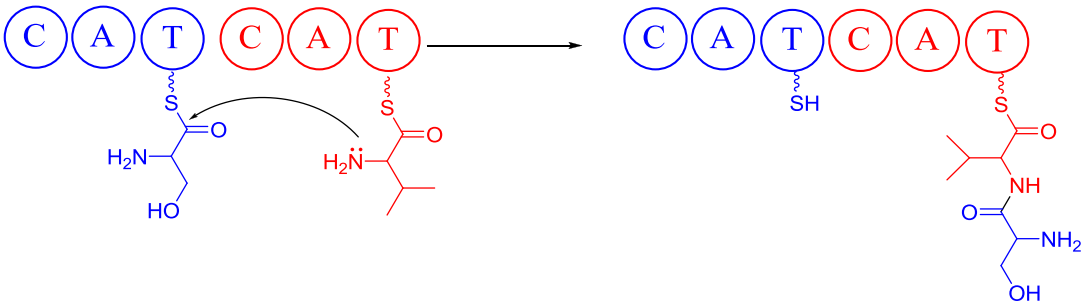



**Figure 2-2. Mechanism of chain elongation in nonribosomal peptide synthetases.**
Colors of the modules correspond to the building blocks added.

The traditional approach of drug discovery of natural product origin from microorganisms involves the successive steps of bioassay-guided fractionation of an organic extract of an inoculated growth culture until a single active compound is obtained. To elucidate the chemical structure, various spectroscopic techniques including 1 and 2-D nuclear magnetic resonance (NMR) and high resolution mass spectrometry (HR-MS) are used. This occasionally leads to the replication and rediscovery of already known compounds. Once the structure has been elucidated, the biosynthetic gene cluster can be identified through cloning and sequencing strategies.

Recently, and with the improvement in sequencing technology, rather than using the chemical structure to predict the biosynthetic pathway, the opposite approach has been used in what is called as genome mining (Figure 2-3). Genome mining for natural product metabolites and biosynthetic pathways uses bioinformatic analysis of microbial genomes to predict the potential production of secondary metabolites. It takes advantage of the modular architecture of PKSs and NRPSs and has rapidly advanced in the last few years with the huge increase in the number of microbial genomes sequenced (1863 microbial genome have been sequenced while 5230 are in progress http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi ; accessed 01/2012).

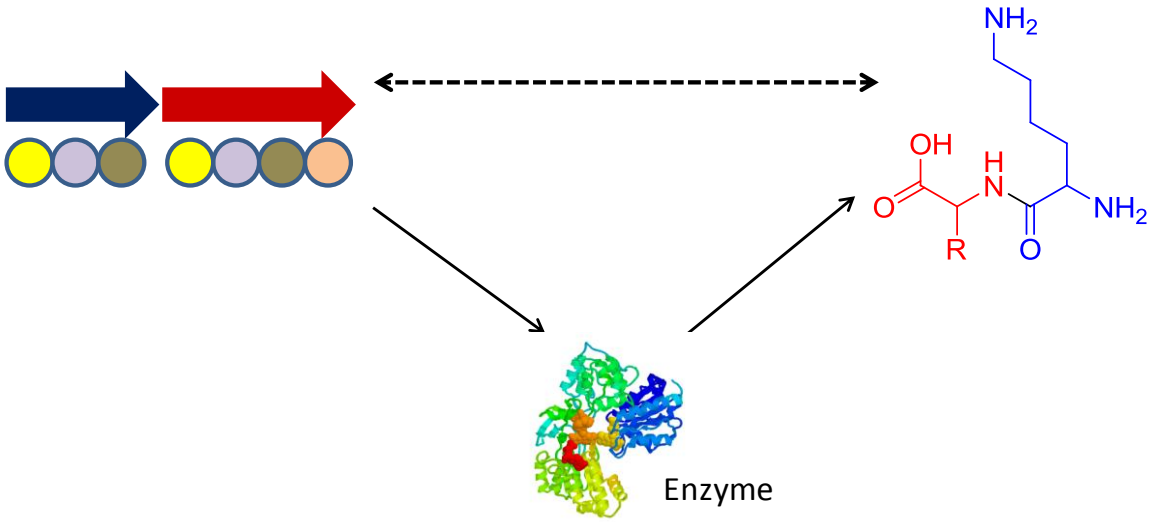



**Figure 2-3. The relation between genes and small molecules.**
Genes organized closely together encode large enzymes such as PKSs and NRPSs. Arrows represent open reading frames while circles represent domains organized in modular architecture. A representative metabolite is shown.

There are several approaches utilized to isolate and identify the possible products of such "cryptic" or "orphan" clusters in these strains. One is the mutational approach where a gene in the target pathway is disrupted through gene deletion or insertion. The drawbacks of this approach are the difficulty of genetic manipulation of some strains and the repression of some pathways under laboratory growth conditions (silent clusters) (Laureti et al., 2011). Another approach is the regulatory approach involving manipulation of transcription regulators, especially Lux homologues (Mesak et al., 2010). Most gene clusters contain either positive or negative regulators that control the transcription of these genes, and thus can be altered to express the genes and consequently the related metabolite. Deletion of negative regulators has been used before to increase the biosynthesis of natural products as in the case of kinamycins (Bunet et al., 2011), thailandamides (Ishida et al., 2010), and bactobolin (Seyedsayamdost et al., 2010). On the

other hand, induction of positive regulators has led to increased production of other metabolites such as fosfomycin (Woodyer et al., 2006) and pimaricin (Antón et al., 2007). In fact, constitutive expression of promoters in *S. ambofaciens* led to the activation of a silent gene cluster and the identification of new compounds: stambomycins (Laureti et al., 2011). In both approaches, comparative metabolic profiles of the genetically manipulated strain and wild type crude extracts are analyzed through HR-MS or HPLC/MS to identify the compound biosynthesized by the target cluster.

A phylogeny approach could also be used to confirm or identify products of gene clusters by comparing the similarity of gene sequences with other clusters where the product has been identified. This could lead to successful detection of predicted compounds or similar ones such as in the case of frontalamides (Blodgett et al., 2010) and clifednamides (Cao et al., 2010). Finally, a chemical structure approach could be used. This is based on analyzing the gene sequences and conserved residues of domains and predicting the compound that could be produced and identifying it using a characteristic spectroscopic feature of the predicted compound or through a predicted labeled substrate feeding study. This approach led to identification of trichamide based on HR-MS (Sudek et al., 2006), salinilactam A and thailandamides using a specific UV absorbance (Udwary et al., 2007; Nguyen et al., 2008) and orfamide A through isotope-guided fractionation (Gross et al., 2007). The drawback of the previous two approaches is that it is not always possible to accurately predict the specificity or functionality of these domains which may lead to unexpected products.

The total microbial community of *Bankia setacea* shipworm was sequenced. In addition, the genomes of the four shipworm symbionts; *Teredinibacter turnerae* T7901, and three

other undescribed shipworm isolates species; BS02, BS08, and BS12 isolated from the shipworms *B. gouldi* in case of *T. turnerae* T7901*, and B. setacea* in case of the other three, were sequenced, assembled, and autoannotated. These bacteria are Gram negative, gammaproteobacterial intracellular endosymbionts that are located in the shipworm gills (Distel et al., 2002; Betcher, 2011). These strains were chosen to be sequenced because they contribute significantly to the composition of the microbial community of the shipworm gill as indicated by cultivation independent analysis. My goal was to investigate the potential biosynthesis of natural products, especially those originating from PKSs and NRPSs, in these microorganisms. In addition, thorough analysis of the domains and modules of the regions, in addition to the conserved residues of the adenylation domains might provide some insights of the structure produced.

## 2.2. Materials and Methods

*B. setacea*, and subsequently the isolated strains BS02, BS08, and BS12 were collected from Puget Sound, Washington, USA. *T. turnerae* T7901 was isolated as previously described (Waterbury et al., 1983). The genome of the latter strain was sequenced using Sanger and Roche-454 GS20 and assembled using Celera Assembler and J. Craig Venter Institute's in-house hybrid assembly method as reported previously (Yang et al., 2009). The genomes of the three strains: BS02, BS08, and BS12 were sequenced using Illumina GAii. Velvet was used to assemble BS02 and BS08 while MERaculous was used to assemble BS12. Libraries were produced by Illumina with average reads of 76 bp for BS02 and BS08 and of 100 bp for BS12. The metagenome of *B. setacea* was sequenced using a combination of 454 Titanium and Illumina. This metagenome was assembled using Newbler and Velvet. The *B. setacea* metagenome analyzed was deposited under the name: *Bankia setacea* microbiome (JGI August sample). All genomes were deposited in the integrated microbial genomes (IMG) database. Genes involved in the biosynthesis of natural products, especially those originating from PKSs and NRPSs, were detected manually through keyword search of the autoannotated genome. Since genes involved in a function are usually clustered together in microorganisms, neighboring genes that could be involved in the biosynthesis of secondary metabolites, upstream and downstream were also analyzed. In addition, the amino acid sequences of the adenylation domains of NRPSs were carefully analyzed. The particular residues that confer substrate specificity (Stachelhaus et al., 1999; Challis et al., 2000) were identified using the online PKS/NRPS tool; (http://nrps.igs.umaryland.edu/nrps). The substrate loaded by each adenylation domain was predicted. The genomes were also analyzed using antiSMASH (antibiotics

and Secondary Metabolite Analysis SHell) (Medema et al., 2011) and local Blast using known PKS and NRPS genes. The *B. setacea* metagenome was also analyzed by the Metagenomics RAST analysis server (MG-RAST http://metagenomics.anl.gov) (Meyer et al., 2008).

## 2.3. Results

### 2.3.1. *T. turnerae* T7901

The total number of genes involved in the biosynthesis, transport, and breakdown of secondary metabolites was estimated to be 380 representing approximately 7% of the total gene (Yang et al., 2009). The genome of *T. turnerae* T7901 is 5.2 Mbp with a GC percentage of 51% (Table 2-1). It was found to contain nine gene clusters predicted to encode PKSs and NRPSs (Figure 2-4). Three of these regions are PKS, three are NRPS, and three appear to be a hybrid of PKS/NRPS. The largest region detected is region 4 which is 74 kbp and is an NRPS while the smallest is the ~ 9 kbp region 9 encoding a PKS open reading frame (ORF) The largest ORF is TERTU_2858 in the NRPS of region 6 that is more than 22 kbp.

**Table 2-1. Genome properties of sequenced shipworm strains.**
*, approximate size

| Property \ Strain | *T. turnerae* T7901 | BS02 | BS08 | BS12 |
|---|---|---|---|---|
| **Scaffold Count** | 1 | 141 | 123 | 240 |
| **Genome Size (Mbp)** | 5.2 | 3.9* | 4.8* | 4.6* |
| **GC Percentage (%)** | 51 | 48 | 47 | 46 |
| **Gene Count** | 4,308 | 3,320 | 4,119 | 4,202 |
| **Protein Coding Genes** | 4,254 | 3,277 | 4,074 | 4,153 |
| **Protein Coding Genes Percentage (%)** | 98.75 | 98.70 | 98.91 | 98.83 |
| **Percentage of Proteins with Predicted Functions (%)** | 70.01 | 72.71 | 67.64 | 66.49 |

The architecture of the PKS and NRPS modules in each region is shown in Table 2-2. A total of 44 adenylation domains were detected. Their conserved residues were determined and their substrate specificity was predicted (Table 2-3). Region 1 consists of 17 ORFs that encodes a mixed derived compound from amino acids and ketides and the NRPS ORF was shown to be closely related to an NRPS ORF in lysobactin and iturin gene clusters from *Lysobacter* species (Hou et al., 2011) and *B. amyloliquefaciens* and *B. subtilis* (Tsuge et al., 2001) with a similarity percentage of 47% and 49%, respectively. Region 2 contains eleven ORFs and was found to be responsible for the biosynthesis of a group of macrolides produced by PKSs (Chapter 3). The two giant clusters, regions 3 and 4 are predicted to be responsible for the formation of large peptides. Region 3 contains 23

ORFs and encodes a mixed PKS/NRPS that is highly similar to the cryptic cluster Sce8257 from *S. cellulosum*. Region 4 contains 11 ORFs encoding NRPS (Figure 2-5) while region 5 contains 13 ORFs and encodes a PKS that is closely similar to the gene cluster of tetronomycin and tetrocarcin A from *Streptomyces* and *Micromonospora*, respectively (Demydchuk et al., 2008; Fang et al., 2008). The NRPS Region 6 contains 13 ORFs (Figure 2-5) while region 7 potentially encode the biosynthesis of siderophores (Han, 2011). The two smallest regions: regions 8 and 9, encode a hybrid PKS/NRPS and a PKS, respectively. The largest ORF in region 9 gave a 51% similarity to *epoD* from *S. cellulosum* involved in the biosynthesis of the anticancer compound, epothilone (Tang et al., 2000).

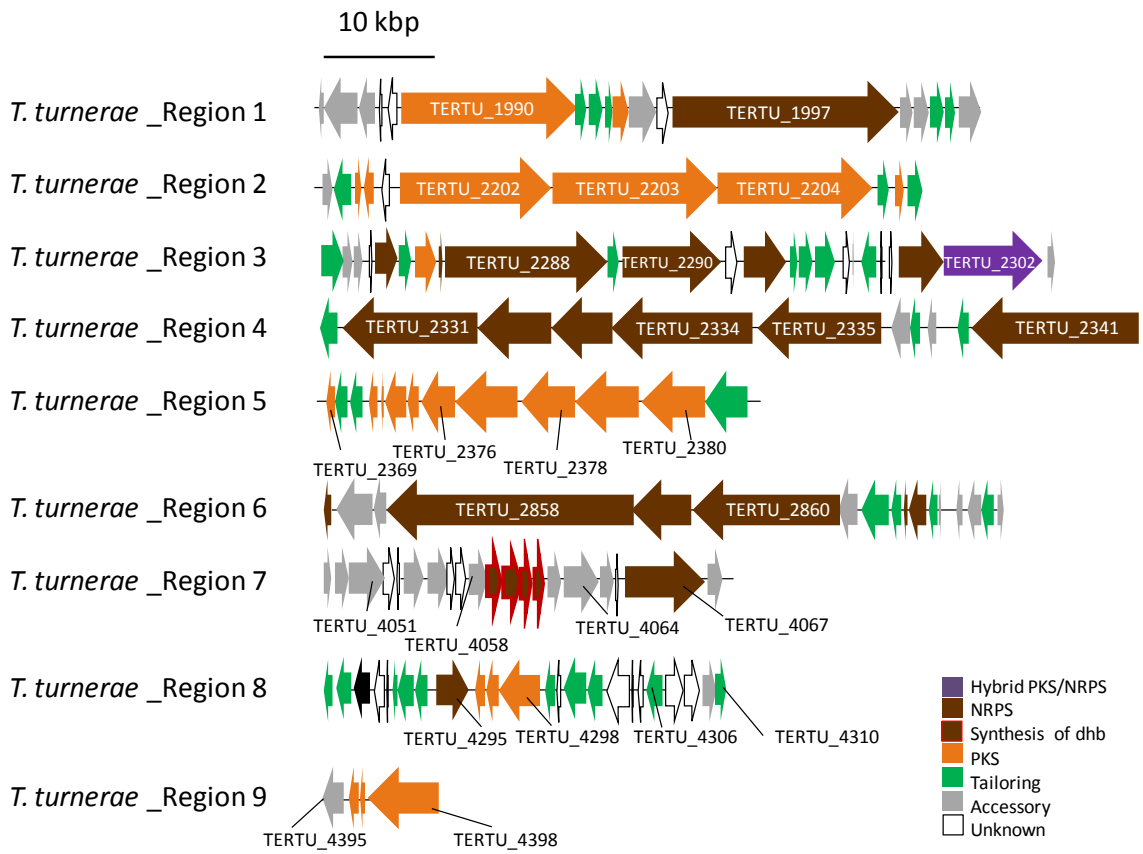



**Figure 2-4. PKS and NRPS gene clusters detected in *T. turnerae* T7901.**
Figure modified from that published previously (Yang et al., 2009); dhb, dihydroxybenzoic acid.

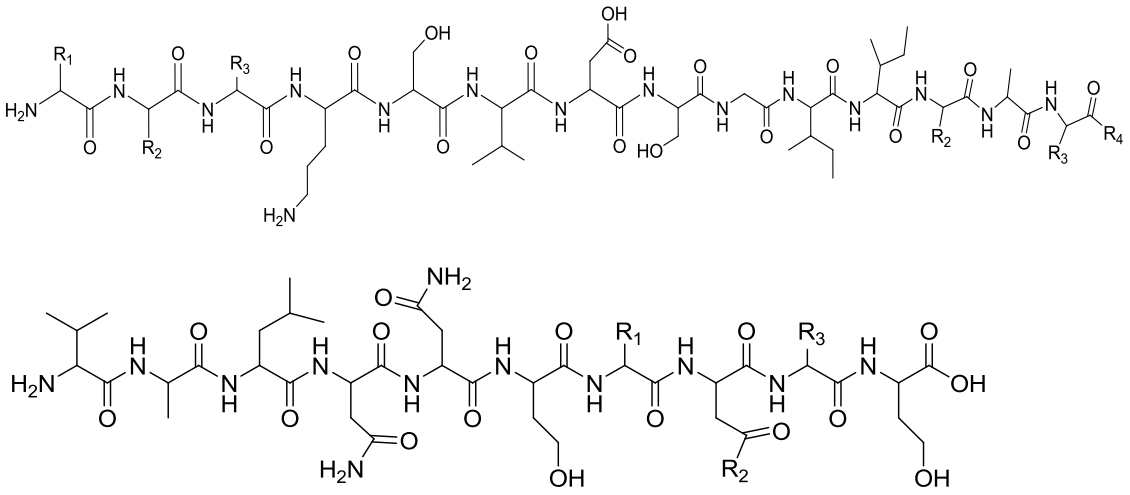


**Figure 2-5. Structures of predicted compounds produced by two regions in *T. turnerae* T7901.**
Predicted compounds produced by regions 4 (top) and 6 (bottom).

**Table 2-2. The domain organization in each ORF of regions of *T. turnerae* T7901.**
*Am: Amino transferase.

| Strain/ ORF Name | Domain Architecture |
|---|---|
| **T7901_R1_TERTU_1990** | KS-DH-KR-ACP-KS-DH-ACP-ACP-KS-DH-KR-ACP-KS |
| **T7901_R1_TERTU_1997** | A-T-E-C-A-T-E-C-A-T-C-A-T-C-A-T-C-ABC transporter |
| **T7901_R2_TERTU_2202** | KS-KR-ACP-KS-DH-KR-ACP-KS |
| **T7901_R2_TERTU_2203** | ACP-KS-DH-KR-ACP-KS-ACP-KS-KR-ACP-KS-ACP-KS |
| **T7901_R2_TERTU_2204** | KR-ACP-KS-DH-KR-MT-KS-DH-KS-ACP-TE |
| **T7901_R3_TERTU_2288** | C-A-T-C-A-T-C-A-T-C-A-T-C |
| **T7901_R3_TERTU_2290** | A-T-E-C-A-T-C |
| **T7901_R3_TERTU_2292** | A-T-C |
| **T7901_R3_TERTU_2301** | C-A-T |
| **T7901_R3_TERTU_2302** | KS-AT-T-C-A-T-C |
| **T7901_R4_TERTU_2340** | A-T-C-C-A-T-C-A-T-C-A-T-C |
| **T7901_R4_TERTU_2335** | C-A-T-E-C-A-T-C-A-T |
| **T7901_R4_TERTU_2334** | C-A-T-E-C-A-T-E-C-A-T |
| **T7901_R4_TERTU_2333** | C-A-T-C-A-T |
| **T7901_R4_TERTU_2332** | C-A-T-E-C-A-T |
| **T7901_R4_TERTU_2331** | C-A-T-C-A-T-E-C-A-T-TE |
| **T7901_R5_TERTU_2381** | Am*-ACP |
| **T7901_R5_TERTU_2380** | KS-AT-DH-KR-ACP |
| **T7901_R5_TERTU_2379** | KS-AT-DH-KR-ACP |
| **T7901_R5_TERTU_2378** | KS-AT-KR-ACP |
| **T7901_R5_TERTU_2377** | KS-AT-DH-KR-ACP |
| **T7901_R5_TERTU_2376** | KS-AT-ACP |
| **T7901_R6_TERTU_2860** | C-A-T-C-C-A-T-C-T-C-A-T |
| **T7901_R6_TERTU_2859** | C-T-C-A-T |
| **T7901_R6_TERTU_2858** | C-T-C-A-T-C-A-T-C-A-T-C-T-C-A-T-C-A-T-C |
| **T7901_R7_TERTU_4067** | C-A-T-C-A-T |
| **T7901_R8_TERTU_4295** | C-A-T |
| **T7901_R8_TERTU_4298** | KS-AT-ACP-TE |
| **T7901_R9_TERTU_4398** | KS-AT-DH-ER-KR-ACP |

**Table 2-3. Adenylation domains in *T. turnerae* T7901, their location, binding pocket residues, and predicted substrate loaded.**

* confirmed by compound isolation (Han, 2011); NA, Not accurately predicted.

| Strain/Region | Gene Object ID | Binding Pocket Residues | Substrate Prediction |
|---|---|---|---|
| **T7901_R1** | TERTU_1997-A1 | D A A V A A A V | NA |
| **T7901_R1** | TERTU_1997-A2 | D V Y X M G G - | Valine |
| **T7901_R1** | TERTU_1997-A3 | D A G A L A A V | Leucine/Serine |
| **T7901_R1** | TERTU_1997-A4 | D I W E L T L - | NA |
| **T7901_R1** | TERTU_1997-A5 | D I W E L T A - | Serine/Dab |
| **T7901_R3** | TERTU_2288-A1 | D G F W V G G T | Valine |
| **T7901_R3** | TERTU_2288-A2 | D G F W V G G T | Valine |
| **T7901_R3** | TERTU_2288-A3 | D S F T L G A I | Aromatic amino acid |
| **T7901_R3** | TERTU_2288-A4 | D L Y N M S L I | Cysteine |
| **T7901_R3** | TERTU_2290-A1 | D L T K I G H V | Aspartate/Asparagine |
| **T7901_R3** | TERTU_2290-A2 | D L T K I T H V | Dab |
| **T7901_R3** | TERTU_2292 | D A S T I A A V | Tyrosine |
| **T7901_R3** | TERTU_2301 | D A G A L A G V | NA |
| **T7901_R3** | TERTU_2302 | D V F V H T D D | NA |
| **T7901_R4** | TERTU_2341-A1 | D A A A L A A I | NA |
| **T7901_R4** | TERTU_2341-A2 | D V F W M G G A | Valine |
| **T7901_R4** | TERTU_2341-A3 | D A A A L A G V | Tyrosine/Leucine |
| **T7901_R4** | TERTU_2341-A4 | X X X X V G H V | Aspartate/Alanine |
| **T7901_R4** | TERTU_2335-A1 | D V G E I G S V | Ornithine/Arginine |
| **T7901_R4** | TERTU_2335-A2 | D V W H F S L I | Serine |
| **T7901_R4** | TERTU_2335-A3 | D A L A L G X X | Valine |
| **T7901_R4** | TERTU_2334-A1 | D L T K V G H I | Aspartate |
| **T7901_R4** | TERTU_2334-A2 | D V W Q M I G D | Arginine/Serine |
| **T7901_R4** | TERTU_2334-A3 | D F W N V G M V | Threonine |
| **T7901_R4** | TERTU_2333-A1 | D I F Q X X C I | Hydrophobic amino acid |
| **T7901_R4** | TERTU_2333-A2 | D A F X L G V T | Isoleucine |
| **T7901_R4** | TERTU_2332-A1 | D A F X R G V T | Hydrophobic amino acid |
| **T7901_R4** | TERTU_2332-A2 | D L K N Y G S D | Basic amino acid |
| **T7901_R4** | TERTU_2331-A1 | D L F T I L L T | Alanine |
| **T7901_R4** | TERTU_2331-A2 | D A F X R G V T | Isoleucine |
| **T7901_R4** | TERTU_2331-A3 | D S - D L G V V | NA |
| **T7901_R6** | TERTU_2865 | D A F W L G X X | Valine |
| **T7901_R6** | TERTU_2860-A1 | D L F T N A L - | Alanine |
| **T7901_R6** | TERTU_2860-A2 | D A F L L G X X | Hydrophobic amino acid |
| **T7901_R6** | TERTU_2860-A3 | D A T - L G E V | Asparagine |
| **T7901_R6** | TERTU_2859 | D A T - L G E V | Asparagine |
| **T7901_R6** | TERTU_2858-A1 | D F W N I G M V | Threonine |
| **T7901_R6** | TERTU_2858-A2 | D V E S V G S V | NA |
| **T7901_R6** | TERTU_2858-A3 | D I P G X X H V | Aspartate/Asparagine |
| **T7901_R6** | TERTU_2858-A4 | D A F L L G X X | Hydrohphobic amino acid |
| **T7901_R6** | TERTU_2858-A5 | D F W N V G M V | Threonine |
| **T7901_R7** | TERTU_4067-A1 | D - W D I I L V | Ornithine* |
| **T7901_R7** | TERTU_4067-A2 | D V W H F S L V | Serine* |
| **T7901_R8** | TERTU_4295 | D V W H I S L I | Serine |

*T. turnerae* T7901 also contains the genes for the synthesis of bacteriocins. In addition, the terpenoid pathway for the synthesis of squalene/hopanoid molecules was detected (Figure 2-6) and the genes involved in their biosynthesis were determined (Table 2-4). In addition, the genes *ectA/B/C* that are responsible for the biosynthesis of ectoine were also detected (Figure 2-7).

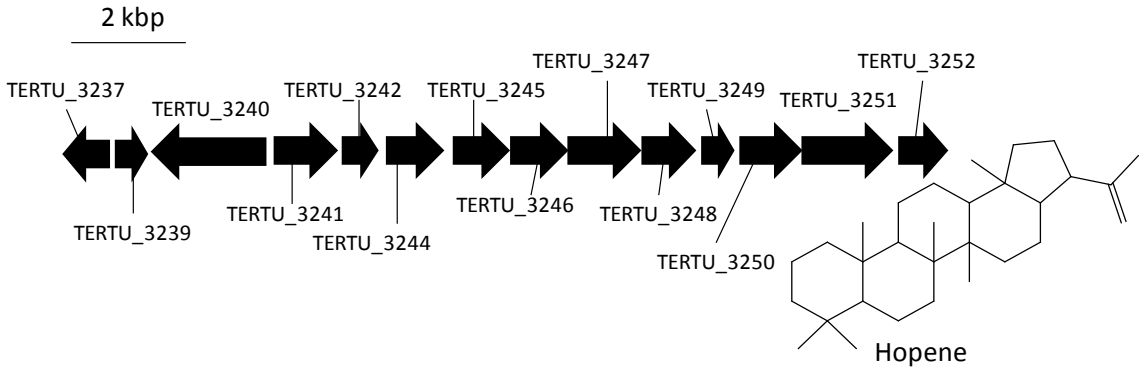



**Figure 2-6. Gene cluster and the predicted chemical structure of the terpenoid molecule in *T. turnerae* T7901.**

**Table 2-4. Name, function, and size of genes involved in the terpenoid biosynthesis.**

| Gene Name | Number of amino acids | Proposed function |
|---|---|---|
| **TERTU_3237** | 282 | hopanoid biosynthesis associated protein HpnK/Unkn |
| **TERTU_3239** | 194 | Unknown |
| **TERTU_3240** | 671 | squalene-hopene cyclase |
| **TERTU_3241** | 375 | squalene synthase |
| **TERTU_3242** | 218 | hopanoid associated phosphorylase |
| **TERTU_3244** | 341 | hopanoid associated epimerse/DH |
| **TERTU_3245** | 335 | diphosphomevalonate decarboxylase |
| **TERTU_3246** | 346 | phosphomevalonate kinase |
| **TERTU_3247** | 432 | hydroxymethylglutaryl-CoA reductase |
| **TERTU_3248** | 318 | mevalonate kinase |
| **TERTU_3249** | 196 | isopentenyldiphosphate synthase |
| **TERTU_3250** | 372 | hopanoid biosynthesis associated radical SAM protein HpnH |
| **TERTU_3251** | 535 | hydroxymethylglutaryl-CoA synthase |
| **TERTU_3252** | 295 | Trans-isoprenyl diphosphate synthases |


**Figure 2-7. Genes responsible for the biosynthesis of ectoine and its predicted chemical structure.**

### 2.3.2. BS02

The genome of BS02 is estimated to be 3.9 Mbp with a GC percentage of 48% (Table 2-1). The genome sequencing of BS02 indicates that there are three ketosynthase domains; one of them resides in a hybrid PKS/NRPS pathway. In addition, three adenylation domains are arranged in what appear to be three clusters. Region 1 (Figure 2-8) seems to be arranged similarly as that of region 8 in *T. turnerae* T7901. It is an NRPS having the three typical domains C-A-T with a thioesterase for the release of the compound (Table 2-5). Upstream of this NRPS module are potential glycosyl transferases, involved in the possible introduction of sugar units to the aglycone and the formation of glycosides. The gene cluster seems to be highly similar to region 8 in *T. turnerae* T7901. Downstream of the NRPS module is a prenyl transferase predicted to load a prenyl unit suggesting a compound composed of terpenoid, peptide, and sugar units, assuming that the terpenoid and glycosyl transferase are included in the same cluster. The same arrangement including the NRPS, prenyl, and glycosyl transferase were similar to corresponding regions in region 8 of *T. turnerae* T7901. However, no PKS is detected within region 1 of BS02 as with that in region 8.

Region 2 of BS02 (Figure 2-8) is a hybrid PKS/NRPS with architecture organized as C-T-TE-A-KS-AT (Table 2-5). Blast analysis shows that the closest match to this ORF is from *Paenibacillus mucilaginosus* KNP414 chromosome. The adenylation domain is predicted to load serine (Table 2-6) while the specificity of the AT is predicted to load a malonyl unit. Region 3 is an NRPS (Figure 2-8) where potential modifying domains are detected downstream, including MT, E, DH, and reductase. BS02 also contains the genes for the biosynthesis of ectoine (Figure 2-7).

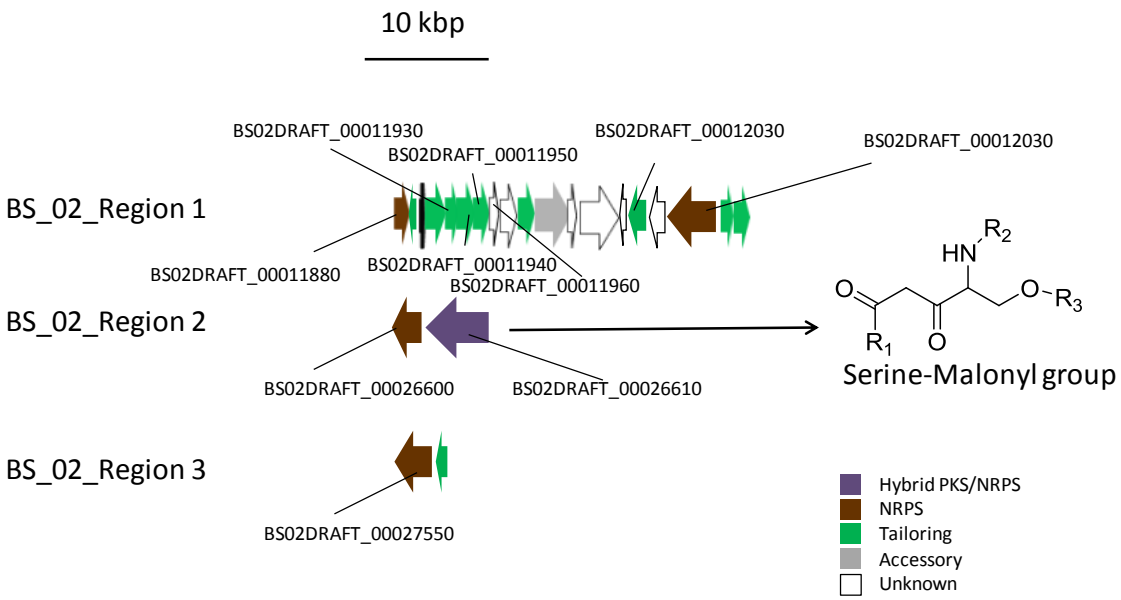



**Figure 2-8. PKS and NRPS gene clusters detected in BS02 and predicted intermediate produced.** Genes are identified by locus tags.

### 2.3.3. BS08

The genome size of BS08 is estimated to be 4.8 Mbp with a GC percentage of 47% consistent with those from gammaproteobacteria (Table 2-1). The genome contains six ORFs coding for NRPSs and five for PKSs. The genome of BS08 contains five potential clusters specifying PKSs and/or NRPSs.

The sequence of region 1 in BS08 (Figure 2-9) suggests the presence of a siderophore. This cluster is highly similar to the vanchrobactin gene cluster from *L. anguillarum* and has similar gene arrangement to the turnerbactin gene cluster from *T. turnerae* T7901. It contains ORFs responsible for the biosynthesis of dihydroxybenzoic acid, the main subunit of catechol siderophores, in addition to NRPS modules with two adenylation domains loading serine and another amino acid not accurately predicted (Table 2-6). This potential siderophore cluster also contains a gene encoding a TonB-like receptor.

The second region is a PKS/NRPS hybrid cluster. It contains two large ORFs; the first ORF is a PKS while the second ORF is an NRPS. It contains five complete modules of NRPSs where the third module contains an epimerase domain. This region, as with region 1 from *T. turnerae* T7901, contains the unusual ABC transporter integrated in the same ORF together with five other modules of NRPS, not as a separate gene. Region 3 (Figure 2-9) is a hybrid PKS/NRPS. The third ORF contains three adenylation domains and a PKS module suggesting a compound with an unsaturated moeity. The fourth ORF is an NRPS.

Region 4 in BS08 (Figure 2-9) is a hybrid PKS/NRPS. It contains two large ORFs. The first is a complete PKS made out of nine modules (Table 2-5) with size of more than 42

kbp (Table 2-7). The second ORF is a hybrid PKS/NRPS with one adenylation domain and four modules of PKS. In addition, several other modifying genes are scattered upstream and downstream of these ORFs. This cluster contains an unusual N-terminus TE domain. The fifth region in BS08 (Figure 2-9) is also a hybrid PKS/NRPS that contains one module of PKS and one module of NRPS, in addition to a 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) synthetase and genes encoding modifying enzymes with DH and two domains of MT present upstream and downstream of this region. BS08 also contains genes necessary for the synthesis of ectoine as in *T. turnerae* T7901 and BS02.

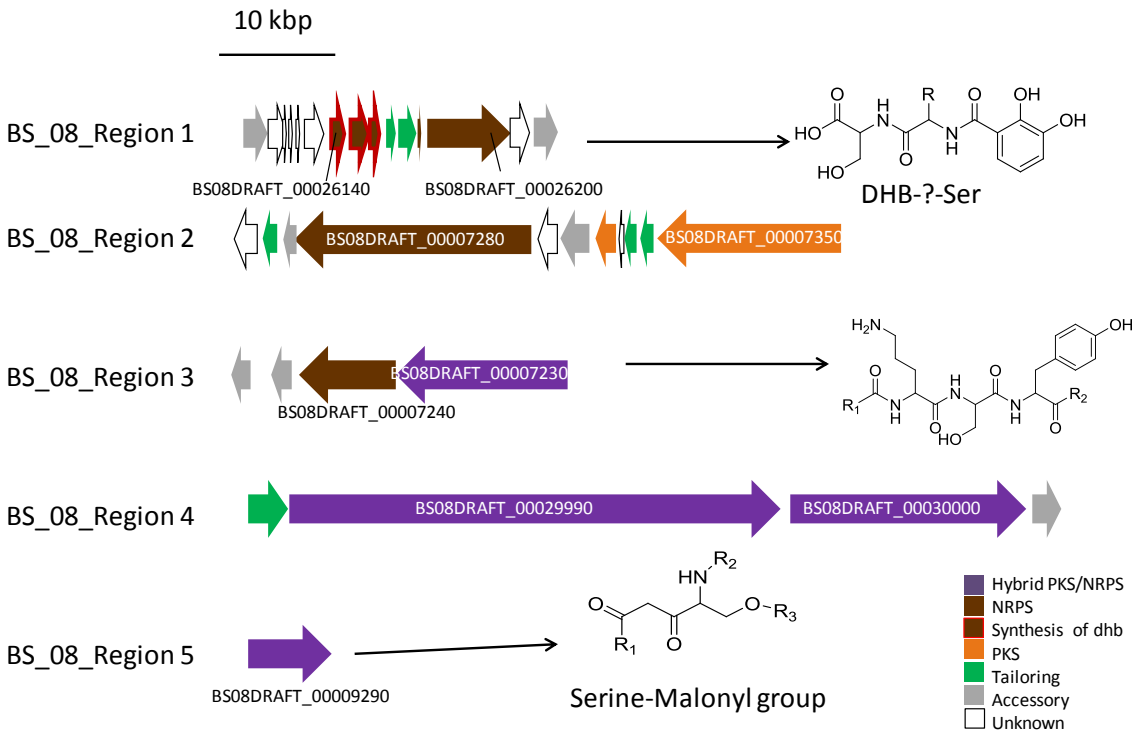



**Figure 2-9. PKS and NRPS gene clusters detected in BS08 and predicted intermediates produced.** Genes identified by locus tags.

### 2.3.4. BS12

The genome size of BS12 is estimated to be 4.6 Mbp with a GC percentage of 46% (Table 2-1). The genome of BS12 contains nine genes predicted to encode NRPSs and six to encode PKSs. Eight of the NRPSs are arranged in what appears to be at least three clusters and one in a hybrid form with PKS (Table 2-5). These nine NRPS genes contain 13 adenylation domains (Table 2-6).

Region 1 is a PKS containing five genes forming a total of five modules that are clustered together in one region preceded by an amino transferase domain (Figure 2-10). The second region is an NRPS (Figure 2-10) where its architecture suggests a potential siderophore biosynthesis cluster of the type involved in catecholate synthesis. This is supported by the presence of chorismate synthase genes and iron siderophore transport receptors within the cluster. Region 3 contains two modules with two adenylation domains in addition to an ORF encoding a multidrug efflux transporter. Region 4 is a one module NRPS suggesting an iterative NRPS enzyme. Region 5 is a hybrid PKS/NRPS. The NRPS portion contains two adenylation domains, in addition to one KS module with modifying domains KR, DH, and a possible ER suggesting a saturated moiety in the product. Region 6 is an NRPS that contains six modules. The largest ORF in this region is more than 13 kbp. In addition, a dioxygenase gene is detected that is not involved with any cluster and is similar to the PhyH predicted to be involved in the biosynthesis of mitomycin antibiotics. This gene might exist within one of the secondary metabolite regions but might have been misassembled.

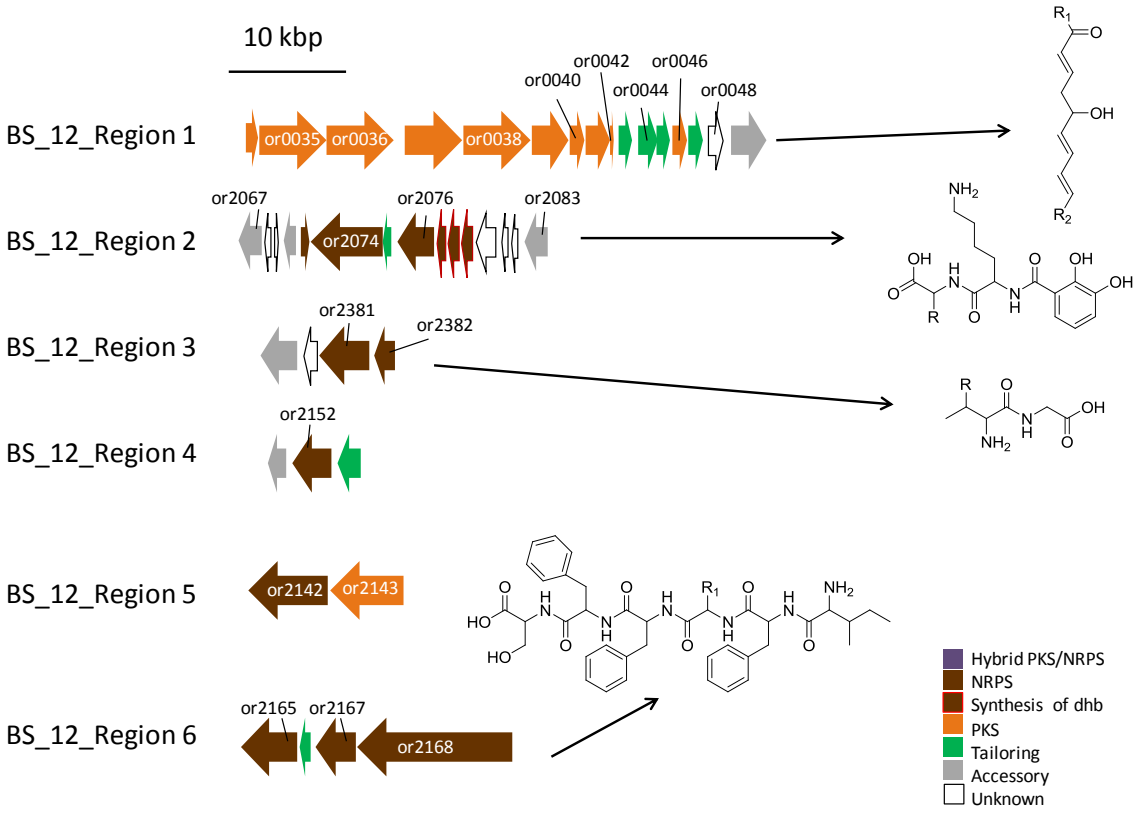



**Figure 2-10. PKS and NRPS gene clusters detected in BS12 and predicted intermediates produced.** Genes identified by locus tags.

**Table 2-5. The domain organization of each ORF in regions present in BS02, BS08, and BS12.** Am*, amino transferase.

| Strain/Region | Locus Tag | Domain Architecture |
|---|---|---|
| **BS_02_R1** | BS02DRAFT_00012030 | C-A-T-TE |
| **BS_02_R2_ORF1** | BS02DRAFT_00026610 | A-T-KS-AT-ACP |
| **BS_02_R2_ORF2** | BS02DRAFT_00026600 | C-P-TE |
| **BS_02_R3** | BS02DRAFT_00027550 | C-A-T |
| **BS_08_R1** | BS08DRAFT_00026200 | C-A-T-C-A-T-TE |
| **BS_08_R2_ORF1** | BS08DRAFT_00007350 | KS-DH-KR-ACP-KS-DH-ACP-ACP-KS-DH-KR-ACP-KS |
| **BS_08_R2_ORF8** | BS08DRAFT_00007280 | A-T-E-C-A-T-E-C-A-T-C-A-T-C-A-T-C-ABC transporter |
| **BS_08_R3_ORF3** | BS08DRAFT_00007230 | A-T-KS-AT-DH-ACP-AmTr-C-A-T-C-A-T |
| **BS_08_R3_ORF4** | BS08DRAFT_00007240 | C-A-T-E-C-T-TE |
| **BS_08_R4_ORF1** | BS08DRAFT_00029980 | ACP-TE |
| **BS_08_R4_ORF2** | BS08DRAFT_00029990 | ACP-KS-DH-KR-ACP-KS-ACP-KS-DH-KR-ACP-KS-DH-ACP-KS-KR-ACP-KS-DH-ACP-KS-DH-KR-ACP-KS-ACP-ACP-ACP-KS-ACP-C-C |
| **BS_08_R4_ORF3** | BS08DRAFT_00030000 | A-ACP-ACP-ACP-KS-ACP-KS-DH-KR-ACP-KS-DH-KR-ACP-KS-DH-ACP-KS |
| **BS_08_R5** | BS08DRAFT_00009290 | A-T-KS-AT-ACP |
| **BS_12_R1_ORF1** | or0034 | Am*-ACP |
| **BS_12_R1_ORF2** | or0035 | KS-AT-DH-KR-ACP |
| **BS_12_R1_ORF3** | or0036 | KS-AT-DH-KR-ACP |
| **BS_12_R1_ORF4** | or0037 | KS-AT-KR-ACP |
| **BS_12_R1_ORF5** | or0038 | KS-AT-DH-KR-ACP |
| **BS_12_R1_ORF6** | or0039 | KS-AT-ACP |
| **BS_12_R2_ORF8** | or2076 | C-A-T-E-C-T |
| **BS_12_R2_ORF10** | or2074 | C-A-T |
| **BS_12_R3_ORF1** | or2382 | A-T |
| **BS_12_R3_ORF2** | or2381 | C-A-T-TE |
| **BS_12_R4** | or2152 | C-A-T |
| **BS_12_R5_ORF1** | or2143 | KS-AT-DH-ER-KR-ACP |
| **BS_12_R5_ORF2** | or2142 | C-A-T-C-A |
| **BS_12_R6_ORF1** | or2168 | C-A-T-C-A-T-C-A-T-C-A-T |
| **BS_12_R6_ORF2** | or2167 | C-A-T |
| **BS_12_R6_ORF4** | or2165 | C-A-T |

**Table 2-6. Adenylation domains in BS02, BS08, and BS12.**

Gene ID, binding pocket residues, and predicted substrates loaded are described. NA, Not accurately determined.

| Strain/ Region | Gene Object ID | Binding Pocket Residues | Substrate Specificity |
|---|---|---|---|
| **BS_02_R1** | 2503998025 | D V E N V G G I | NA |
| **BS_02_R2** | 2503998956 | D V W H M S L I | Serine |
| **BS_02_R3** | 2504000774 | M G Y V I G X X | NA |
| **BS_08_R1** | 2503993942-A1 | D S W D I I X X | Glutamate/ lysine |
| **BS_08_R1** | 2503993942-A2 | D V W H F S L V | Serine |
| **BS_08_R2** | 2503997029-A1 | D A A V A A A V | Tyrosine/ Alanine |
| **BS_08_R2** | 2503997029-A2 | D V Y X M G G - | Val |
| **BS_08_R2** | 2503997029-A3 | D A G A L A A I | Valine/ Alanine |
| **BS_08_R2** | 2503997029-A4 | D I W E L T A - | Serine |
| **BS_08_R2** | 2503997029-A5 | D I W E L T A - | Serine |
| **BS_08_R2** | 2503997024-A1 | D V G E I G T I | Ornithine |
| **BS_08_R2** | 2503997024-A2 | D S W D L G C I | Serine/ Glycine |
| **BS_08_R2** | 2503997023 | D A S T I A A V | Tyrosine |
| **BS_08_R3** | 2503994486 | D V W H F S L V | Serine |
| **BS_08_R4** | 2503995728 | D V W H F S L V | Serine |
| **BS_12_R2** | 2505680920 | D A M V G G C V | Lysine |
| **BS_12_R2** | 2505680922 | D A E D I G T V | Leucine/ Phenylalanine |
| **BS_12_R3** | 2505681226 | D A F F L G G T | Isoleucine/ Valine |
| **BS_12_R3** | 2505681225 | D I L Q – G M V | Glycine |
| **BS_12_R4** | 2505680997 | G V L H L G L V | HPG/ Alanine |
| **BS_12_R5** | 2505680988-A-1 | D V W N N A M I | Serine/ Threonine |
| **BS_12_R5** | 2505680988-A2 | D A L F I G A T | Valine |
| **BS_12_R6** | 2505681013–A1 | D A F F L G V T | Isoleucine/ Valine |
| **BS_12_R6** | 2505681013–A2 | D A W T V A A V | Phenylalanine |
| **BS_12_R6** | 2505681013–A3 | D S A X I A E V | No hit |
| **BS_12_R6** | 2505681013–A-4 | D A W T V A A V | Phenylalanine |
| **BS_12_R6** | 2505681012 | D A S T I A A V | Tyrosine/ Phenylalanine |
| **BS_12_R6** | 2505681010 | D V W H F S L I | Serine |

### 2.3.5. *Bankia setacea* metagenome

#### 2.3.5.1. MG RAST data

The total microbial community in the gills of *B. setacea* was analyzed using MG-RAST. The majority of the reads has a GC percentage in the range of 45-50% consistent with bacterial sequences (Figure 2-11) belonging to proteobacteria. The sequence reads of the proteins and rRNA predicted to belong to bacteria represent 97.8% of total reads as expected. About 1.7% of the reads belong to eukaryotes, probably from the shipworm itself. Surprisingly, reads that belong to phylum *Euryarcheota* are detected, although rare (0.3%). No archaea have ever been detected in shipworms before (Betcher, 2011). Proteobacteria present 86.3% of the total reads while actinobacteria and cyanobacteria, known to produce small molecules of interesting pharmacological activities present 2.3% and 1.5% of total phyla, respectively (Figure 2-12). On the class level, gammaproteobacteria is the major contributor to the microbial community in the shipworm with 73.3%, while alpha and betaproteobacteria represent 4.5 and 5.1% of the total reads, respectively. Genus level analysis shows similarity to *Saccharophagus*, *Pseudomonas*, *Teredinibacter*, *Cellvibrio*, and *Shewanella* ranging from 12.7 to 5% of the total bacterial community (Figure 2-13). In addition, sequences that are predicted to belong to *Photorhabdus* and *Xenorhabdus* based on gene similarity are detected. Bacteria that belong to these genera are known to produce secondary metabolites (Bode, 2009).

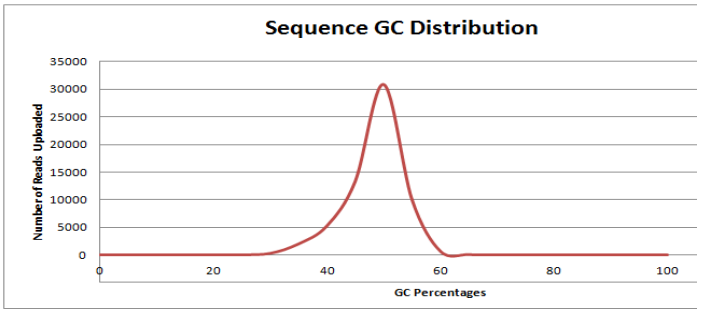



**Figure 2-11. Sequence GC distribution showing the average percentage of GC content of all reads of the *B. setacea* metagenome.**

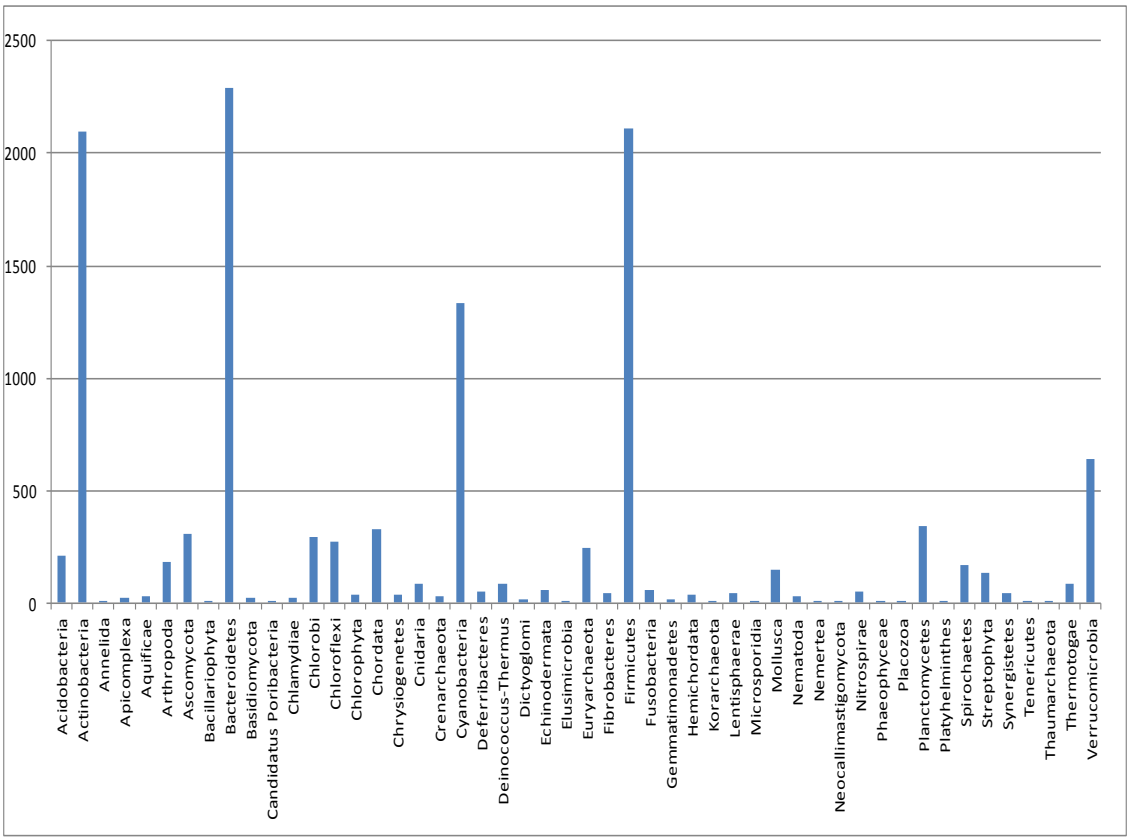



**Figure 2-12. Phyla distribution of the non-proteobacterial microbial community in the shipworm metagenome.**

The Proteobacterial phylum was excluded due to its high abundance.

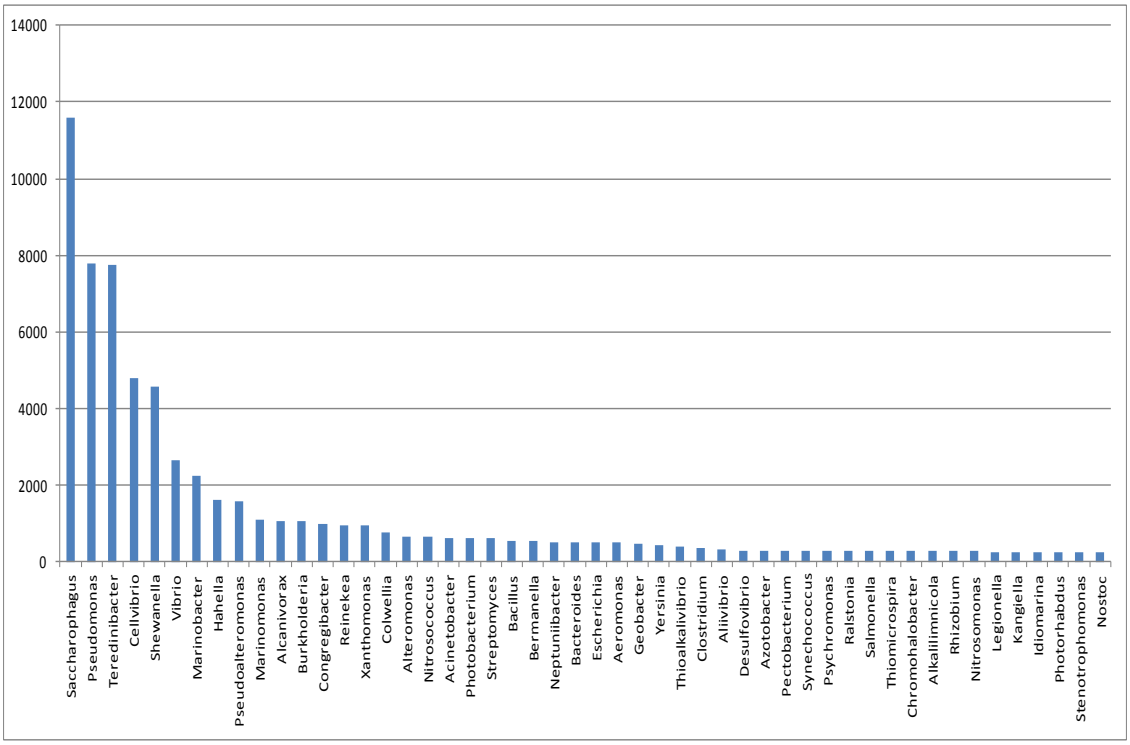



**Figure 2-13. Genus distribution of the microbial community in the shipworm metagenome.** Based on the top 50 hits.

#### 2.3.5.2. Analysis for secondary metabolites

The size of this metagenome is estimated to be 26.5 Mbp. The total gill microbial community of *B. setacea* metagone sequences provided 16 ORFs encoding NRPSs and 16 encoding PKSs. The metagenome was analyzed more closely for large genes arranged that might encode the production of PKs or NRPs. Three gene clusters were detected (Figure 2-14). Region 1 encodes an NRPS that includes genes potentially involved in a siderophore production. The arrangement of this region is similar to that of BS12 but lacking one module which might be due to misassembly of complex metagenomes. Region 2 is an NRPS that contains three ORFs where the largest ORF is more than 6 kbp (Table 2-7). The region is formed from five modules expected to load five amino acids (Table 2-8). Region 3 encodes a hybrid NRPS/PKS in the form of A-P-KS (Table 2-9). All binding pocket residues were detected and the loaded substrates were predicted (Table 2-8). In addition, the genes for the synthesis of ectoine (Figure 2-7) are detected as expected from the presence of these genes in all previous strains. Moreover, genes for the synthesis of a lantibiotic are detected.

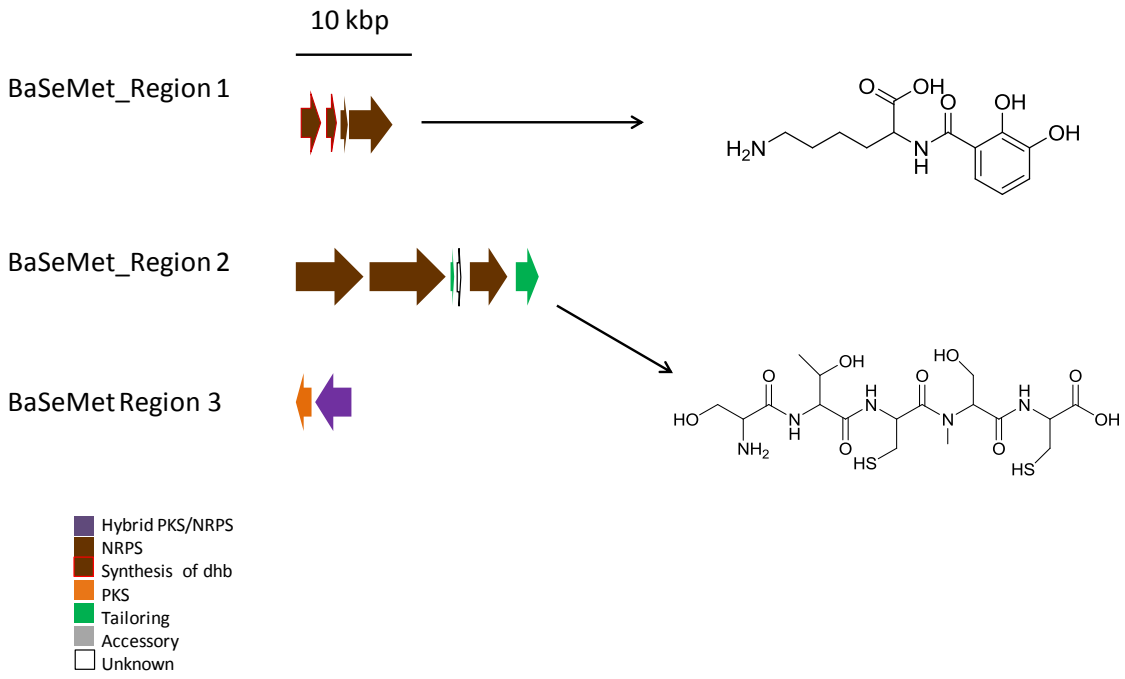



**Figure 2-14. PKS and NRPS gene clusters detected in *B. setacea* metagenome and the predicted intermediates produced.**

**Table 2-7. Secondary metabolite regions in symbionts indicating size of largest ORF and most similar compounds predicted.**

NA, not accurately determined

1: confirmed, chapter 3

2: confirmed (Han, 2011).

| Strain | Region | Cluster Type | IMG Gene ID or locus tag of largest ORF (size in bp) | Most similar compound |
|---|---|---|---|---|
| **T7901** | 1 | NRPS/PKS (trans AT) | TERTU_1997 (20442) | Lysobactin/Iturin |
| **T7901** | 2 | Type I PKS (trans AT) | TERTU_2203 (14844) | Tartrolons/ Antibacterial[1] |
| **T7901** | 3 | NRPS/Type I PKS | TERTU_2302 (8892) | NA |
| **T7901** | 4 | NRPS | TERTU_2331 (12078) | NA |
| **T7901** | 5 | Type I PKS | TERTU_2380 (5688) | Tetronomycin/ Antibacterial |
| **T7901** | 6 | NRPS | TERTU_2858 (22341) | NA |
| **T7901** | 7 | NRPS | TERTU_4067 (7200) | Turnerbactin/ Siderophore[2] |
| **T7901** | 8 | NRPS/Type I PKS | TERTU_4298 (3735) | NA |
| **T7901** | 9 | Type I PKS | TERTU_4398 (6426) | Epothilone/ Anticancer |
| **BS_02** | 1 | NRPS | 2503998025 (4053) | NA |
| **BS_02** | 2 | NRPS/Type I PKS | 2503998956 (5061) | NA |
| **BS_02** | 3 | NRPS | 2504000774 (3204) | NA |
| **BS_08** | 1 | NRPS | 2503993942 (7200) | Vanchrobactin/ Siderophore |
| **BS_08** | 2 | NRPS/Type I PKS-trans AT | 2503997029 (20463) | Lysobactin/Iturin |
| **BS_08** | 3 | NRPS/Type I PKS | 2503997024 (14820) | NA |
| **BS_08** | 4 | NRPS/Type I PKS-trans AT | 2503994485 (42339) | NA |
| **BS_08** | 5 | NRPS/PKS | 2503995728 (7122) | NA |
| **BS_12** | 1 | Type I PKS | 2506460487 (5775) | Tetronomycin/ Antibacterial |
| **BS_12** | 2 | NRPS | 2506459223 (6243) | Siderophore |
| **BS_12** | 3 | NRPS | 2506459536 (4335) | NA |
| **BS_12** | 4 | NRPS | 2506459301 (3408) | NA |
| **BS_12** | 5 | NRPS/Type I PKS | 2506459291 (6828) | NA |
| **BS_12** | 6 | NRPS | 2506459317 (13242) | NA |
| **BaSet_MG** | 1 | NRPS | 2079088531 (3795) | Siderophore |
| **BaSet_MG** | 2 | NRPS | 2079087845 (6630) | NA |
| **BaSet_MG** | 3 | PKS/NRPS | 2079087776 (3054) | NA |

**Table 2-8. Adenylation domains organized in a gene cluster form in *B. setacea* metagenome.**
The ID of these genes, binding pocket residues, and predicted substrates loaded are recorded.

| Strain/Region | Gene Object ID | Binding Pocket Residues | Substrate Specificity |
|---|---|---|---|
| **BaSet_MG_R1** | 2079088531 | D A E D I G T V | Lysine/Arginine |
| **BaSet_MG_R2** | 2079087844-A1 | D V W H V S L I | Serine |
| **BaSet_MG_R2** | 2079087844-A2 | D F W N I G M V | Threonine |
| **BaSet_MG_R2** | 2079087845-A1 | D L Y X L S X X | Cysteine |
| **BaSet_MG_R2** | 2079087845-A2 | D F W N I G M V | Threonine |
| **BaSet_MG_R2** | 2079087848 | D L Y N L S X X | Cysteine |
| **BaSet_MG_R3** | 2079087776 | D V W H F S L V | Serine |

**Table 2-9. The domain organization in each ORF of regions in *B. setacea* metagenome.**

| Strain/Region | Domain Architecture |
|---|---|
| **BaSet_MG_R1_ORF3** | C |
| **BaSet_MG_R1_ORF4** | A-T-TE |
| **BaSet_MG_R2_ORF1** | A-T-C-A-T-C |
| **BaSet_MG_R2_ORF2** | A-T-C-A-MT |
| **BaSet_MG_R2_ORF3** | C-A-T |
| **BaSet_MG_R3_ORF1** | A-T-KS |
| **BaSet_MG_R3_ORF2** | AT |

## 2.5. Discussion

Based on extensive microscopy and phylogeny (Betcher, 2011), the microbial community in the shipworm gill consists of intracellular symbionts which are gammaproteobacteria related to *T. turnerae.* In addition to the gammaproteobacterial symbionts, environmental bacteria that are introduced to the system via the siphons and those that are located on the exterior of the gills or from other tissues contribute to the shipworm microbial community. It is worth to mention that genes are classified by similarity to other known genes in other organisms but may actually reside in a different microorganism than that identified by MG-RAST.

Analysis of the metagenome of *B. setacea* using MG-RAST shows that it is composed mainly of bacterial sequences with minor contamination from the host genes. Gammeproteobacteria represented the major bacterial group in the shipworm gill community as expected. Genes that were similar to those from non-proteobacterial phyla are mainly, besides bacteroidetes and firmicutes, actinobacteria and cyanobacteria. These two phyla are known to produce a large array of natural products suggesting a wide range of interaction among the microbial population within the shipworm.

Although one might need to perform PCR or construct genomic libraries to confirm that the genes are present within one cluster as predicted, the bioinformatic analysis of shipworm symbionts' genomes reveals that they represent a rich reservoir for secondary metabolites with the presence of the characteristic large size of ORFs encoding PKSs and NRPSs. The *ectA/B/C* genes responsible for the biosynthesis of ectoine are also present in the four strains and the shipworm metagenome. The gene *ectA* is a 2,4-diaminobutyric acid acetyltransferase, e*ctB* is a 2,4-diaminobutyric acid transaminase, while *ectC* is an L-

ectoine synthase. This compound acts as an osmoprotectant, similar to the glycine-betaine system in other organisms (Yancey et al., 1982; Imhoff, 1986). Ectoine accumulates in the cytoplasm of cells in marine environments to account for the osmotic differences within the surroundings (Jebbar et al., 1992). This suggests that supplementing the growth cultures with ectoine might be a successful strategy for the cultivation and reviving of other shipworm isolates. Ribosomal peptide genes involved in synthesis of bacteriocins are detected in one symbiont; BS08 in addition to the shipworm metagenome. Bacteriocins are proteinaceous toxins, usually with narrow spectrum of activity, produced by microorganisms to antagonize other closely related bacteria (Riley and Wertz, 2002). This prevents one organism from dominating the community and is expected in such a complex shipworm gill environment where the majority of the microbial community belongs to gammaproteobacteria.

Genes involved in the biosynthesis of terpenoid-derived hopanoids were detected in *T. turnerae* T7901. The key enzyme in the production of these compounds is the squalene cyclase that causes cyclization of the 30-carbon unit squalene to the pentacyclic hopene nucleus of the hopanoids. The role of hopanoids in prokaryotes is thought to resemble that of sterols in eukaryotes in increasing the membrane strength. Deletion in the squalene cyclase enzyme in other bacteria showed loss of hopanoid production leading to increased sensitivity to low pH and growth inhibitors (Schmerk et al., 2011). In addition, hopanoids have been shown to be essential in nitrogen fixing organisms, a key feature of *T. turnerae*, where their role is to decrease oxygen diffusion across the membrane of nitrogenase-containing vesicles (Berry et al., 1993).

Three out of the four strains: *T. turnerae* T7901, BS08 and BS12 showed potential for siderophore production. In fact, this was confirmed experimentally in *T. turnerae* (Han, 2011). Siderophores are iron chelators produced by microorganisms and secreted into the surrounding medium. They bind iron and provide it to the microbial producer in a bioavailable form. The presence of genes encoding potential siderophores organisms is not unusual in marine organisms. However, the catecholate-type siderophores found in both strains are uncommon in marine organisms. Although the cluster architecture of the siderophore in BS08 is similar to that of *T. turnerae* T7901, the adenylation domains are predicted to load different amino acids.

The genes specifying the two-component system PhoB-PhoR (Martín, 2004) are detected in *T. turnerae* T7901, BS02, and BS08 but are absent from BS12. This suggests that growing these strains under phosphate limiting conditions could lead to overexpression of secondary metabolites. The gene clusters in the symbionts' genomes show interesting features. One is the presence of an N-terminus thioesterase in region 4 of BS08. This is unusual as most TEs are located in the C-terminus of the region. This is similar to the TE in the biosynthetic gene cluster of the *Streptomyces* tripeptide, phosphinothricin suggesting that the C-terminus of the third ORF is arranged in a special manner such that it is close to the N-terminus of the first ORF (Eys et al., 2008). Another feature is present in the similar regions 1 and 2 in *T. turnerae* and BS08. A domain that functions as an ABC transporter is integrated in each of these NRPS ORFs instead of a standalone ORF. Since transporters are associated with the membrane, this might suggest that the biosynthetic enzymes are membrane associated and the compound is exported out of the cell once synthesized.

The assembly of large gene clusters by the assembler in the *B. setacea* metagenome suggests that they belong to a microbe that is predominant in the shipworm microbial community because of the fact that greater sequencing depth of an abundant organism might lead to more overlap and longer contigs. Nevertheless, large contigs could also be the result of misassembly of conserved regions in non-contiguous sequences due to the repetitive modular nature of these genes.

Although the microbial community analysis of the *B. setacea* gill did not show any evidence for the presence of the species *T. turnerae* T7901 (Dr. Daniel Distel, Ocean Genome Legacy, personal communication), it appears that several regions from the three *B. setacea* symbionts BS02, BS08, and BS12 are similar to that of the *B. gouldi* symbiont *T. turnerae* T7901 and predicted to produce similar compounds (Figure 2-15). The first region in BS02 for example, seems to be similar to that of region 8 in *T. turnerae* T7901 (Table 2-10). In fact, several ORFs upstream and downstream of the NRPS in this cluster are highly similar to corresponding genes of region 8 in *T. turnerae* T7901, including the glycosyl transferases, involved in the introduction of sugar units to the aglycone and the formation of glycosides. It is worth mentioning that the glycosyl transferases were not included as part of region 8 in *T. turnerae* T7901 when reported previously (Yang et al., 2009). It is possible that these genes are part of the cluster in both organisms making the cluster larger than originally thought.

Regions 1 and 2 in *T. turnerae* and BS08 respectively are highly similar (Table 2-10). Analysis of each ORF within the region shows homology to the corresponding gene in the other region, suggesting that both strains are capable of producing compounds that might be related to lysobactin or iturin, especially in the peptide portion of the molecule.

Lysobactin showed strong antibacterial activity against aerobic and anaerobic Gram positive organisms (Bonner et al., 1988), while iturin possesses antifungal properties (Tsuge et al., 2001). In addition, the homology of regions 1 and 5 in BS12 and *T. turnerae* T7901 (Table 2-11) respectively suggest that both strains are capable of producing spirotetronate compounds as tetronomycin (Demydchuk et al., 2008). These polyketides are polyether ionophoric compounds that also possess antibacterial activity (Keller-Juslén et al., 1982). Thirteen genes including the five PKS ORFs and other modifying genes and an exporter all show high similarity to corresponding genes in region 5 of *T. turnerae* T7901 (Table 2-10). The only difference is the orientation of the genes. Finally, regions 6 in both *T. turnerae* and BS12 show similarity in the largest ORF in each region (Table 2-11: Figure 2-15).

**Table 2-10. Similar genes in BS02 and BS08 to *T. turnerae* T7901.**
*, Corresponding gene in *T. turnerae* T7901 was not highly similar.

| Locus Tag | Predicted Function | Similar gene in *T. turnerae* T7901 | Identity/ similarity % |
|---|---|---|---|
| **BS02DRAFT_00011880** | thioesterase | Region 8_TERTU_4298 | 39/60 |
| **BS02DRAFT_00011930** | prenyltransferase | Region 8_TERTU_4288 | 61/79 |
| **BS02DRAFT_00011940** | oxidoreductase | Region 8_TERTU_4287 | 62/77 |
| **BS02DRAFT_00011950** | dehydrogenase | Region 8_TERTU_4286 | 59/73 |
| **BS02DRAFT_00011960** | hypothetical | Region 8_TERTU_4308 | 49/70 |
| **BS02DRAFT_00012000** | hypothetical | Region 8_TERTU_4305 | 45/72 |
| **BS02DRAFT_00012010** | oxidoreductase | Region 8_TERTU_4306 | 68/79 |
| **BS02DRAFT_00012030*** | NRPS | Region 8_TERTU_4295 | 29/48 |
| **BS08DRAFT_00007350** | PKS | Region 1_TERTU_1990 | 54/70 |
| **BS08DRAFT_00007340** | dehydratase | Region 1_TERTU_1991 | 67/82 |
| **BS08DRAFT_00007330** | aminotransferase | Region 1_TERTU_1992 | 83/93 |
| **BS08DRAFT_00007320** | hypothetical | Region 1_TERTU_1993 | 76/86 |
| **BS08DRAFT_00007310** | AT | Region 1_TERTU_1994 | 65/81 |
| **BS08DRAFT_00007300** | exporter | Region 1_TERTU_1995 | 64/80 |
| **BS08DRAFT_00007290** | hypothetical | Region 1_TERTU_1996 | 56/77 |
| **BS08DRAFT_00007280** | NRPS | Region 1_TERTU_1997 | 61/75 |
| **BS08DRAFT_00007270** | ABC transporter | Region 1_TERTU_1998 | 71/86 |
| **BS08DRAFT_00007260** | hypothetical | Region 1_TERTU_1999 | 60/73 |
| **BS08DRAFT_00007250** | aminotransferase | Region 1_TERTU_2000 | 76/88 |
| **BS08DRAFT_00007240** | TonB dep. receptor | Region 1_TERTU_2002 | 69/84 |

**Table 2-11. Similar genes in BS12 to *T. turnerae* T7901.**
*, Corresponding gene in *T. turnerae* T7901 was not highly similar.

| Locus Tag | Predicted Function | Similar gene in *T. turnerae* T7901 | Identity/ similarity % |
|---|---|---|---|
| **or0034** | dehydrogenase-ACP | Region 5_TERTU_2381 | 51/68 |
| **or0035** | PKS | Region 5_TERTU_2380 | 44/60 |
| **or0036** | PKS | Region 5_TERTU_2379 | 45/62 |
| **or0037** | PKS | Region 5_TERTU_2378 | 45/61 |
| **or0038** | PKS | Region 5_TERTU_2377 | 47/63 |
| **or0039** | PKS | Region 5_TERTU_2376 | 45/63 |
| **or0040** | Oxoacyl-ACP | Region 5_TERTU_2375 | 68/82 |
| **or0041** | FkbH-like protein | Region 5_TERTU_2374 | 42/66 |
| **or0042** | ACP | Region 5_TERTU_2373 | 37/62 |
| **or0043** | Dehydrogenases-AT | Region 5_TERTU_2372 | 52/71 |
| **or0044** | hydroxylase | Region 5_TERTU_2371 | 52/71 |
| **or0045** | hypothetical | Region 5_TERTU_2370 | 51/66 |
| **or0046** | TE | Region 5_TERTU_2369 | 49/69 |
| **or0047** | polyketides cyclase | Region 5_TERTU_2368 | 49/64 |
| **or0048** | hypothetical | Region 5_TERTU_2367 | 40/58 |
| **or2168*** | NRPS | Region 6_TERTU_2858 | 36/54 |

The region similarity of secondary metabolites between *T. turnerae* T7901 and the *B. setacea* symbionts suggests the possibility of previous horizontal gene transfer event between a once present *T. turnerae* T7901 and other symbionts in *B. setacea* (Table 2-10; Table 2-11). Another possibility that *T. turnerae* might have been introduced into the shipworm-microorganisms symbiosis later and acquired the secondary metabolite clusters. Possibly, these microorganisms might have suffered genome reduction and might have been excluded from the symbiosis. Another possibility is that since these strains are closely related shipworm symbionts, they could be a linear descendant from a common ancestor where these genes were conserved due to their essential functions. Another scenario that would explain this resemblance, although less likely, is convergent evolution. These similarities could shed some light on similar environmental conditions of these strains.

Based on the genomic information, the shipworm symbionts are expected to present a rich reservoir for bioactive drugs. Secondary metabolites produced by shipworm symbionts, different from those produced by soil bacteria, most probably have evolved to perform certain functions in a way not to harm the host. This suggests that the symbionts' metabolites might serve as compounds with interesting pharmacological activities with less toxicity. These potential compounds could be targeted using different approaches based on the bioinformatic data presented. The similarity in some of the regions could be used as an advantage in their isolation. Compounds of regions 1 and 2 in *T. turnerae* T7901 and BS08, respectively could be isolated by growing both strains in different culture conditions and comparing their LC/PDA/MS profiles. The same approach could be used for *T. turnerae* and BS12 for the spirotetronate compounds. In both cases, the growth conditions need to be optimized, since some regions are silent under specific growth conditions. Regions 1 and 2 in BS08 and BS12 respectively, predicted to produce siderophores, could be detected based on Chrome azurol S assay activity guided isolation after growing in iron limited media (Schwyn and Neilands, 1987). The chemical and phylogeny approach could be suitable for products specified by region 3 of BS08 in addition to those of regions 2 and 6 in BS12 using the characteristic absorbance for aromatic compounds as a guide. Feeding labeled substrates and amino acids could also be used to isolate a specific compound. For other regions where straight-forward predictions are not available or when the predicted compounds are not produced by the previous approaches, either mutational or regulatory approaches could be used. The drawback of this might be the difficulty in genetic manipulation of these strains, in spite of two previous successful attempts in *T. turnerae* (Han, 2011; chapter 3).

Actinobacteria have always been a rich source of interesting compounds especially antibiotics and hence have attracted major attention for their chemical and genomic analysis. The current data produced from four bacteria that belong to the same class of gammaproteobacteria suggest that the latter deserve similar attention. The rich reservoirs of genes encoding secondary metabolites and especially potential PKS and NRPS genes suggest possible sources for diverse natural products that possess therapeutic activities.

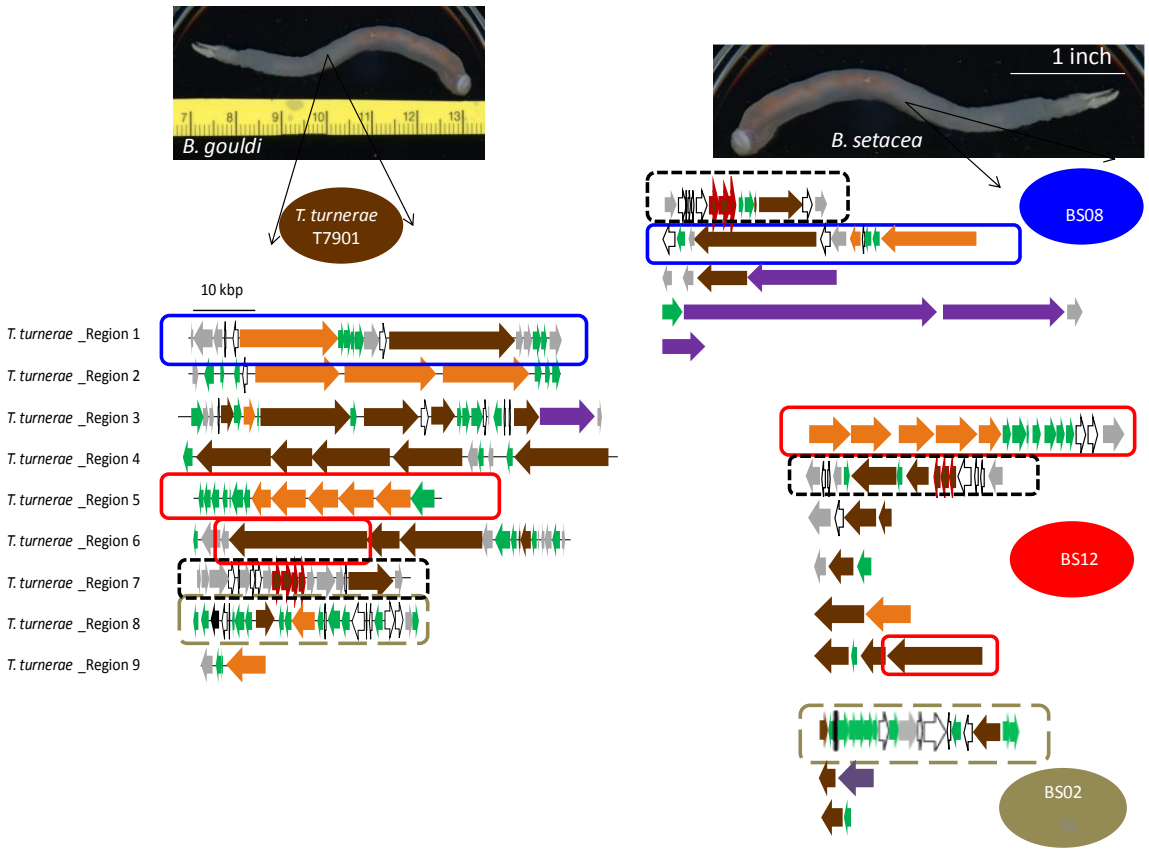



**Figure 2-15. Common biosynthetic clusters between *T. turnerae* T7901 and each of BS02, BS08, and BS12.**

Colored rectangles indicate the similar clusters in BS08 (blue), BS12 (red), and BS02 (grey). Black dotted rectangles indicate the presence of clusters encoding the biosynthesis of catecholate siderophores. Shipworm photos provided by Dan Distel.

# 3. Isolation, Biosynthesis, and Bioactivities of Tartrolons from *Teredinibacter turnerae*; a Shipworm Endosymbiont

## 3.1. Introduction

Shipworms are marine bivalve mollusks that belong to the family Teredinidae. They feed on wood and can also filter feed; some species have been shown to be capable of using wood as a sole source for food. They burrow into wood using their valves as drills to convert the wood into small particles with a high surface area. These particles are further digested into simpler units using cellulose degrading enzymes. There are excellent reviews on shipworm anatomy and biology (Turner, 1966; Distel, 2003).

Shipworms have been shown to contain a microbial community of closely related gammaproteobacterial intracellular symbionts in the gill that are proposed to provide benefits to the host. As many as four genetically distinct symbionts may exist in the gill of an individual shipworm (Distel et al., 2002). Cultivated strains of one of these symbionts, *Teredinibacter turnerae,* secrete cellulose digesting enzymes and fix nitrogen (Waterbury et al., 1983), functions that have been hypothesized in the case of cellulose digestion and shown in the case of nitrogen fixation, to assist the host in consuming wood (Distel et al., 2002; Yang et al., 2009).

More than 58 strains of *T. turnerae* have been isolated from different shipworm hosts collected from distantly separated habitats. The genome of one of these strains; T7901, was sequenced (Yang et al., 2009) and revealed genes that encode enzymes specific for cellulose degradation and nitrogen fixation as expected. The genome sequencing also revealed at least nine regions that encode megaenzymes for the biosynthesis of polyketides and nonribosomal peptides. This rich source of metabolites could play a role

in the symbiotic relationship by acting as signaling molecules to defend the host against predation, or to deter microbial competitors.

Bioactive metabolite symbiosis (Hildebrand et al., 2004) is a term used to describe a symbiotic relationship between organisms based on chemical compounds. Usually one of the organisms produces one or more of secondary metabolites that has the potential of protecting the host or the rest of the community from environmental threats. Several examples in the field have been reported recently. First, a wide spectrum of nine antibiotics produced by a group of symbiotic actinobacteria seems to protect the offspring of a wasp from fungal and bacterial pathogens (Kroiss et al., 2010). Another example comes from the leaf cutting ants that grow a specific type of fungus and use it as the major food source. To protect its food against other competitive microorganisms and fungi, the ants harbor a group of actinobacterial symbionts that produce antimicrobial and antifungal compounds (Schoenian et al., 2011). Finally, the gene cluster for bryostatin 1, originally isolated from the marine bryozoan, *Bugula neritina*, was cloned and sequenced (Sudek et al., 2007; Hildebrand et al., 2004). The bryozoan was found to harbor a gammaproteobacterial endosymbiont, *Candidatus* Endobugula sertula, that was proposed to be the true producer for this compound (Davidson et al., 2001). This metabolite was found to be deposited on the larvae to act as feeding deterrent to protect the larvae (Lopanik et al., 2004). Many of these symbiont metabolites were found to possess therapeutic activities, highlighting the importance of marine natural products.

Boron is a semi-metal that is known to exist in the form of borate or orthoborate in nature. The concentration of boron in the ocean is estimated to be 0.4 mM (Uppström, 1974; Carrano et al., 2009). Boron is known to play important roles for living organisms

including animals (Nielsen, 2000), plants (Marschner, 2011), yeast (Bennett et al., 1999), and bacteria (Bolaños et al., 2002). Boron-containing secondary metabolites have been isolated and identified from soil and marine bacteria from different bacterial groups and were found to possess interesting pharmacological activities. In addition, synthetic and naturally present boron containing compounds were found to participate in quorum sensing (Dembitsky et al., 2011).

The shipworm symbiont *T. turnerae* was isolated and cultivated about three decades ago; yet no reports have characterized secondary metabolites from this bacterium. We hypothesized that this microorganism may play an important role, through its secondary metabolites, in interspecies interactions with other symbionts, other microbes and with the host. This study reports the first secondary metabolite to be identified from *T. turnerae* or any shipworm symbiont, describes the biosynthetic gene cluster linked to it, and presents evidence that this metabolite is produced in the symbiotic state.

## 3.2. Materials and Methods

### 3.2.1. Strains and culturing

Eleven of the thirteen strains of *T. turnerae* (Table 3-1) used in this study were previously described (Distel et al., 2002) and provided by the Waterbury lab at the Woods Hole Oceanographic Institute, Massachusetts, USA. The strain CS30 was also isolated and characterized previously (Trindade-Silva et al., 2009).

**Table 3-1. Strains and plasmids used in this study.**
The strain numbers of *T. turnerae* starts with a T designates an isolate from Teredinidae followed by a number, the first 2 digits refers to the year the strain was isolated while the 2 other digits refer to the number of isolate in that year.

| Strain or plasmid | Description | Reference |
|---|---|---|
| *T. turnerae* | | |
| T7901 | isolated from *Bankia gouldi* and wild type for AH02 mutant | (Distel et al., 2002) |
| AH02 | T7901 gene disrupted in TERTU_2202 | This study, constructed by Andrew Han, OHSU |
| T7902 | isolated from *Lyrodus pedicellatus* | (Distel et al., 2002) |
| T7903 | isolated from *Teredo navalis* | (Distel et al., 2002) |
| T8201 | isolated from *Psiloteredo healdi* | (Distel et al., 2002) |
| T8202 | isolated from *Teredo furcifera* | (Distel et al., 2002) |
| T8203 | isolated from *Teredo bartschi* | (Distel et al., 2002) |
| T8304 | isolated from *Nototeredo edax* | (Distel et al., 2002) |
| T8401 | isolated from *Bankia rochi* | (Distel et al., 2002) |
| T8402 | isolated from *Teredora malleolus* | (Distel et al., 2002) |
| T8506 | isolated from *Teredo triangularis* | (Distel et al., 2002) |

| | | |
|---|---|---|
| T8602 | isolated from *Dicyathifer manni* | (Distel et al., 2002) |
| CS30 | isolated from *Neoteredo reynei* | (Trindade-Silva et al., 2009) |
| *Escherichia coli* strains | | |
| TOP10 | F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L ($Str^R$) *end*A1 *nup*G | Invitrogen |
| S17-1 λpir | *thi pro hsdR hsdM*$^+$ *recA* RP4-2-Tc::Mu-Km::Tn7λpir | (Simon et al., 1983) |
| Plasmids | | |
| pCR2.1 | T-vector, $Km^r$, $Amp^r$ | Invitrogen |
| pDM4 | Suicide plasmid, *sacB* gene, R6K origin, $Cm^r$ | (Milton et al., 1996) |
| pDMrg2KS | portion of first KS domain from TERTU_2202 gene cloned into pDM4 | This study |

### 3.2.2. Media recipes, growth conditions, and growth monitoring

Optimization of preferred media for the production of tartrolons was monitored by disc diffusion assay against *B. subtilis* and measuring the zone of inhibition produced by the organic extract of the *T. turnerae* T7901 culture. We slightly optimized a sucrose and basal medium (SBM) (Waterbury et al., 1983) previously used. We tested the use of different carbon sources including sucrose, Sigmacell cellulose Type 101 (Sigma Aldrich) and glucose. Since phosphate has been shown to regulate the production of antibiotics in some microorganisms (Martín, 2004), a range of inorganic phosphate concentrations was also tested. Shaking was varied from 100 to 225 rpm to test for the highest activity.

Each culture was extracted with ethyl acetate and dried under vacuum. The crude extract was dissolved in a mixture of 50% chloroform in methanol at a concentration of 10 mg/ml and a volume of 10 µl of this solution was applied to a disc. The disc was applied on a *B. subtilis* lawn and incubated at 37$^{o}$C overnight. Activity was measured by the diameter of the inhibition zone around the disc with media extract. Negative controls including solvents and uninoculated media under the same conditions were used.

*T. turnerae* T7901 was streaked on agar plates and grown at 30$^{o}$C for three days. Colonies were then used to inoculate a seed culture of 50 ml and was grown for 48 hours. This culture was used to inoculate a large scale culture at 0.1%. Six-2.8 L vessels were used; each contained 2 L of culture. The culture was incubated at 30$^{o}$C for four days at 125 rpm. This process was repeated for a second time making a total of 24 L of growth culture. The recipe of the medium that gave highest activity was follows: sucrose (5 gm/L), NaCl (19.8 gm/L), $NH_4Cl$ (267.5 mg/L), $MgCl_2.6H_2O$ (8.95 gm/L), $Na_2SO_4$ (3.31

gm/L), $CaCl_2.2H_2O$ (1.25 gm/L), $NaHCO_3$ (0.162 gm/L), $Na_2CO_3$ (10 mg/L), KCl (0.552 gm/L), KBr (81 mg/L), $H_3BO_3$ (21.5 mg/L), $SrCl_2.6H_2O$ (19.8 mg/L), $KH_2PO_4$ (3.82 mg/L), NaF (2.48 mg/L), $Na_2MoO_4.2H_2O$ (2.5 mg/L), $MnCl_2.4H_2O$ (1.8 mg/L), $ZnSO_4.7H_2O$ (0.22 mg/L), $CuSO_4.5H_2O$ (0.079 mg/L), $Co(NO_3)_2.6H_2O$ (0.049 mg/L), Fe-EDTA complex (4.15 mg/L), HEPES (4.76 gm/L) to pH = 8.0.

### 3.2.3. Isolation and identification of tartrolons

The culture was centrifuged to separate the supernatant from the cell pellet. The cell pellet was sonicated in 50% chloroform in methanol three times while the supernatant was extracted with 5% HP-20 w/v for 2-4 hrs. The resin was collected and washed with 25% isopropanol in water. The adsorbed molecules were collected by washing with 100% methanol and 100% acetone and dried under vacuum. Both the cell pellet and supernatant extracts were combined. The extract was partitioned between ethyl acetate and water three times. The non-polar fraction was then filtered over anhydrous sodium sulfate. This was fractionated on a silica column using hexane and ethyl acetate in a gradient elution. The active fraction was then subjected to purification by normal phase HPLC using a normal phase silica column (Supelco) 25 × 10 mm and hexane and isopropanol as mobile phases.

### 3.2.4. General

IR spectra were recorded on a Perkin-Elmer System 2000 FTIR spectrometer. NMR data were collected using a Bruker 600 ($^1H$ 600 MHz, $^{13}C$ 150 MHz) NMR spectrometer equipped with a microprobe with a z-axis gradient and utilized residual solvent signals for referencing. High-resolution mass spectra (HR-MS) were obtained using ThermoElectron LTQ-Orbitrap high resolution mass spectrometer. AntiMarin database

(http://www.chem.canterbury.ac.nz/marinlit/marinlit.shtml) was used as the natural product database for dereplication.

### 3.2.5. Gene disruption

Gene disruption mutant AH02 was constructed to disrupt region 2 in *T. turnerae* T7901 (Figure 3-1). For construction of the plasmids used for gene disruption, a region within the first KS domain of the PKS gene TERTU_2202 was amplified with specific PCR primers, Rg2KS693F and Rg2KS1218R (Table 3-2) and the high-fidelity Phusion DNA polymerase (Finnzymes). The addition of 3' A-overhangs for TA cloning was carried out by DyNAzyme II DNA polymerase (Finnzymes) following the manufacturer's protocol. The amplicon was purified using the QIAquick PCR Purification Kit (Qiagen) and then cloned into pCR2.1 (Invitrogen). Plasmid DNA was isolated with QIAprep Spin Miniprep Kit (Qiagen) and double digested with the restriction enzymes XhoI and SacI. The resulting fragment was gel purified with the GENECLEAN II Kit (MP Biomedicals) and then ligated with the Quick Ligation Kit (NEB) into suicide vector pDM4, which had been double digested with the same restriction enzymes, resulting in plasmid pDMrg2KS. The plasmid pDMrg2KS was transformed into *E. coli* S17-1 λpir. This plasmid was then conjugated into *T. turnerae* T7901, and plasmid cointegrates were selected on SBM plates with Sigmacell cellulose Type 101 0.5% final concentration (Sigma-Aldrich) as the sole carbon source and 10 μg/ml chlorampenicol. The location of integration by pDMtart into the chromosome of *T. turnerae* T7901 was confirmed by PCR and DNA sequencing.

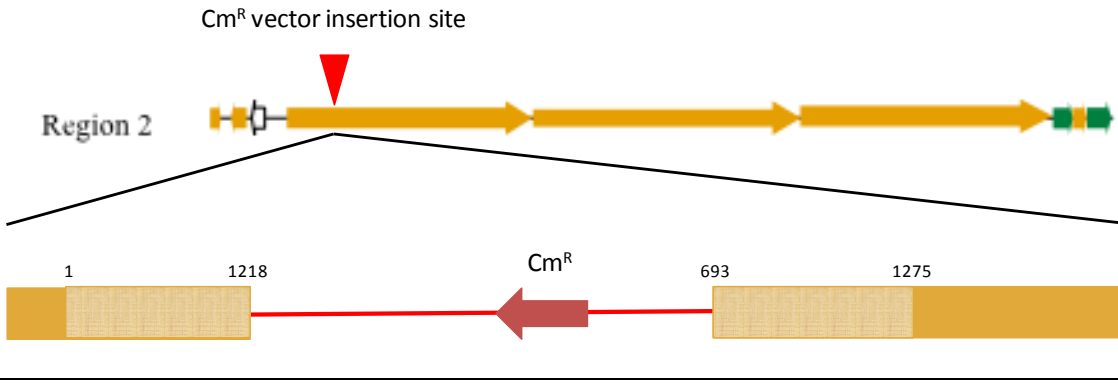



**Figure 3-1. Mutation scheme of region 2.**

### 3.2.6. Mass spectra analysis of wild type and mutant AH02

Both, *T. turnerae* T7901 wild type and AH02 mutant were grown in glucose 0.5% final concentration and basal medium as described above. Sucrose used previously in cultures to isolate the compound was substituted with glucose in this experiment in both the wild type and the mutant due to the presence of sucrose sensitivity gene *sacB* in the vector of AH02. Both strains were extracted similarly. Cells were extracted with 50% chloroform in methanol while the supernatant was mixed with 5% HP20 w/v (Sigma-Aldrich) and washed with methanol and acetone. The combined dried extract was partitioned between ethyl acetate and water and the organic fraction was dried. The extracts were dissolved in methanol at equal concentrations and subjected to HR-MS and MS/MS.

### 3.2.7. Prevalence of tartrolons in different *T. turnerae* strains

The media used is described earlier in the methods section. Each strain was streaked on agar plates and then inoculated in 50 ml liquid cultures. The same growth conditions were used to grow all twelve strains in order to test for the prevalence of tartrolon E. For MS analysis, cells were centrifuged at 10,000 ×g for 20 min, collected and sonicated in 50% chloroform in methanol. The combined fractions were filtered and concentrated under reduced pressure. All dry extracts were dissolved in the same volume of methanol

and analyzed by HR-MS and MS/MS. Methanol (25 µl) as a negative control was injected before each sample. Background subtraction from the previous methanol sample was used to process the data. The region between *m/z* 800-930 was analyzed. In addition, MS/MS was used to fragment the characteristic peaks for tartrolon E. For PCR analysis; DNA was isolated from twelve *T. turnerae* strains using phenol and chloroform as reported previously (Sambrook and Russell, 2001). PCR reactions were performed with 10-100 ng of *T. turnerae* genomic DNA, using Platinum Taq DNA polymerase (Invitrogen). PCR was used for checking the prevalence using primers targeting each of the three biosynthetic ORFs *trtDEF* (Table 3-2).

**Table 3-2. Sequence of primers used in this study and genes targeted.**

| Primer Number | Primer Name | Sequence | Target ORF |
|---|---|---|---|
| **1** | Rg2KS693F | 5'-AAT GGG GGA GGG CGT CGG GTT-3' | *trtD* |
| **2** | Rg2KS1218R | 5'-ACC CGC ACG TTT AGG CTC CG-3' | *trtD* |
| **3** | ORF 03363-F | 5’-GCG GTA AAT TCC GGC TTT CTC GTT-3’ | *trtD* |
| **4** | ORF 03363-fkhb-PR | 5’- CAA GTA AGC TGC TTT GCC GAG A-3’ | *trtD* |
| **5** | AH-ORF03363-8396-F | 5’- TAA TCA ACA GCC AAC GCA CGC TTC -3’ | *trtD* |
| **6** | AH-ORF03363-8963-R | 5’- TGA GAG ACA GGT TGC TGC CCA ATA -3’ | *trtD* |
| **7** | AS_orf03363-fkhb-PF | 5’- CCA GTA GAA ATA CTG GAA GTG CAT CAG C-3’ | *trtD* |
| **8** | AS_orf03363-fkhb-PR | 5’- CAA GTA AGC TGC TTT GCC GAG A-3’ | *trtD* |
| **9** | AH-ORF03362-6012-F | 5’-TTT AGT TTG GTG CAA ACC GCC GAC-3’ | *trtE* |
| **10** | AH-ORF03362-6843-R | 5’-AAA TAC ACC GAG CGA GTG AGA GCA-3’ | *trtE* |
| **11** | AS_Orf03362-M1-ID-PF | 5’- AAG CAA CCT CGC TGA TGC TTT CAC-3’ | *trtE* |
| **12** | AS_Orf03362-M1-ID-PR | 5’- TTA CTC AGC AAT TCC GCC AAT GCG -3’ | *trtE* |
| **13** | AH-dPKSORF03362-2663-F | 5’- TGG AGC GAA CGC CTA CTT GAT TAT CG -3’ | *trtE* |
| **14** | AH-dPKSORF03362-2892-R | 5’- CAA TCG ATG GGC CAA ACG CTT TCT -3’ | *trtE* |
| **15** | AH-ORF03361-12757-F | 5’- ATG AGT CCA TGA GCG AGC CAG AAA-3’ | *trtF* |
| **16** | AH-ORF03361-13393-R | 5’- GTT CAG CGA GTT GGC GAC ACA TTT-3’ | *trtF* |
| **17** | AS_Orf3361-MT-PF | 5’- AGA TTG GTG CAG GTA CTG GGT-3’ | *trtF* |
| **18** | AS_Orf3361-MT-PR | 5’- AAT GTC ACA GTT TGG GAT TCC CGC-3’ | *trtF* |

### 3.2.8. Analysis of different shipworms for the presence of tartrolons

Different individuals of shipworms of the species *Bankia setacea* and *Lyrodus pedicellatus* were extracted on ice from the wood using appropriate tools and sterile seawater. Shipworms were washed with sterile seawater and frozen rapidly at -80 $^{o}$C. The samples were then lyophilized and sonicated in 50% chloroform in methanol three times using Branson digital sonifier. The combined organic fraction was concentrated under vacuum and partitioned between ethyl acetate and water. The ethyl acetate fraction was filtered over anhydrous sodium sulfate and dried under vacuum. This fraction was then dissolved in methanol before injecting into HR-MS. Proper controls containing solvents were used to detect any contamination from solvents or instruments.

### 3.2.9. Bioactivities of tartrolons

#### 3.2.9.1. Disc diffusion assays

About two to three colonies of each of the test organisms *B. subtilis*, *E. coli*, or *V. anguillarium*, were picked from the corresponding plate and grown overnight in liquid culture. 1 ml of each culture was transferred to 50 ml and left to grow until the turbidity reached an OD600 of 0.7-1. A volume of 20 µl was applied on each LB plate and spread. Fractions and pure compounds were dissolved in chloroform at concentrations of 20 mg/ml and 1 mg/ml, respectively. In addition, ten other shipworm symbiont isolates representing different clades were tested including BS02 on cellulose and basal medium agar plates (Waterbury et al., 1983). The phylogeny of these symbionts has been reported elsewhere (Betcher, 2011). A volume of 20 µl of each sample to be tested was applied on a disc and allowed to air dry on a sterile glass slide. Discs were then applied on the

streaked plates and incubated at 37$^{o}$C, 30$^{o}$C, or at room temperature depending on the test organism.

#### 3.2.9.2. Determination of minimum inhibitory concentration (MIC) of tartrolon E against *B. subtilis* and *E. coli*

The protocol used for the determination of MIC was previously described with slight modifications (Wiegand et al., 2008). Briefly, the experiment was performed in a 96 well plate using Mueller-Hinton media as recommended by Clinical and Laboratory Standards Institute (CLSI) for testing antimicrobials using serial dilutions of tartrolon E, in addition to chloramphenicol as a control. The plate was incubated overnight at 37$^{o}$C and then checked by a plate reader. The well with the minimal concentration of tartrolon E that did not show growth and whose reading was similar to the negative control was recorded as the MIC.

#### 3.2.9.3. Antimicrobial activity against pathogenic bacteria

The antimicrobial assays were performed on four pathogens; *P. aeruginosa* (ATCC 10145), *K. pneumoniae* (ATCC 13883), methicillin-sensitive *S. aureus* (ATCC 12600), and methicillin-resistant *S. aureus* (MRSA) (ATCC 43300). Tartrolon E was tested at 100 µg/ml and 10 µg/ml using the standard microdilution method (Andrews, 2002; Garcia, 2010; Brown et al., 2005). Resazurin, a blue non-fluorescent and non-toxic dye that is reduced to a pink and fluorescent resorufin within viable cells, was used as indicator of cell growth (Sarker et al., 2007). Fluorescence for each well was measured using the Synergy HT Multi-Mode Microplate Reader at 530/25 ex and 590/35 em. Minimum inhibitory concentrations (MIC) of compounds positive for antibacterial activity at high and low doses were determined from the two-fold serial dilution of compounds with

range between 0.08 and 10 µg/ml. All compounds were tested in triplicate and the positive controls used are as follow: nalidixic acid for *P.aeruginosa* and *K. pneumoniae*, oxacillin for methicillin sensitive *S. aureus*, and vancomycin for methicillin resistant *S. aureus*.

#### 3.2.9.4. Dorsal root ganglion assay

Dorsal root ganglion (DRG) cells from cervical and lumbar regions were obtained from C57B1 mice and used in an assay with pure compounds as previously described (Peraud et al., 2009; Lin et al., 2011). Briefly, DRG cells were suspended in medium with additives and loaded with a fluorescent dye used to measure intracellular calcium levels. After baseline measurements, the cells were treated with 25 mM KCl solution and then washed. After return to baseline, pure compounds were applied. The pure compounds and KCl were injected together. This solution was then later replaced with 25 mM KCl solution 5 min later. To differentiate pain-sensing and TRPV1-expressing neurons from other neuronal types, capsaicin was applied after return to baseline of the compounds. Finally, additional 25 mM KCl was applied to determine whether cells were still viable with normal action potentials.

#### 3.2.9.5. Cytotoxicity assay

Human breast adenocarcinoma cell line (MCF-7) was obtained from the American Type Culture Collection (Manassas, VA) and was grown as a monolayer in minimum essential medium (MEM) containing 10% fetal bovine serum and 1% of 1× of the antibiotic-antimycotic, penicillin-streptomycin in fungizone. Cell cultures were maintained at 37$^{o}$C in a humidified 5% $CO_2$ atmosphere. Cytotoxicity assay was done using a modified method used previously (Mosmann, 1983). Cells were seeded into 96-well plates at 2 ×

$10^4$ cells/ well and incubated overnight. Cells were then exposed to a range of concentrations of tartrolon E. After 72 hours of drug treatment, cell viability was determined by measuring the metabolic conversion of 3-(4,5-dimethyl-2-thiazoyl)-2,5-diphenyl-2H-tetrazolium bromide (MTT) into purple formazan crystals by viable cells. The absorbance of each well was read at 570 nm using the Synergy HT Multi-Mode Microplate Reader. Tartrolon E was tested in quadruplicate and solubilized in 100% DMSO with a final DMSO concentration of 1% in each well. Doxorubicin was used as positive control. The $IC_{50}$ value, or the concentration of compound causing 50% inhibition of cell growth, was obtained using Graph Pad Prism 5.

## 3.3. Results

### 3.3.1. Growth conditions

The modified version of the shipworm basal medium was found to cause the highest activity (see method). Sucrose was found to give better results than either cellulose or glucose as sole carbon sources. In addition, decreasing the concentration of phosphate to 25% of that reported previously (Waterbury et al., 1983) seemed to cause larger inhibition zone against *B. subtilis*. This medium showed the largest zone of inhibition and was used to scale up.

### 3.3.2. Isolation and identification of the antibacterial metabolites

The modified version of the shipworm basal medium with lower inorganic phosphate (see methods) was used to scale up for the isolation of the antibacterial compound. A combined liquid culture of 24 L was extracted. The crude extract was used for bioactivity guided isolation to identify the antibacterial compound. Different chromatographic techniques led to the isolation of **1.** Careful analysis of the mass spectrum of **1** revealed the presence of a characteristic isotopic pattern for the presence of boron (Figure 3-2). A search in natural products database, AntiMarin (updated 03/2011), did not retrieve any formula corresponding to the HR-MS for **1**. Thus, 1 and 2D NMR spectroscopy data (Table 3-3) were used to elucidate the structure by combining partial structures based on 2D NMR spectroscopy. Compound **1** was found to be related to tartrolons and thus we gave it the name tartrolon E. A slightly more polar fraction also showed antibacterial activity. Purification of this fraction led to the isolation of **2** with abundance of approximately 10% of **1**. HR-MS, $^{1}H$ and $^{13}C$ NMR spectra of **2**, corresponded to a compound previously reported from a *Streptomyces* species (Pérez et al., 2009) which is

the deboronated version of **1**. Indeed, the same authors reported the borate ester derivative as a minor contaminant by analyzing the set of peaks that co-appeared with tartrolon D but did not isolate it or name it. Thus no full published spectroscopic or biological data exist for tartrolon E. These two compounds are related to a group of macrodiolide dimers that have been shown to possess interesting biological activities (Table 3-4; Figure 3-3).

The fractions of *T. turnerae* also suggested the presence of other borate ester derivatives. Analysis of other polar fractions in the negative mode revealed peaks of boronated compounds of masses 813.42, 843.43, and 859.43 (Figure 3-4; Figure 3-5). This suggests the presence of other tartrolon derivatives. Since the boronated derivative tartrolon E had a mass of 827 (Figure 3-2) with a formula of $C_{44}H_{64}O_{14}B^-$, the compound with a mass of 813.42 suggests a reduction of C=O to $CH_2$ with a formula of $C_{44}H_{66}O_{13}B^-$ (Figure 3-4) or a loss of one of the methylated groups of tartrolon E with a formula of $C_{43}H_{62}O_{14}B^-$, while the compound with mass 843.43 suggests a formula of $C_{44}H_{64}O_{15}B^-$ with an addition of oxygen (-$CH_2$ to -CH-OH). The compound with mass 859.43 might be $C_{44}H_{64}O_{16}B^-$ with the conversion of two methylene groups to methanol (Figure 3-5). These compounds were present in nanogram quantities not sufficient to analyze using NMR spectra to confirm the structure.

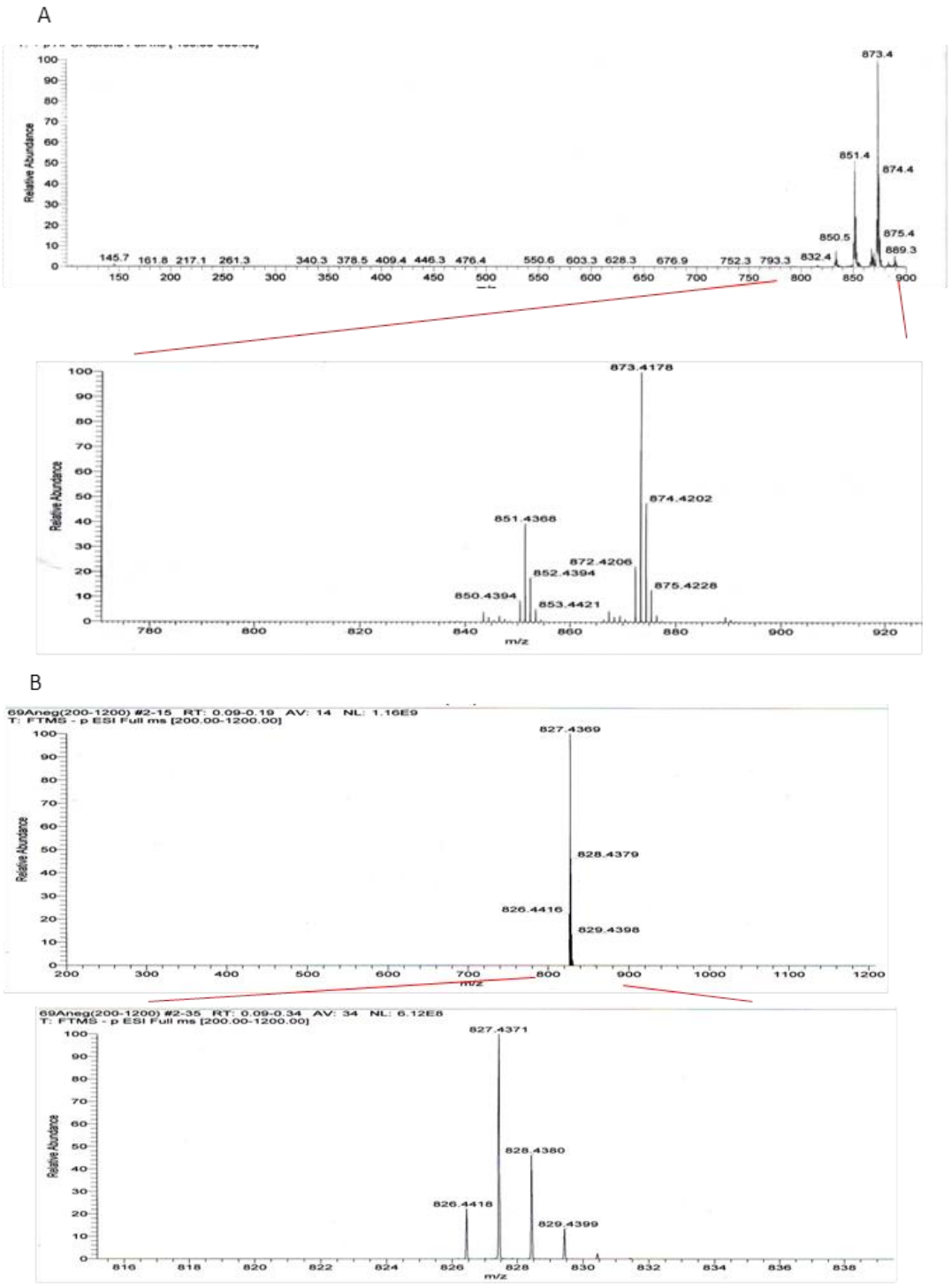
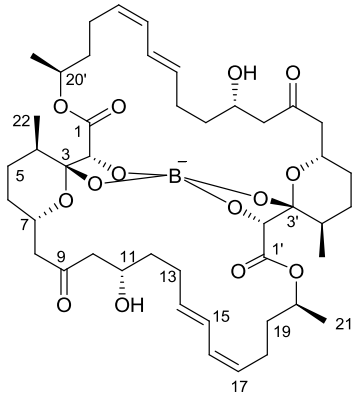



**Figure 3-2. High resolution mass spectrum of tartrolon E.**

HR-MS in A) positive mode and B) negative mode showing the characteristic natural abundance isotope pattern of boron.

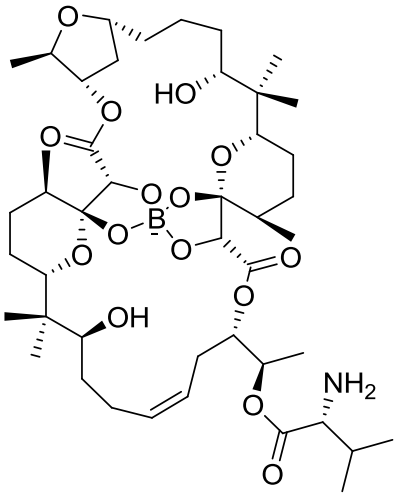
**Table 3-3. $^{1}H$, multiplicities, coupling constants, and $^{13}C$ NMR of tartrolon E.**

| Position | Tartrolon E (in $CDCl_3$) | |
|---|---|---|
| | $\delta_C$ | $\delta_H$, multiplicities, ($J$ in Hz) |
| 1,1' | 174.0, s | |
| 2,2' | 78.0, d | 4.45, s |
| 3,3' | 104.0, s | |
| 4,4' | 35.3, d | 1.81, m |
| 5,5' | 28.3, t | 1.68, m<br>1.54, m |
| 6,6' | 32.6, t | 1.60, m<br>1.16, m |
| 7,7' | 65.9, d | 4.45, m |
| 8,8' | 50.6, t | 2.52, dd, (13.5, 4.5)<br>2.15, dd, (13.0, 12.3) |
| 9,9' | 210.0, C | |
| 10,10' | 46.8, t | 3.40, dd (18.9, 10.5)<br>2.26, d (18.5) |
| 11,11' | 67.5, d | 4.03, t (10.0) |
| 12,12' | 31.6, t | 1.66, m<br>1.26, m |
| 13,13' | 28.2, t | 2.36, m<br>1.98, m |
| 14,14' | 134.8, d | 5.90, dt (15.1, 4.3) |
| 15,15' | 123.0, d | 5.98, dd (15.3 ,11.0) |
| 16,16' | 131.0, d | 6.07, t (11.0) |
| 17,17' | 127.8, d | 5.25, sextet (5.5) |
| 18,18' | 23.0, t | `2.42, dd<br>1.93, dd |
| 19,19' | 36.0, t | 1.83, m<br>1.33, t |
| 20,20' | 69.5, d | 4.65, m |
| 21,21' | 20.5,q | 1.20, d (6.0) |
| 22,22' | 16.8, q | 1.01, d(6.4) |

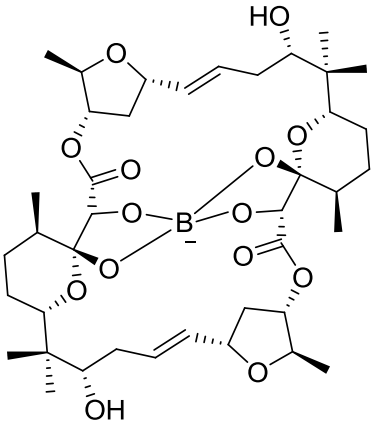


Structural characterization of tartrolons E **1** and D **2**.

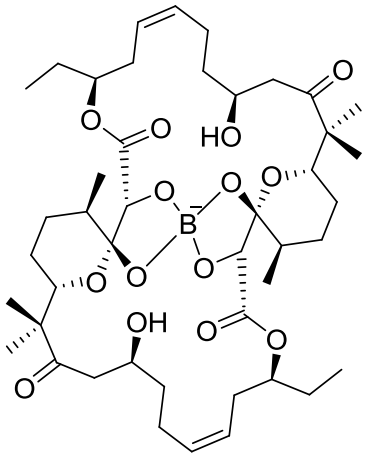
**1** White solid; UV (MeOH) $\lambda_{max}$ (log ε) 230 (4.0), IR (neat) $\nu_{max}$: 3500, 2950, 2923, 2830, 1753, 1730, 1430, 1380, 1100, 1005 $cm^{-1}$. HRESIMS m/z 873.4165 $[M + Na]^+$ (calcd for $C_{44}H_{66}O_{16}Na_2$, 873.4172) and 851.4475 $[M+H]^+$ (calcd for $C_{44}H_{67}O_{16}Na$). $^1H$ and $^{13}C$ NMR see Table 3-3. HRMS in addition to $^1H$ and $^{13}C$ NMR of **2** matched those reported previously (Pérez et al., 2009).



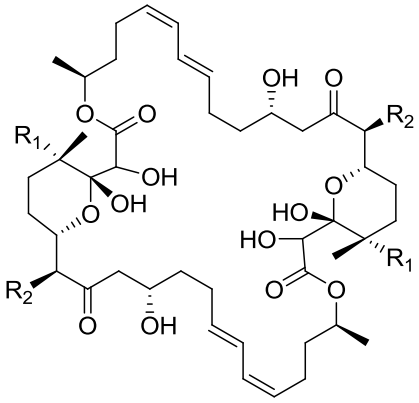
boromycin



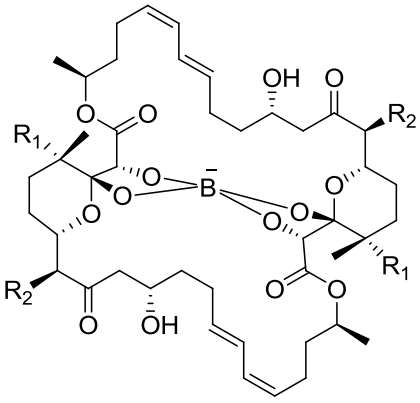
aplasmomycin



borophycin



tartrolon A: $R_1$=H, $R_2$=$CH_3$
tartrolon D: $R_1$= $R_2$=H



tartrolon B: $R_1$=H, $R_2$=$CH_3$
tartrolon C:$R_1$=OH, $R_2$=$CH_3$
tartrolon E: $R_1$=$R_2$=H

**Figure 3-3. Chemical structure of tartrolons and structurally related compounds.**

**Table 3-4. Structurally related compounds to tartrolons with their source and reported biological activities.**

| Compounds | Boron | Source | Biological Actiivity | Reference |
|---|---|---|---|---|
| boromycin | present | *Streptomyces antibioticus* (terrestrial) | -antibacterial<br>-anti-HIV | (Hütter et al., 1967; Kohno et al., 1996) |
| aplasmomycin | present | *Streptomyces griseus* (marine) | -antiplasmodium<br>-antibacterial | (Okami et al., 1976; Nakamura et al., 1977) |
| borophycin | present | Cyanobacteria *Nostoc linckia* *Nostoc spongiaeforme* var. tenue (marine) | -antibacterial<br>-cytotoxic | (Hemscheidt et al., 1994; Banker and Carmeli, 1998) |
| tartrolon A | absent | myxobacterium *Sorangium cellulosum* (terrestrial) | -antibacterial | (Schummer et al., 1994; Irschik et al., 1995) |
| tartrolon B | present | myxobacterium *Sorangium cellulosum* (terrestrial) | -antibacterial | (Schummer et al., 1994; Irschik et al., 1995) |
| tartrolon C | present | *Streptomyces* species (terrestrial) | -insecticidal activity<br>-inhibition of HIF-1 | (Lewer et al., 2003) |
| tartrolon D | absent | *Streptomyces* species (marine) | -cytotoxic<br>-antibacterial | (Pérez et al., 2009)<br>This study |
| tartrolon E | present | *T. turnerae* (marine) | -antibacterial<br>-cytotoxic | This study |

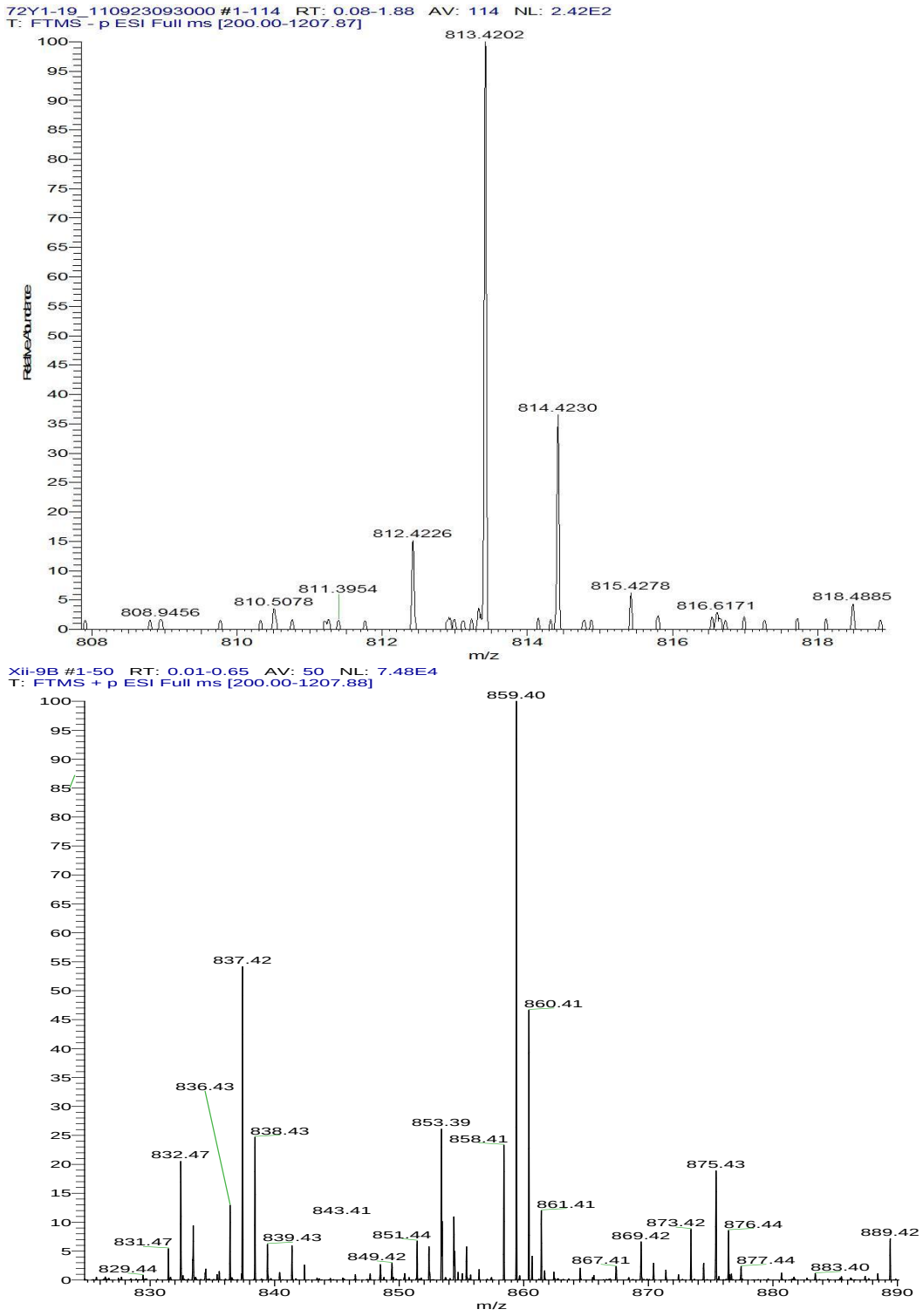



**Figure 3-4. Mass spectra of a boronated derivative.**
MS in negative (top) and positive (bottom) modes.

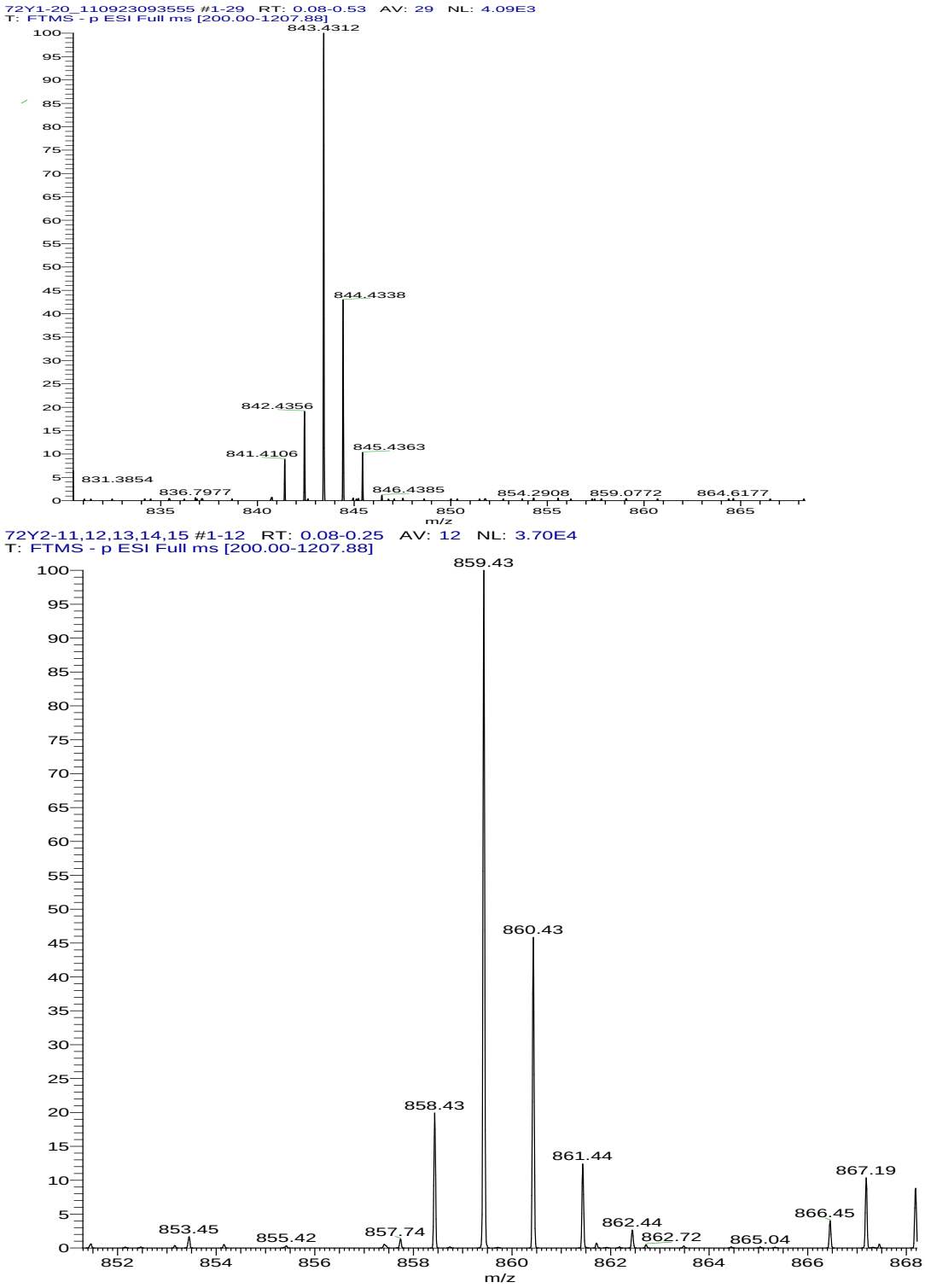



**Figure 3-5. Mass spectra of two boronated derivatives in negative modes.**

### 3.3.3. Gene disruption

Nine secondary metabolite regions were found during the analysis of the genome of *T. turnerae* T7901 that encode the biosynthesis of a siderophore in addition to eight other secondary metabolites (Yang et al., 2009). The chemical structure of tartrolons suggested a polyketide synthase (PKS) biosynthesis mechanism. Thorough analysis of the domains and modules of the three regions that encode PKS (regions 2, 5, and 9) and comparing the predicted intermediate formed with the tartrolon chemical structure strongly suggested that region 2 is responsible for the biosynthesis of tartrolons. To confirm that region 2 is responsible for this biosynthesis, a mutant AH02 was constructed by integrating a plasmid containing a chloramphenicol resistance cassette into the KS1 domain of *trtD* by a single crossover recombination. HR-MS and tandem MS/MS of the ethyl acetate fractions of both the wild type and the mutant were compared and showed absence of tartrolons in the mutant (Figure 3-6).

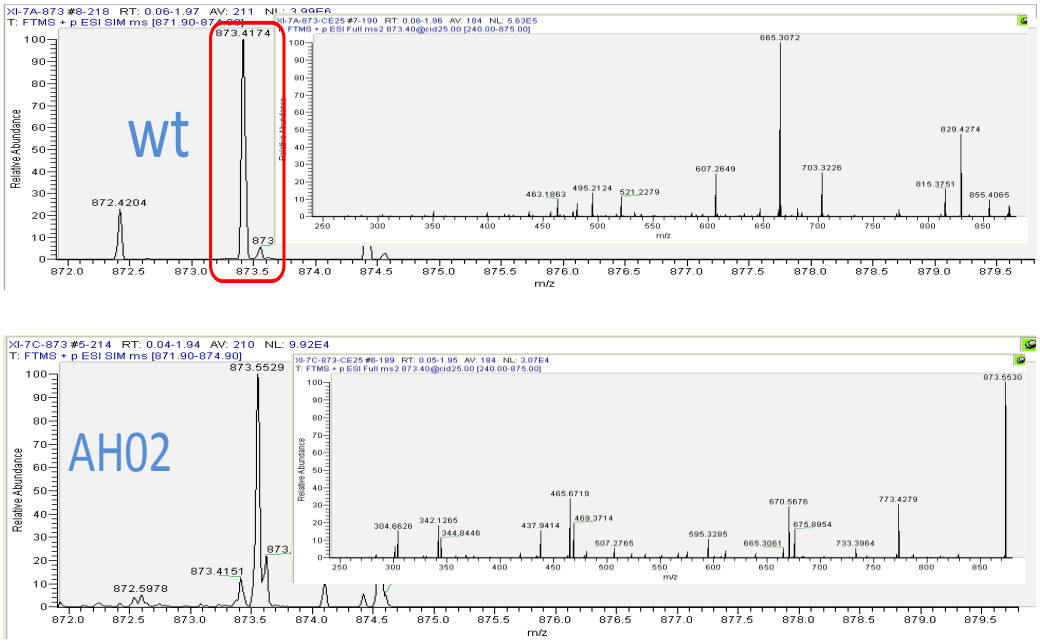



**Figure 3-6. Comparison of the wild type and region 2 mutant mass spectra.**

HR-MS and MS/MS fragmentation patterns of wild type (top) and AH02 mutant (bottom) showing the absence of tartrolons in AH02. The high resolution of the main peak in the wt spectrum is 873.4174 while that in the mutant AH02 is 873.5529 suggesting different compounds. In addition, the MS/MS fragmention patterns of each peak were different confirming that they are different compounds.

### 3.3.4. Analysis of the *trt* gene cluster

Region 2 of the previously sequenced *T. turnerae* T7901 (Yang et al., 2009) (accession number NC_012997.1) was analyzed using the BLAST tool in addition to the SearchPKS tool (Yadav et al., 2003) and the PKS/NRPS server (Bachmann and Ravel, 2009). It was found to be approximately 50 kbp containing 20 ORFs possibly involved in the tartrolons biosynthesis. Of these, ten (*trtA*-*trtJ*) seem to be the core biosynthetic genes and are organized in an apparent operon (Figure 3-7) including PKS domains (Table 3-5). More than 42.5 kbp of this region are formed by *trtDEF*; three large genes that encode *trans* AT type I PKSs. In addition, other biosynthetic genes are present including two hypothetical acyl transferases (AT) *trtAB*, two putative oxidoreductase *trtGI*, a putative polyketide cyclase *trtJ* and a type II thioesterase (TE) *trtH*. All PKS genes *trtDEF* lacked AT domains in their modules characteristic of a typical *trans* AT type I PKS gene cluster. No transport related coding sequences are present in the cluster or within the cluster's neighboring DNA segments.

The upstream and downstream flanking regions include ten more genes that might be involved in the transcription or resistance of tartrolon (Table 3-5). TERTU_2194 and TERTU_2212 may function as transcription regulators upstream and downstream, respectively. TERTU_2194 is similar to the transcription factor LysR, while TERTU_2212 is similar to a putative *Rho* independent transcription terminator (Lesnik et al., 2001). TERTU_2190 and TERTU_2191 encode possible aldolases that might be involved in the synthesis of the loading substrate. A transposase TERTU_2188 is encoded by a gene located upstream of the core genes suggesting a possible horizontal gene transfer event. TERTU_2193 is a potential ACP phosphodiesterase required for the

turnover of the ACP prosthetic group (Thomas and Cronan, 2005). TERTU_2211 is predicted to be a glycoside hydrolase for which we cannot assign a function in the tartrolon biosynthesis. Three ORFs with no predicted functions are also present; TERTU_2189, TERTU_2195, and TERTU_2209 (Table 3-5).

The three multimodular PKS ORFs contain eleven modules in addition to the loading module. The first of these PKS ORFs, *trtD*, was most similar to *dfnG* from *B. amyloliquefaciens* with a similarity percentage of 50%, involved in the biosynthesis of the macrolide difficidin (Chen et al., 2006; Chen et al., 2009). This is followed by *trtE* the largest ORF in this region which is similar to a PKS from *C. cellulolyticum* (Table 3-5). It was also found similar to *baeN* involved in the biosynthesis of bacillaene from *Bacillus amyloliquefaciens* FZB42 (Chen et al., 2009; Butcher et al., 2007) with a similarity of 44% (Table 3-5). The third PKS ORF, *trtF*, was similar to a PKS from *Paenibacillus polymyxa.* This ORF was also similar to *dfnD* involved in the biosynthesis of difficidin from *B. amyloliquefaciens* (Chen et al., 2006; Chen et al., 2009) with a similarity percentage of 48%. Other genes include *trtG* and *trtI* that seem to encode proteins for oxygenases. The integrated thioesterase in *trtF* is expected to cause release of the elongated chain while *trtH* is a type II thioesterase. The ORF *trtJ* encodes a potential polyketide cyclase (Kallio et al., 2006) (Table 3-5).

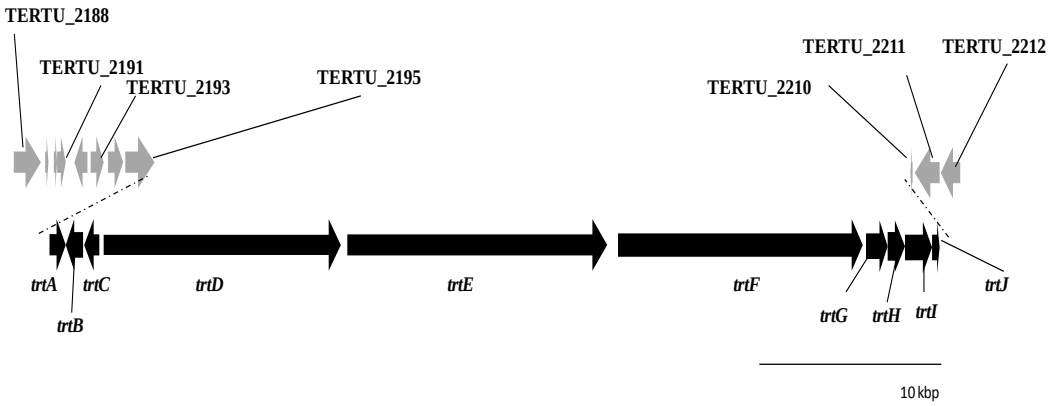



**Figure 3-7. Organization of *trt* biosynthetic gene cluster from *T. turnerae* T7901.**
Black arrows indicate genes assumed to be directly involved in the core biosynthesis of tartrolons.

**Table 3-5. Genes and functions of region 2.**

Size, proposed functions, source of closest homologs and their accession number and percentage of identity and similarity are determined. GNAT, Gcn 5-N-acetyl transferase; AT, acyl transferase; PKS, polyketide synthase; KS, ketosynthase; KR, ketoreductase; ACP, acyl carrier protein; DH, dehydratase; ER, enoyl reductase; MT, methyl transferase; TE, thioesterase.

| Gene Name | Number of amino acids | Proposed function | Organism | Identity/ Similarity Percentage (I/S) % | Accession Number |
|---|---|---|---|---|---|
| TERTU_2188 | 519 | transposase | *Cellvibrio japonicus* | 68/81 | YP_001983638.1 |
| TERTU_2189 | 62 | hybothetical protein | | | |
| TERTU_2190 | 68 | hybothetical protein | *Saccharophagus degradans* | 49/63 | YP_526685.1\| |
| TERTU_2191 | 171 | hybothetical protein | *Saccharophagus degradans* | 66/78 | YP_526685.1 |
| TERTU_2193 | 234 | ACP phosphodiesterase | *Idiomarina loihiensis* | 86/94 | YP_155995.1 |
| TERTU_2194 | 290 | Transcription Regulator LysR | *Marinobacter adhaerens* | 80/91 | ADP97201.1\| |
| TERTU_2195 | 549 | peptidase | *Plesiocystis pacifica* | 39/55 | ZP_01910797.1 |
| *trtA* | 196 | GNAT | *Burkholderia thailandensis* | 46/69 | ZP_05587728.1 |
| *trtB* | 287 | AT | *Streptomyces cattleya* | 47/62 | CCB76280.1 |
| *trtC* | 236 | Lipoprotein/enoyl coA isomerase hydratase | *Actinosynnema mirum* | 43/61 | YP_003099559.1 |
| *trtD* | 4539 | PKS<br>-Hyd-KR-FkbH-ACP<br>-KS1-DH-KR-ACP<br>-KS2-KR-ACP<br>-KS3 | *Bacillus amyloliquefaciens* | 35/50 | YP_001421792.1 |
| *trtE* | 4947 | PKS<br>-DH-ACP<br>-KS4-KR-ACP-<br>-KS5-ACP<br>-KS6-KR-ACP<br>-KS7-ACP<br>-KS8 | *Clostridium cellulolyticum* | 29/47 | YP_002505214 |
| *trtF* | 4663 | PKS<br>-KR-ACP<br>-KS9-DH-KR-MT-ACP<br>-KS10-ACP<br>-KS11-ACP-TE | *Paenibacillus polymyxa* | 30/47 | YP_003871371 |
| *trtG* | 377 | oxygenase | *Dickeya dadantii* | 44/64 | YP_002988660.1 |
| *trtH* | 268 | TE | *Xenorhabdus bovienii* | 31/51 | YP_003468018.1 |
| *trtI* | 455 | dioxygenase | *Pseudomonas fluorescens* | 56/72 | AAM12912.1 |
| *trtJ* | 151 | Polyketide cyclase | *Acidithiobacillus ferrivorans* | 35/55 | ZP_08490092.1 |
| TERTU_2209 | 51 | hypothetical | | | |
| TERTU_2211 | 486 | glycoside hydrolase | *Sorangium cellulosum* | 61/75 | YP_001617495.1 |
| TERTU_2212 | 378 | transcription regulator AraC family | *Saccharophagus degradans* | 33/54 | YP_527963.1 |

### 3.3.5. Biosynthesis of tartrolons

Correlation of the bioinformatic analysis of the domains in each module of region 2 with that of the tartrolon chemical substructure, in addition to mutational analysis confirmed that region 2 is the gene cluster responsible for the biosynthesis of tartrolons. We propose a route to the biosynthesis of tartrolons (Figure 3-8). Tartrolon biosynthesis starts by the loading of the three-carbon unit, D-lactate. A conjugated diene formed at C14-C17 is predicted to be formed by the KS2-KR-ACP-KS3-DH-ACP of modules 2 and 3 through a stuttering mechanism. Module 4 loads a saturated intermediate while KS5 of module 5 appears to be non-functional, based on the absence of conserved residues. Modules 6, 7, and 8 are responsible for the formation of the β-dihydroxy ketone region at C7-C11, while module 9 forms a saturated derivative. In addition, a C-MT domain in module 9 appears to load a methyl group at C20 possibly using S-adenosyl methionine. KS10, similar to KS5 appears to be non functional, while KS11 forms the ketone group at C1. We propose that the two ORFs *trtG* and *I* downstream of the cluster and predicted to encode oxygenases that are responsible for the formation of the acidic hydroxyl group at C2. The ketone group at C3 is reduced to a hydroxyl group by the oxygen at C7 to form a pyran ring. This leads to the formation of an α, β-dihydroxy acid moiety at C2 and C3. This is followed by the dimerization of two identical monomers to form a molecule with four hydroxyl groups capable of forming borate ester. Tartrolon D was easily transformed to its boron ester derivative, tartrolon E, just by the addition of boric acid as evidenced by TLC and MS (Figure 3-9). This suggests that the unboronated dimer binds boron in a Boesken complex form (Nakamura et al., 1977) without the need of an enzymatic reaction. This is also supported by the fact that deboronated derivatives of structurally

related compounds were converted to the boronated derivative through the addition of boric acid (Chen et al., 1981).

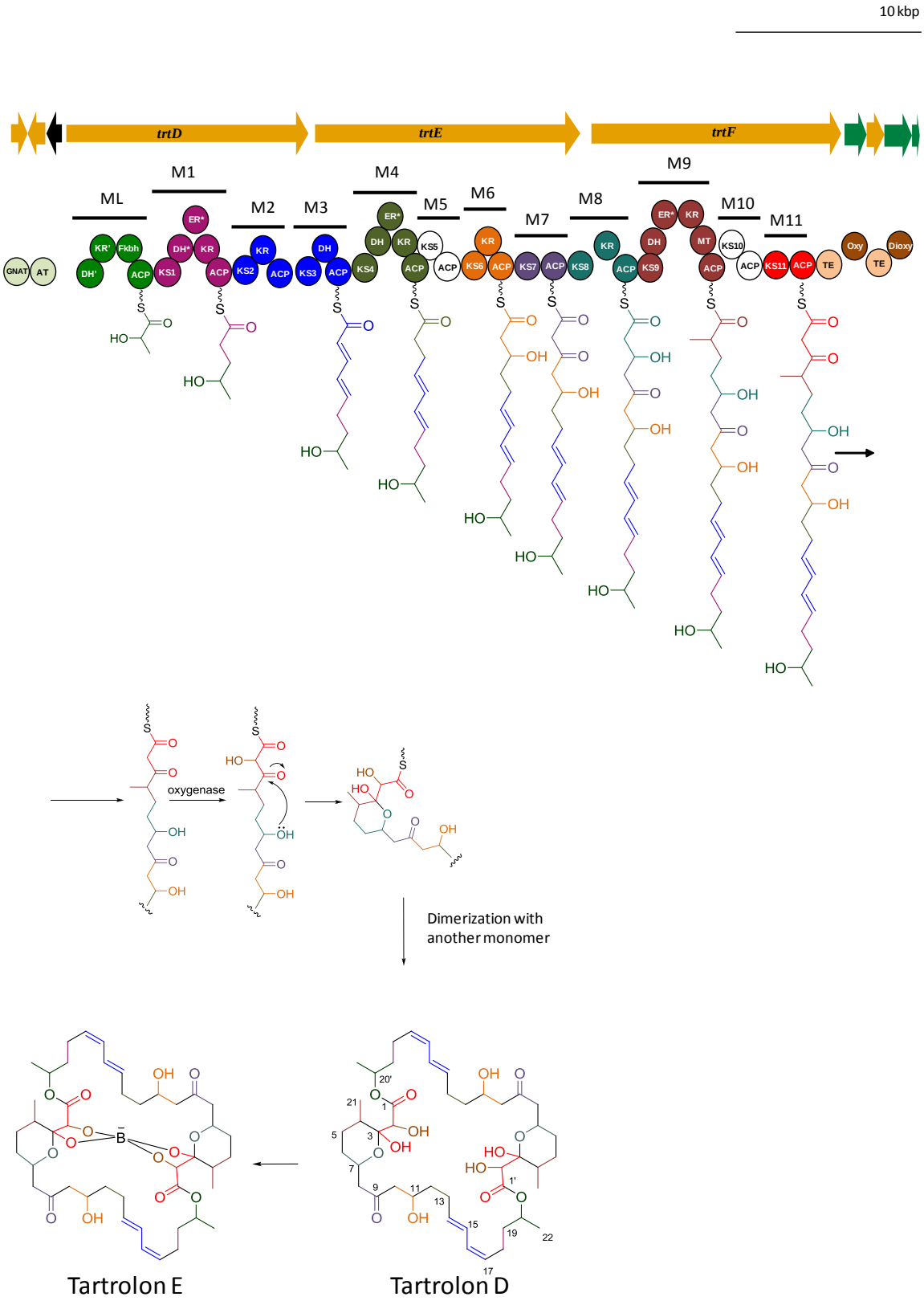



**Figure 3-8. Biosynthesis scheme of tartrolons from the *trt* cluster.**

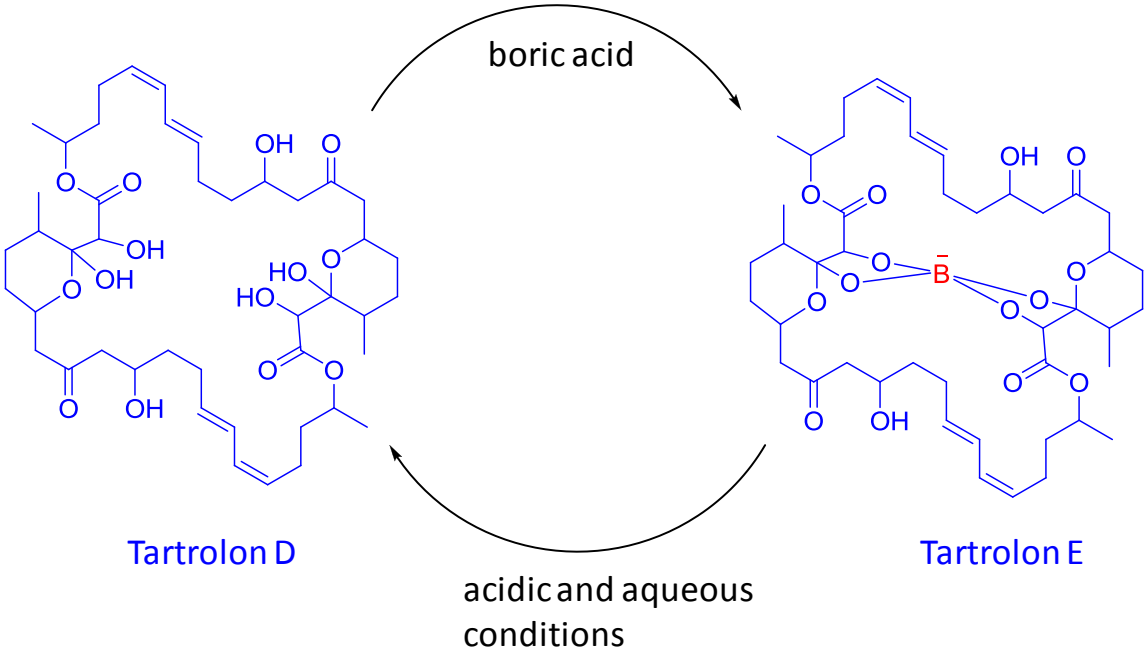



**Figure 3-9. Deboronation and formation of borate ester derivative of tartrolons E and D.**

### 3.3.6. Prevalence of tartrolons in different strains of *T. turnerae*

To determine whether tartrolon E, the major tartrolon in *T. turnerae*, is produced only in *T. turnerae* T7901 or is widespread in other *T. turnerae* strains isolated from different hosts, we examined eleven additional strains that were reported previously (Distel et al., 2002; Trindade-Silva et al., 2009) from different host species. DNA extraction of each of the *T. turnerae* strains followed by amplification of specific regions in the *trt* gene cluster with 2-3 primer pairs (Table 3-2) designed to target each of the three *trtDEF* ORFs of the biosynthetic gene cluster through polymerase chain reaction yielded products that were analyzed by agarose gel electrophoresis. Analysis of the amplified regions suggested that differences were mainly at the C-terminus of the *trtF* product in the same location. All strains showed evidence of the presence of the *trtDEF* ORFs (Table 3-6) suggesting the prevalence of the *trt* gene cluster in *T. turnerae*. However, presence of some of the genes belonging to the gene cluster does not confirm the presence of the entire gene cluster within the genome, or the chemical identity of the product.

To use an independent approach to assess prevalence of *trt* in *T. turnerae*, cultures were extracted and analyzed by HR-MS and tandem MS/MS. *T. turnerae* T7901 in addition to eleven other strains (Table 3-6) were analyzed under the same conditions and several measures were taken to confirm the presence or absence of tartrolons to avoid cross contamination in these strains. The analysis showed that tartrolon E is present in at least eight out of twelve strains tested (Table 3-6). The presence of tartrolons strongly suggests that these strains have the gene cluster within their genome, which is supported by the PCR results. However, strains that did not show the presence of tartrolons might contain a *trt* gene cluster within the genome that is silent under these growth conditions. All but

one of the MS positive samples showed strong PCR evidence of the *trt* cluster. The one exception, T8203, failed to amplify one fragment in *trtF*, but was positive for all the others. In this case, a slight difference in the gene sequence may be present. Two samples (T8602 and CS30) that showed positive PCR amplification for all fragments in the three ORFs did not show the compound using MS analysis, and thus have either non-functional or silent *trt* clusters. All of the strains showed at least some evidence of genes related to *trt*. Overall, tartrolon production appears to be common in *T. turnerae*.

**Table 3-6. Prevalence study of tartrolons in different *T. turnerae* strains.**

The prevalence of tartrolons was studied using HR-MS and MS/MS fragmentation, in addition to polymerase chain reaction. The host shipworm and habitat are stated. Primer numbers used refer to that in Table 3-2.

| Strain | Host | Host Source | MS, MS/MS | *trtD* | | | *trtE* | | | *trtF* | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 3/4 | 5/6 | 7/8 | 9/10 | 11/12 | 13/14 | 15/16 | 17/18 |
| **T7901** | *Bankia gouldi* | NC, USA | + | + | + | + | + | + | + | + | + |
| **T7902** | *Lyrodus pedicellatus* | CA, USA | + | + | + | + | + | + | + | + | + |
| **T7903** | *Teredo navalis* | MA, USA | + | + | + | + | + | + | + | + | + |
| **T8201** | *Psiloteredo healdi* | Maricaibo, Venezuela | + | + | + | + | + | + | + | + | + |
| **T8202** | *Teredo furcifera* | Hamilton Island, Bermuda | + | + | + | + | + | + | + | + | + |
| **T8203** | *Teredo bartschi (Clapp)* | Hamilton Island, Bermuda | + | + | + | + | + | + | + | + | - |
| **T8304** | *Nototeredo edax* | Andhra Pradesh, India | - | + | + | + | + | + | + | + | - |
| **T8401** | *Bankia rochi* | Pakistan | - | + | + | + | + | + | + | + | - |
| **T8402** | *Teredora malleolus* | MA,USA | + | + | + | + | + | + | + | + | + |
| **T8506** | *Teredo triangularis* | Hawaii, USA | + | + | + | + | + | + | + | + | + |
| **T8602** | *Dicyathifer manni* | Australia | - | + | + | + | + | + | + | + | + |
| **CS30** | *Neoteredo reynei* | Brazil | - | + | + | + | + | + | + | + | + |

### 3.3.7. Detection of tartrolons in the shipworm host using LC/MS and MS/MS

In order to determine whether tartrolons can be detected in shipworms using HR-MS and MS/MS, individuals belonging to two species of shipworms were used for extraction and analysis (Table 3-7). One of them, *L. pedicellatus,* is known to contain *T. turnerae*; and the other, *B. setacea,* lacks *T. turnerae* (in all specimens thus far examined). The relatively larger size of *B. setacea* allowed us to dissect it and isolate the caecum in one of our trials and use it for analysis. In case of *L. pedicellatus,* due to its small size, 4-7 shipworms were pooled and extracted. In each case samples were extracted with organic solvents, and then partitioned between ethyl acetate and water. The organic fraction was dissolved in a minimum amount of methanol and analyzed by HR-MS and MS/MS. The characteristic peaks for tartrolons were detected in the *L. pedicellatus* whole animal extract, and one of three *B. setacea* examined (Figure 3-10) confirming the expression of tartrolons in shipworms (Table 3-7).

Table 3-7. Detection of tartrolons in different shipworms.

| Sample | Shipworm sample | Characteristics | Presence/Absence of tartrolon |
|---|---|---|---|
| **1** | *B. setacea* | caecum of 1 shipworm | tartrolon E |
| **2** | *B. setacea* | pool of 5 shipworms | absent |
| **3** | *B. setacea* | 1 whole shipworm | absent |
| **4** | *L. pedicillatus* | pool of 7 shipworms | tartrolon E |
| **5** | *L. pedicillatus* | pool of 4 shipworms | tartrolon E |
| **6** | *L. pedicillatus* | 1 whole shipworm | tartrolon D |

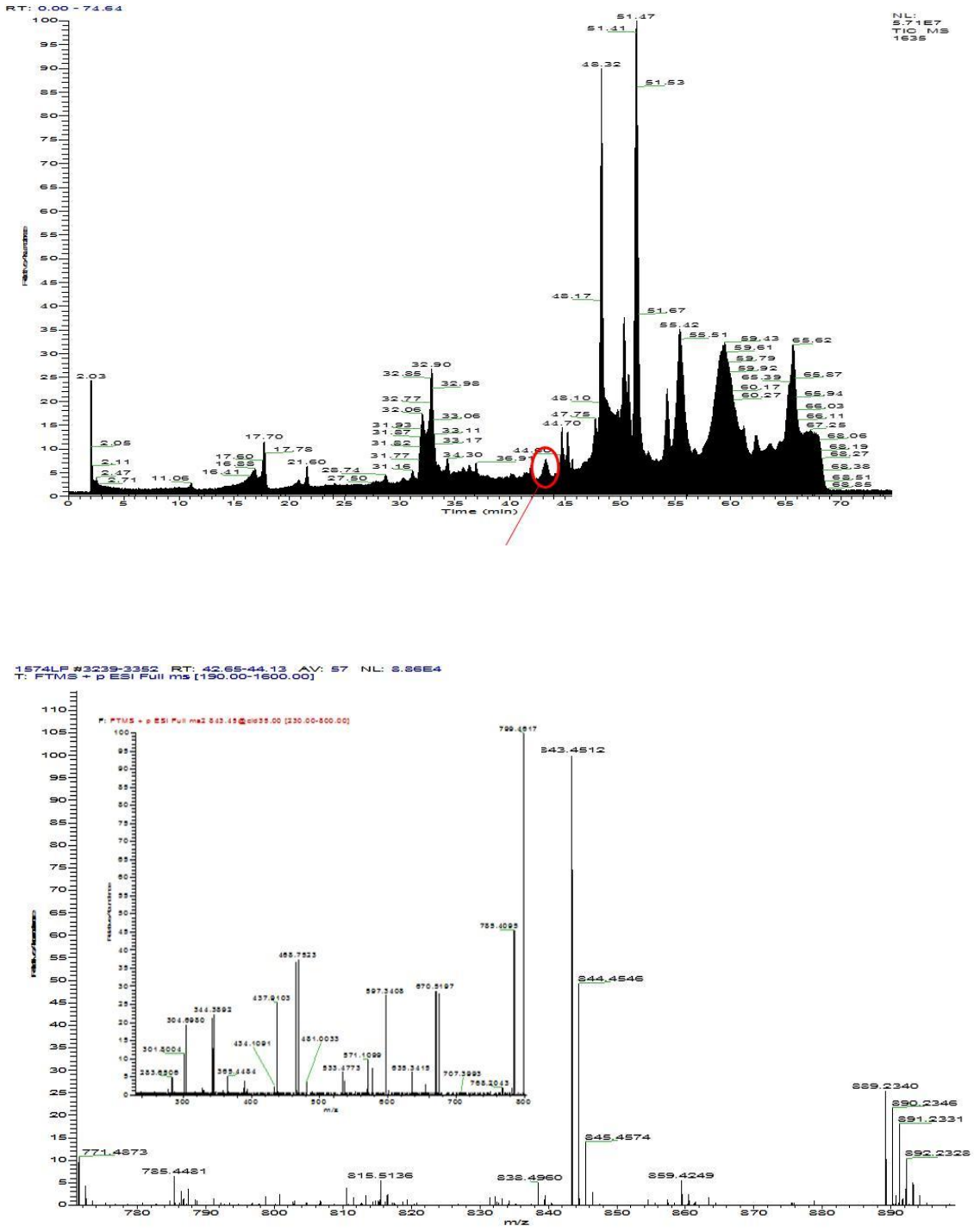



**Figure 3-10. HPLC-MS analysis of shipworm crude extract.**
HPLC chromatogram of a crude extract of a *L. pedicellatus* shipworm (top), HR-MS and MS/MS fragmentation pattern (bottom) of the peak corresponding to the tartrolon peak showing the presence of tartrolon D in the shipworm sample.

### 3.3.8. Bioactivities of tartrolons

#### 3.3.8.1. Antibacterial

Tartrolons D and E were found to have antibacterial activity. In fact, they were purified using this activity as a guide in a disc diffusion assay. *B. subtilis* was used as an organism of choice due to its rapid growth. However, both compounds failed to inhibit *E. coli* in disc diffusion assays. To determine the minimum inhibitory concentration (MIC) of tartrolon E, *B. subtilis* and *E. coli* were used as test organisms using a protocol reported before (Wiegand et al., 2008) in Mueller Hinton broth in a 96 well plate. The MIC of tartrolon E against *B. subtilis* was found to be 1 µg/ml (1.17 µM) while no inhibition activity was detected against *E. coli* at a concentration > 32 µg/ml (39 µM).

#### 3.3.8.2. Inhibition activity on other shipworm symbionts and associates

The effect of tartrolon E against several strains of marine shipworm symbionts from our ICBG collections were tested using a disc diffusion assay. Tartrolon E inhibited a representative symbiont from clade 3 (BS02), in addition to the marine pathogen *V. anguillarium* (Figure 3-11) but failed to inhibit any other organism from other clades. This shows that tartrolon E is capable of inhibiting marine organisms including shipworm symbionts and not only the lab test strain *B. subtilis*.

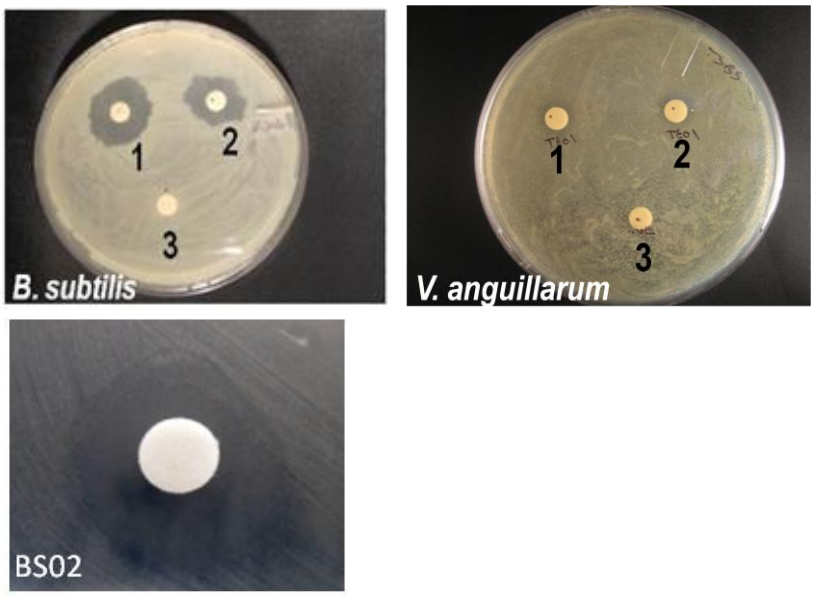



**Figure 3-11. Disc diffusion assays of tartrolons against different bacteria.**
*B. subtilis* (top left), *V. anguillarium* (top right), and BS02 (bottom). 1, tartrolon E; 2, tartrolon D; 3, chloroform and methanol as negative control.

### 3.3.8.3. Antibacterial activity against pathogenic bacteria

Tartrolon E was found to have antibacterial activity against other pathogenic bacteria. When tartrolon E was tested against *P. aeruginosa*, in addition to methicillin sensitive and methicillin resistance *S. aureus*, it possessed bacterial inhibition activity at the nanomolar level (Table 3-8).

**Table 3-8. Results of MIC antibacterial activities of tartrolon E against pathogenic bacteria.**

| Strain Tested | MIC in µg/ml (µM) |
|---|---|
| *P. aeruginosa* | 0.31 (0.36) |
| Methicillin sensitive *S. aureus* | 0.08* (0.095) |
| Methicillin resistant *S. aureus* | 1.25 (1.14) |
| * lowest concentration tested | |

### 3.3.8.4. Dorsal root ganglion assay

The neurotoxicity of tartrolon E was tested in a dorsal root ganglion assay (DRG) (Light et al., 2008; Peraud et al., 2009). The compound was found to increase and decrease the neuronal response in the majority of dorsal root ganglion neurons when tested at two concentrations; 100 µg/ml and 50 µg/ml, respectively (Figure 3-12). The increased and decreased intracellular calcium responses at these two concentrations might suggest that tartrolon E affects specific targets depending on the concentration.

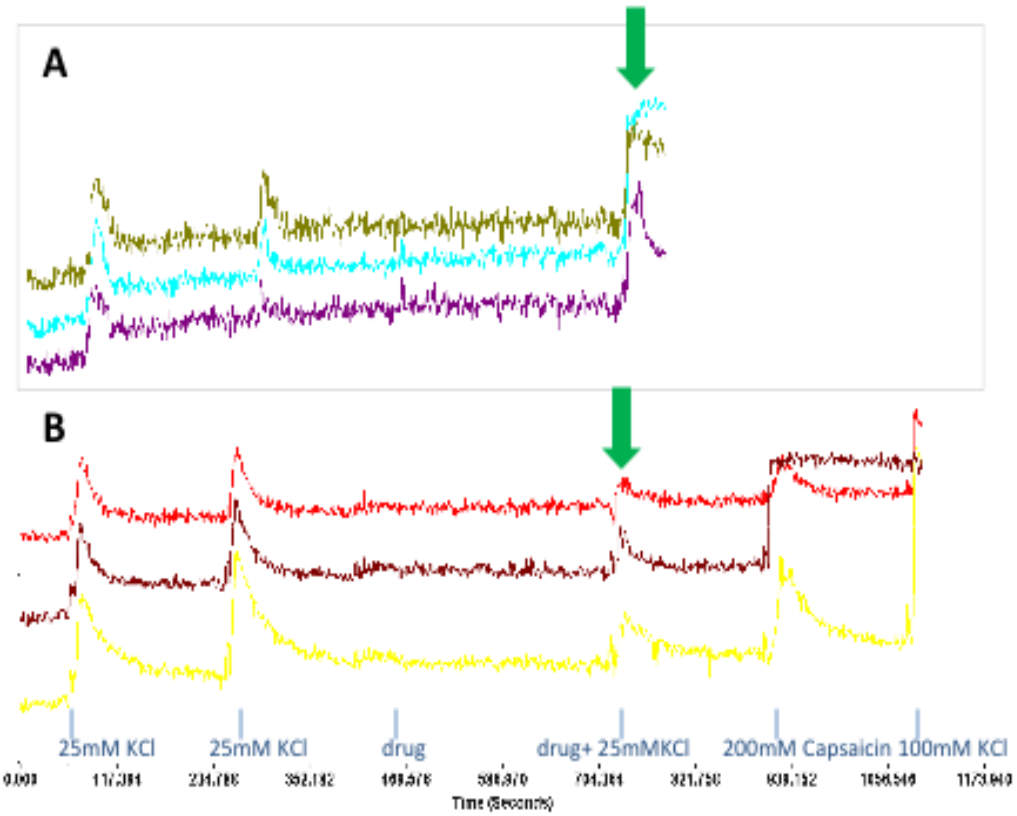



**Figure 3-12. Dorsal root ganglion assay for tartrolon E.**
Concentration at 100 µg/ml (top) and 50 µg/ml (bottom). Each colored trace represents a single neuron in the DRG population. Initial KCl injection is considered the control.

#### 3.3.8.5. Cytotoxicity

The cytoxicity of tartrolon E was measured against the breast cancer MCF-7 cell line. The compound was found to possess an $IC_{50}$ of 1.73 µg/ml corresponding to 2 µM. This suggests that tartrolon E has moderate cytotoxicity activity.

## 3.4. Discussion

Shipworm symbionts are unusual for their degree of wood specialization and their association with their host. Although their contribution to the shipworm-microbe symbiosis via cellulose degradation and nitrogen fixation has been previously proposed or confirmed, nothing has been known in terms of their secondary metabolites. In fact, the portion of the genome devoted to the biosynthesis of secondary metabolites is comparable to other antibiotic rich bacteria such as *Streptomyces* and *B. amyloliquefaciens*. This research constitutes the first analysis of shipworm symbiont secondary metabolites.

Production of these antibiotics increased in lower phosphate conditions. This agrees with previous reports that secondary metabolite production in microorganisms is regulated by the levels of phosphate (Martín, 2004). This mechanism might involve the two component system PhoR-PhoB. Indeed, analysis of the *T. turnerae* T7901 genome led to the detection of two genes. These two genes encode PhoB (TERTU_4563) a phosphate regulon transcriptional regulator and PhoR (TERTU_4562), a phosphate regulon sensor kinase. In addition, cultivation at lower agitation favored the production of these molecules in agreement with the fact that *T. turnerae* is microaerophilic, although capable of growing at higher agitation, this might not favor the production of these metabolites.

The two tartrolons D and E, identified in this study, belong to a group of macrodiolides with well known pharmacological activities (Table 3-4). Compounds structurally related to tartrolons are dimers or pseudodimers consisting of two polyketide chains joined as diesters. Members of this group are nearly identical in their carboxy terminal regions,

differing primarily in oxidation state and in chain length. Because of these close structural relationships, it is possible that the members of this family are biosynthetically related as well. It was interesting to see that the production of this class of compounds is not restricted to marine organisms since soil bacteria are also capable of producing them. In addition, their biosynthesis is not phylogenetically restricted, since their production was found in actinobacteria, cyanobacteria, deltaproteobacteria, in addition to the gammaproteobacterium; *T. turnerae*. This suggests convergent evolution, common ancestor or lateral gene transfer among these species reflecting an important function that this class fulfills.

Only a few examples of bioactive metabolite symbiosis are well-described in which the host harbors one or more bacteria capable of producing metabolites with antibacterial or antifungal activities. One of the most striking examples is detected in an insect-bacteria symbiosis (Kroiss et al., 2010). The female beewolf secretes symbiotic actinobacteria from its antennal glands into the larval cells. These *Streptomyces* produce a group of nine compounds that act at a distance to inhibit potential bacterial and fungal threats to help protect the larvae.

An interesting feature of tartrolons is their ability to bind the semi-metal boron. Most members of this group are reported to have antibacterial activity. Unboronated derivatives are also reported to possess the same activity; however it is difficult to exclude boron from experiments. It is possible that boron contributes to the activity by forming an anionic form that could disrupt the cell membrane and cause ion leakage. The first compound isolated and identified from this class was the D-valine ester; boromycin (Dunitz et al., 1971) which differs from other members in this group in that it is not a

symmetrical dimer and contains an amino group. This compound was found to have different pharmacological activities including anti-HIV activity. Aplasmomycin was isolated from a marine actinobacterium (Okami et al., 1976). It was found to have antiplasmodium activity among other activities, hence the name. Borophycin was identified from two different cyanobacterial *Nostoc* species (Hemscheidt et al., 1994; Banker and Carmeli, 1998) while tartrolons A and B were identified from the soil myxobacterium *S. cellulosum* So ce 678 (Schummer et al., 1994). Compounds structurally related to tartrolons were found to possess diverse pharmacological activities (Table 3-4). Tartrolons A and B were found to inhibit the DNA and RNA synthesis of *S. aureus* but not *E. coli* (Irschik et al., 1995) while tartrolon C was isolated from a soil *Streptomyces* species and found to have insecticidal activity (Lewer et al., 2003). Tartrolon D was identified in a cytotoxicity screening assay from a marine actinomycete species (Pérez et al., 2009). These authors also reported the boronated derivative for this compound as a minor contaminant in their original sample, which we have dubbed tartrolon E. Boron containing compounds seem to have interesting pharmacological activities. In fact, the tripeptide brotezomib, a boron containing small molecule approved by the FDA and marketed by Millennium Pharmaceuticals under the name Velcade®, is used for the treatment of multiple myeloma and mantle cell lymphoma (clinicaltrials.gov). Moreover, boron compounds have been reported in the treatment of different tumors in boron neutron capture therapy (Zhu et al., 2010). Tartrolon E also inhibited the growth of marine organisms such as *V. anguillarium* and the *B. setacea* isolate, BS02 (Figure 3-11). This suggests that tartrolons could play a role in microbial competition in the shipworm system.

The loading unit, D-lactate, is most probably formed from pyruvate which originates from glycerol. This assumption is based on feeding studies on compounds related to tartrolons, in addition to feeding studies performed on bryostatins (Kerr et al., 1996) where the loading module of *trtD* is similar to that in *bryA* in the *bry* gene cluster involved in the biosynthesis of bryostatins. The loading module in *bryA* is predicted to load the same starter unit, D-lactate. In the proposed bryostatin biosynthetic pathway, BryKS1 would recognize a D-lactate starter unit formed by the action of three domains included in the BryA loading module (Hildebrand et al., 2004). FkbH, a dehydratase-homolog (DH*), and a KR-homolog (KR*) would sequester and convert a three-carbon active intermediate from the glycolytic pathway, probably a phosphoglycerate. Interestingly, analyses of domain composition of modular PKSs showed that TrtKS1 is also preceded by the catalytic organization found in the loading module of BryA. Other evidence comes from feeding studies on one of the compounds in this group, aplasmomycin, its biosynthesis was studied and also proved that the loading substrate is D-lactate (Chen et al., 1980; Chen et al., 1981; Lee et al., 1987). Extensive feeding studies using labeled substrates were used to identify the building blocks of these compounds, which were found to be similar in that the building blocks are acetate and methionine in addition to glycerol.

Interestingly, the *trtDEF trans*-AT PKSs showed high similarities to enzymes from *Bacillus* species (Table 3-5) and especially PKSs from *B. amyloliquefaciens* FZB42 and *P. polymyxa* which, in turn, synthesize bioactive compounds related to the role of this bacterium in plant biocontrol (Chen et al., 2006; Chen et al., 2009). Considering that these PKS routes are thought to have evolved from an ancient gene cluster (Chen et al.,

2006) and that populations of *B. amyloliquefaciens* are found in marine and mangrove environments (Siefert et al., 2000), it is possible that the tartrolons biosynthetic route evolved from such ancient gene cluster and that *T. turnerae* acquired such DNA fragment through lateral gene transfer events.

The conjugated diene portion of the molecule is most probably formed by a KS2-KR-ACP-KS3-DH-ACP architecture of modules 2 and 3 through a stuttering mechanism. This mechanism has been reported previously in the biosynthesis of other conjugated dienes such as kalimantacin C (Mattheus et al., 2010), chivasazol (Perlova et al., 2006), macrolactin (Schneider et al., 2007), and difficidin (Chen et al., 2006). The *trt* gene clusters also contains the standalone type-II TE domain. This type of TE causes the hydrolysis of noncognate intermediates from the thiolation domain leading to regeneration of these misprimed domains (Hou et al., 2011).

In plants, boron reacts with some cell membrane constituents to increase the integrity of the membrane. Boron is essential in some bacteria for nitrogen fixation (Bonilla et al., 1990) and in the symbiotic relationship between plant and an actinomycete (Bolaños et al., 2002). There are several compounds that can bind boron including carbohydrates such as mannitol, glycerol, ribose, in addition to cofactors such as NAD. Boron is also chelated by cell surface compounds to increase the integrity of cell wall possibly to prevent gas diffusion (O'Neill et al., 1996). On the other hand, boron is also toxic at high levels to living organisms. Boron can inhibit enzymes and decrease nitrogen fixation at elevated levels. Decreased boron concentration was also correlated with change in nitrogen fixation (Masumi and Yukio, 1994) and the beneficial range seems to be narrow (Carrano et al., 2009). Boron, in the form of borate, reacts with diols to form an anion

that has a decreased permeability relative to the unchelated compound and thus plays an important role in their transport. Although boron was hypothesized, based on its characteristics, to be transported across the cell membrane via passive transport, boron transporters have been reported from other living systems suggesting that its transfer across the cell membrane is regulated through active not passive transport (Takano et al., 2002; Takano et al., 2010). We have searched the genome of *T. turnerae* T7901 for homologs for BOR1 and BOR1p, the recently discovered borate transporters in *A. thaliana* and *S. cerevisiae*, respectively but no hits were found. Just as some microorganisms have evolved biosynthetic pathways to acquire iron in the form of siderophores, others that lack boron transporters or those that have unique internal environments such as shipworms might have evolved molecules to facilitate boron transport or exclude toxic levels of boron. Given that the concentration of boron in seawater is estimated to be 400 µM (Bowmen, 1966), organisms might have evolved a control mechanism to either decrease its toxicity or make use of its abundance. In fact, the crystal structures of the universal quorum sensing molecule, AI-2 bound to its receptor was found to be in the deboronated form in terrestrial bacteria but in the boronated form in marine ones (Miller et al., 2004). Although, boron can be a result of glass contamination in some cultures (Schummer et al., 1994) this is less likely the case with marine organisms, given the high concentration of boron in the sea. *T. turnerae* is a marine organism not a terrestrial one and thus is subjected to high amounts of boron much higher than that present as glass contamination or a growth culture constituent. Another interesting feature of this group of molecules is that boron binding seems to be an important function of this molecule because of the presence of oxygenases needed to

produce the hydroxyl groups at C2 and C3. Boron was also found to bind certain catecholate siderophores (petrobactin) and dicarboxylate siderophores (vibrioferrin and rhizoferrin) isolated from marine microorganisms but not hydroxamates (aerobactin), much stronger than iron suggesting a role in binding this seawater abundant semimetal as a way to detoxify high concentrations of boron (Harris et al., 2007). The difference in boron binding between siderophores and tartrolons is that in the former and in other boron binding dicitrates, the formation of boron ester derivative appears to be through the α-hydroxy acid moiety while in the latter the formation of the boron bound complex appears to be through the two hydroxyl groups of α,β-dihydroxy ester moiety.

In addition to its possible role in boron transport in *T. turnerae*, tartrolons might also act as signaling molecules within the bacterial community or between the symbiont and its shipworm host. Tartrolons could also play a role in the microbial competition in the gills. Tartrolon E seems to possess a deterrent activity against certain members of the symbiont community, but not others. This might maintain a population of similar strains within distinct bacteriocytes. Another possibility arises from recent results (Betcher, 2011) that show that the caecum, the wood digestion organ of some shipworms, unlike other most xylophagous animals, has few microbes. Antibacterials might be used in the shipworm system to deter invaders that are harmful to the host. Antibacterial tartrolons produced by the symbionts in the gills might contribute to bacterial suppression in the caecum. This suppression could allow the host to maximize efficient uptake of the glucose liberated by the breakdown of cellulose. The mechanism by which products of the symbionts in the gill could be translocated to the caecum is unknown. Tartrolons may well serve multiple functions in the shipworm system.

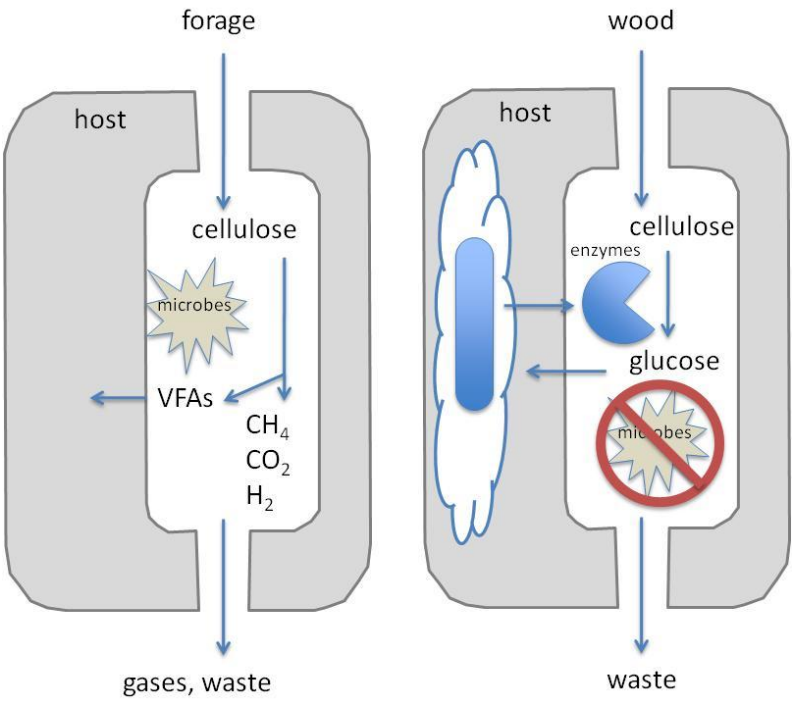



**Figure 3-13. Models of symbiotic cellulose digestion.**
Left: rumen (e.g., cow). Cellulose is digested by a diverse microbial community and fermented to volatile fatty acids (VFA) that are absorbed by the host. Right: proposed model for shipworm caecum. Symbionts in the gill provide enzymes that hydrolyze cellulose to glucose, which is absorbed by the host; competing microbes are excluded by bioactive metabolites produced by the symbionts (Courtesy, Haygood, MG).

This ability of different strains of *T. turnerae* isolated from different shipworms that live in various environmental conditions to produce tartrolons strongly suggests that tartrolons fulfill an important function in these strains either to combat other microorganisms or in the boron transport or detoxification mechanisms. This is also supported by the fact that the *trt* gene cluster was expressed in the shipworms as depicted by HR-MS. It is possible that tartrolons are expressed at a certain life stage or triggered by a specific signal in the system. This confirms that the *trt* gene cluster is not silent *in vivo* and has a potential role in the shipworm-microbial symbiosis. The fact that tartrolons were detected in one sample of *B. setacea* where no *T. turnerae* have yet been detected suggests that the

biosynthetic gene cluster is present in another shipworm symbiont, emphasizing its importance in the symbiosis. The fact that different symbionts might harbor similar gene clusters was shown in chapter II of this thesis. In addition, since lateral gene transfer event has been detected and confirmed between an endosymbionts *W. pipientis* and its host before (Hotopp et al., 2007), it is possible that the same event occurred between *T. turnerae* and the shipworm.

The gene cluster for the biosynthesis of tartrolons in *T. turnerae* was characterized in this study. In spite of the several members of compounds that belong to this class produced from different bacteria, no genes involved in their biosynthesis have been reported. This cluster will not only reveal the mechanism for the biosynthesis of tartrolons D and E but also provide insight for the gene clusters of other boron containing natural products. Moreover, this might increase the possibility of constructing combinatorial libraries for members in this class to form derivatives with higher activities and less toxicity.

In summary, we have isolated and identified two compounds from the marine shipworm symbiont *T. turnerae* T7901. The compounds we isolated were two macrodiolides tartrolon D and its boronated derivative that we called tartrolon E. These compounds belong to a group of structurally related compounds. We identified the biosynthetic gene cluster of these compounds that will shed more light about the biosynthesis of this class. We detected tartrolons in the shipworm host strongly suggesting that it plays a role in the shipworm chemical symbiosis.

1- Dr. Andrew Han (Oregon Health & Science University) constructed the gene insertion mutant AH02 in region 2 of *T. turnerae* T7901 (section 3.2.5 in methods).

2- Dr. Amro Hanora (Suez Canal University, Egypt) tested the prevalence of region 2 in different *T. turnerae* strains using polymerase chain reaction (part of section 3.2.7 in methods).

3- Ms. Malem Flores (University of the Philippines) tested the activity of tartrolon E against the pathogenic bacteria *P. aeruginosa*, and methicillin sensitive and methicillin resistant *S. aureus*. In addition, she tested the cytotoxicity of the compound against the MCF-7 breast cancer cell line (sections 3.2.9.3 and 3.2.9.5 in methods).

4- Ms. Rowena Antemano (University of the Philippines) tested the activity of tartrolon E in the dorsal root ganglion assay (section 3.2.9.4 in methods).

# 4. Geographical Distribution of Et743 in Different *Ecteinascidia* Populations

## 4.1. Introduction

Species of the genus *Ecteinascidia* are ascidians that are present in different places of the world especially the Caribbean, the Atlantic Ocean, in addition to the Mediterranean and the Red Sea. Most often they are found on mangroves, although they can be attached to docks, rocks, or concrete pilings. Each colony of this tunicate is composed of numerous transparent zooids connected at the base through stolons. At the top of each zooid, is a round opening called a siphon that opens to the water column and allows the tunicate to filter feed. These siphons are orange in color due to the presence of carotenoids. The *Ecteinascidia* spp. have ovoviviparous fertilized eggs and release lecithotrophic larvae into the water that, after settlement in a suitable location, metamorphose into a juvenile tunicate (Carballo, 2000). *E. turbinata* is known to be present in the Caribbean and the Mediterranean Sea while *E. thurstoni* is known to be present in the Red Sea (Gab-Alla, 2008). The larvae of the genus *Ecteinascidia* is thought to be protected through a group of compounds that are toxic to most other organisms (Young and Bingham, 1987), except the tiger flatworm *Maritigrella crozieri* that considers *Ecteinascidia* spp. its primary diet (Newman et al., 2000).

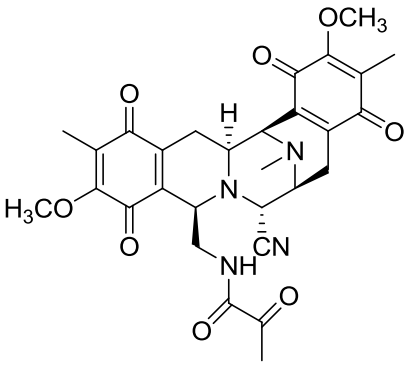
In the late sixties, the extracts of *E. turbinata* showed potent cytotoxic properties but the compounds were not identified until years later. These compounds were found to be a group of tetrahydroisoquinoline alkaloids that were called ecteinascidins. They were isolated from *E. turbinata* (Wright et al., 1990; Sakai et al., 1992) and recently found in *E. thurstoni* (Suwanborirux et al., 2002). One of the member of these compounds, Et743 (trabectedin) (Rinehart et al., 1990; Rinehart, 2000) attracted attention as a potent

anticancer compound that has been approved to treat soft tissue sarcomas and ovarian carcinoma. It is marketed under the name Yondelis®. The mechanism of action of Et743 is thought to be through binding of the DNA through a covalent interaction by alkylating the guanidine nucleotide with its carbinolamine group within the minor groove of the DNA (Aune et al., 2002; Pommier et al., 1996). This leads to interference with transcription, cell division, and DNA repair mechanisms (Takebayashi et al., 2001). Previously, the major supply for this antitumor drug was the ascidian itself, which provides milligram quantities per ton of animal weight. Since then, different approaches have been taken to solve the supply problem. Total synthesis failed to provide enough quantities in an economical manner (Fishlock and Williams, 2008; Vincent et al., 2007). Carballo and coworkers used aquaculture through two approaches based on the sexual and asexual reproduction of the tunicate to maintain a large supply of the Et743 (Carballo et al., 2000). However, the current and relatively economical method of providing Et743 in adequate amounts is through semisynthesis and fermentation starting from the homologous metabolite safracin B produced from the bacterium *Pseudomonas fluorescens* (Cuevas and Francesch, 2009).

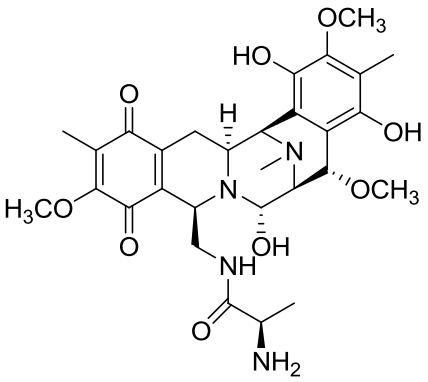
Ecteinascidins seem to be structurally related to the sponge metabolites renieramycins, the nudibranch metabolites jorumycin and jorunnamycins, in addition to three other bacterial metabolites; saframycin A, saframycin MX1, and safracins (Figure 4-1). These peptides seem to be formed from non-proteinogenic amino acids. This raises the possibility that the ecteinascidins, in addition to renieramycins and jorumycin, are produced by endosymbionts living within these animals via a modular nonribosomal peptide synthetase (NRPS) pathway similar to that found in bacteria that produce other

members of this family. Indeed, there have been reports linking a consistent specific endosymbiont with the *E. turbinata* adults and larvae (Salomon, 2002; Moss et al., 2003; Perez-Matos et al., 2007) and the potential producer symbiont was called *Candidatus* Ectoecteinascidia frumentensis. Since bacterial genes of a specific pathway usually cluster together, identifying a unique gene within the pathway might lead to identification of the whole biosynthetic cluster.

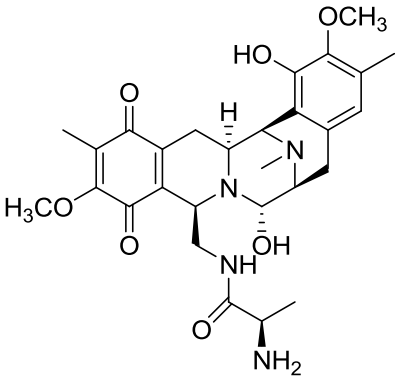
One of the factors impeding studies of marine natural products, is the chemical variation of active compounds in different species within a genus, different locations of similar species, and even sometimes different colonies of the same species and location (Slattery et al., 2001; Schmidt et al., 2012). This variation might be attributed to differences in predation pressure and possibly due to the variation in the complex marine environmental conditions that might affect the biosynthesis of these secondary metabolites. Thus, it is important from the environmental, biotechnological, and ecological point of view to identify tissues that are capable of producing these bioactive compounds. The main aim of this study was to study the distribution of the tetrahydroisoquinoline compounds and especially Et743 in different *Ecteinascidia* samples and to identify potential genes that might be involved in their biosynthesis. During the progress of this work, the Sherman group reported the identification of a gene cluster proposed to be responsible for the biosynthesis of Et743 (Rath et al., 2011).



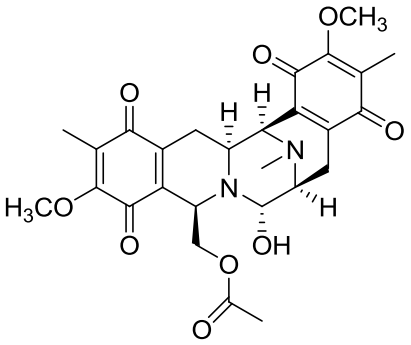
saframycin A
*S. lavendulae* (actinobacteria)



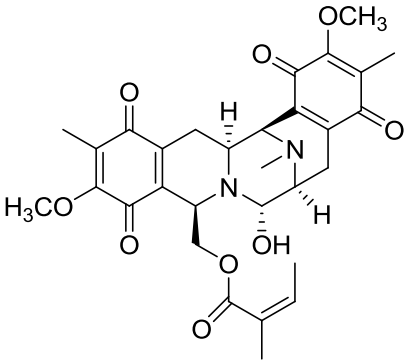
saframycin MX1
*M. Xanthus* (deltaproteobacteria)



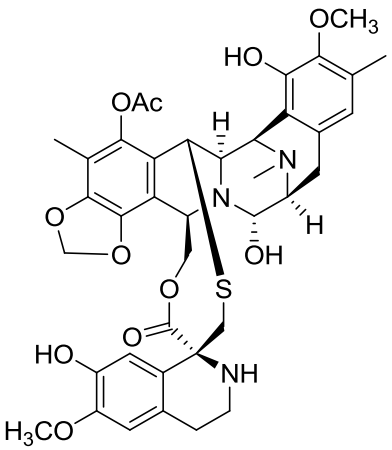
safracin B
*P. fluorescens* (gammproteobacteria)



jorumycin
*J. funebris* (nudibranch)



renieramycin E
*Reneira* (sponge)



Et743
*Ecteinascidia* (ascidian)

**Figure 4-1. Et743 and structurally related compounds and their sources.**

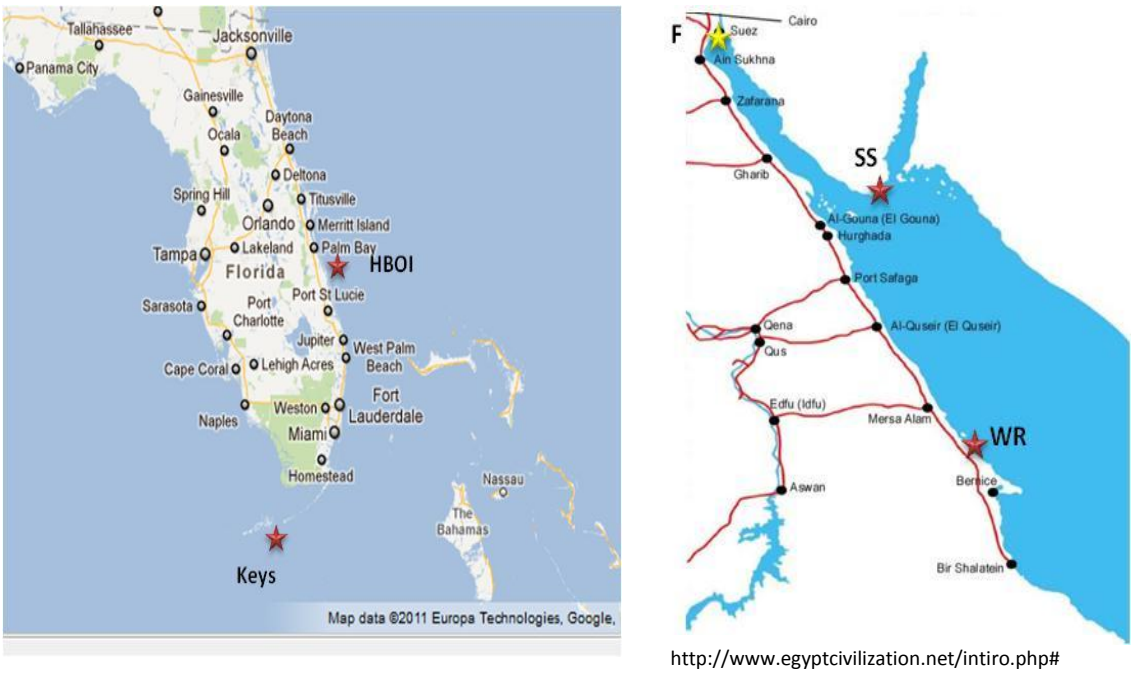
## 4.2. Materials and Methods

### 4.2.1. Sample collection

Samples were collected from two locations in Florida and from three locations in the Red Sea (Figure 4-2) and preserved for study (Table 4-1). *Ecteinascidia* animals collected from Florida sites were shipped to our lab at Oregon Health & Science University and kept alive for a few days in aquaria containing special salts (Red Sea salts). To confirm identity of the samples, 18s rRNA analyses of these samples were performed and reported elsewhere (Elbaz, 2009). Temperature and salinity of the aquaria were adjusted and monitored to be similar to that of its native environment. The ascidians were fed daily with ten drops of Phytofeast microalgal extract (Reef Nutrition, CA). *Ecteinascidia* larvae were collected from live animals from our aquarium using a Pasteur pipette and either stored at -80°C or extracted fresh. The ascidians were then homogenized or freeze dried and DNA organic extracts were prepared as shown below. Animals collected from the Red Sea were described previously (Gab-Alla, 2008). They were collected from the field and either stored in RNAlater (Qiagen) for DNA extraction or kept in seawater until frozen for organic extraction purposes.

**Table 4-1. *Ecteinascidia* samples collected with their location and habitat.**
ND, not determined.
HBOI, Harbor Branch Oceonographic Institute.

| **Sample** | Location | Coordinates | Habitat |
|---|---|---|---|
| ***E. turbinata*** | Florida-HBOI | ND | floating docks |
| ***E. turbinata*** | Florida-Keys | ND | *R. mangle* mangrove |
| ***E. thurstoni*** | Egypt-Wadi El-Ra'da | 24◦19_N 35◦ 20_E | *A. marina* mangrove |
| ***E. thurstoni*** | Egypt-Sharm Elsheikh (El-Rawisia) | 28◦ 11_N 34◦ 27_E | *A. marina* mangrove |
| ***Ecteinascidia* spp.** | Egypt-Fayed | 30◦ 19_N 32◦ 19_E | metal pilings |




http://www.egyptcivilization.net/intiro.php#

**Figure 4-2. Map showing locations of the five sites where *Ecteinascidia* samples were collected from Florida and the Red Sea.**

Red and yellow stars, presence and absence of Et743, respectively. F, Fayed; SS, Sharm El-Sheikh; WR: Wadi El-Ra'da. Maps of Florida (left) and the Red Sea (right).

### 4.2.2. Verification of Et743 content

Each sample was lyophilized and homogenized to powder followed by extraction and sonication with 50% chloroform in methanol (100 ml) twice and finally with 100% methanol (100 ml) twice. The organic extracts were dried under vacuum. All samples were partitioned between water and ethyl acetate. The organic fractions were filtered through anhydrous sodium sulfate and dried under vacuum while the aqueous fractions were lyophilized. Samples were dissolved in methanol and analyzed by direct MS first then confirmed by LC/MS.

For samples that did not show evidence of Et743, a different method previously reported was followed (Suwanborirux et al., 2002) using a different *Ecteinascidia* colony from the same collection. This method is advantageous in that it stabilizes the reactive

carbinolamine group as a cyanide derivative, in this case, Et770. Briefly, the ascidian sample (50 gm wet weight) was freeze-dried. The animals were then homogenized in methanol (50 ml). Phosphate buffer (200 ml) was added to the resulting solution to a pH of 7.0 (78 ml of sodium monobasic phosphate and 122 ml of sodium dibasic phosphate). Potassium cyanide was added to the solution which was allowed to react for six hours to convert the Et743 into the less reactive derivative. Methanol was added to the mixture and the methanol/aqueous extract was concentrated. This mixture was partitioned with ethyl acetate. This portion was evaporated under reduced pressure to yield the crude extract (62 mg).

In addition, a pure form of the standard Et743 was obtained from the National Cancer Institute. This standard (80 μg) was dissolved in methanol (700 μl) and divided to two vials. The first was dried and left as Et743. The second was transformed into the cyanide derivative Et770 using phosphate buffer (300 μl) and potassium cyanide (700 μg). The reaction was allowed to proceed for six hours and extracted by equal amounts of ethyl acetate. The reactions of both the organic extract and the standard were monitored by TLC. Each sample was dried and redissolved in methanol and injected in a ThermoElectron LTQ-Orbitrap HR-MS and HPLC/MS column. A gradient of 20-100% acetonitrile in water over 20 min was used on an Ascentis C18 column 250 × 4.6 mm (Supelco).

### 4.2.3. DNA extraction

Commercial kits failed to produce DNA of high molecular weight and acceptable purity. Thus, we developed a method slightly modified of one reported previously (Sambrook and Russell, 2001). Briefly, ascidian colonies were rinsed twice with filter-sterilized

artificial seawater, frozen for 15-30 sec in liquid $N_2$ and ground with a sterile mortar and pestle. The ascidian colonies (300 mg) were used for extraction of metagenomic DNA. Samples were centrifuged and supernatants were collected in new clean sterile tubes. Lysis buffer (100 mM Tris-HCl, 10 mM EDTA, 37 µM Triton 100×, 100 µg RNase A /ml) was added to the pulverized tissue and allowed to thaw. Lysozyme was added to the homogenate at a concentration of 0.2 mg/ml, and the samples were incubated for 30 min at 37$^{o}$C. Proteinase K (0.5 mg/ml) was added and the suspension was incubated for further 30 min at 60$^{o}$C with occasional shaking. CTAB (cetyltrimethyl ammonium bromide) extraction buffer (1.4 N NaCl, 20 mM EDTA, 100 mM Tris-HCl, 3% CTAB, and 1% β-mercaptoethanol) was added to the suspension and the sample was further incubated for 30 min at 65$^{o}$C. This was followed by the addition of 1× volume of phenol: chloroform: isoamyl alcohol and mixed gently for 15 min by gentle rotation. Samples were centrifuged at 5,000 ×g for 10 min to separate phases. The aqueous layer was mixed with 2× volume of chloroform/IAA and rotated gently for 15 min, followed by centrifugation at 5,000 ×g for 10 min at room temperature to obtain separate phases. DNA was precipitated by the addition of 0.1× 3 M sodium acetate and 0.8× isopropanol. DNA was either spooled or mixed by inverting gently until a stringy white precipitate was visible. This was followed by centrifugation at 15,000 × g for 20 min and the supernatant was discarded. DNA was washed with 70% ethanol and left to air dry for ten min. The DNA was resuspended in $T_{10}E_1$ buffer in 4$^{o}$C overnight. DNA concentrations were measured using ND-1000 NanoDrop spectrophotometer. The quality of the DNA was confirmed using $A_{260/280}$ and $A_{260/230}$, gel electrophoresis, and its ability to amplify using universal bacterial primers.

### 4.2.4. Further purification of DNA

Although DNA obtained by the above extraction method was amplifiable, metagenome libraries construction require high quality of pure insert DNA in order to maximize the titer number. Thus, two methods were used for further DNA purification; sucrose gradient and formamide denaturant buffer.

#### 4.2.4.1. Sucrose density gradient

Sucrose gradient density has the advantage of size enrichment, which is beneficial for the construction of genomic libraries, in addition to the removal of inhibitory pigments. We used a method slightly modified from that reported elsewhere (Hildebrand et al., 2004). Briefly, different concentrations of sucrose solutions were prepared; 10, 20, 30, and 40% of sucrose w/v in a buffer of 1 M NaCl, 20 mM Tris Cl (pH 8.0), and 5 mM EDTA (pH 8.0). The four solutions were filter sterilized using a 0.22 µm nitrocellulose filter. The metallic centrifuge tubes were cleaned by 10% bleach and 95% alcohol followed by sterile deionized water and left to dry. A gradient was formed by adding equal amounts of different sucrose solutions starting from the highest concentration to the lowest. The tubes were left to settle for two hours and then frozen at -20$^{o}$C. These were left to thaw at 4$^{o}$C to allow the gradient to linearize. Sucrose solutions were added as described above and 30 µg DNA was added followed by centrifugation. The conditions used were $rpm_{SW60}$ = 25,000, t=22hrs, and T = 22$^{o}$C. Aliquots of 100-150 µl were taken and checked by gel electrophoresis. This method proved successful in size separation and getting rid of impurities. Fractions containing DNA of higher molecular weight than 30 kbp were combined. DNA was separated from the sucrose using an Amicon 10K column

(Millipore) to concentrate the sample. This was followed by centrifugation through a zeba desalting spin column (Thermo Scientific).

#### 4.2.4.2. Formamide denaturant buffer

We used a method slightly modified of that reported before (Liles et al., 2008) which is a modification of another reported elsewhere (Sambrook and Russell, 2001). This method involves the use of two ingredients; formamide which causes DNA denaturation and helps to dissociate DNA-protein complexes and sodium chloride that increases the DNA stability.

Briefly, DNA was run into 1% low melting point agarose in 1× TAE using gel electrophoresis. The gel part containing the DNA was excised and placed in formamide denaturation buffer (20 mM Tris-Cl pH=8.0, 0.8 M NaCl, 80% v/v formamide). The gel plug was incubated overnight at 10$^{o}$C followed by dialysis in dialysis buffer 1 (20 mM Tris Cl (pH = 8), 0.1 M NaCl, 10 mM EDTA) for 12 hours followed by 12 hours dialysis in dialysis buffer 2 (10 mM TrisCl, 10 mM NaCl, 0.5 mM EDTA). DNA was recovered from the agarose plug using GELase™ enzyme (Epicentre Biotechnologies).

### 4.2.5. Identification of a gene candidate

Based on the conserved and unique reductase domain in the gene clusters of safracin B, and saframycins A and MX1, two pairs of degenerate primers were used, TGf/PVr, as an external pair and another internal, QAf/KWr (Table 4-2). Nested PCR was performed using the external degenerate primers for the reductase domain using Phusion polymerase (NEB). PCR conditions were as follows: initial denaturation at 98$^{o}$C for 3 min followed by 35 cycles at 98$^{o}$C for 15 sec, 58$^{o}$C for 15 sec, and 72$^{o}$C for 20 sec followed by final

extension at 72$^{o}$C for 10 minutes. The PCR product was purified using Qiaquick PCR purification kit (Qiagen) and subjected to another round of amplification using the internal degenerate primers for the reductase domain. PCR conditions were as follows: initial denaturation at 98$^{o}$C for 3 min, then 35 cycles at 98$^{o}$C for 15 sec, 62$^{o}$C for 15 sec and 72$^{o}$C for 15 sec and final extension at 72$^{o}$C for 10 min. The PCR product of expected size (458 bp) was gel purified and cloned into pJET1.2 vector (Fermentas) and transformed into DH5α competent cells (NEB). Cells were screened by PCR using insert primers to confirm the correct insert and plasmids containing the expected insert were purified using PureYield plasmid miniprep system (Promega) using manufacturer's recommendations. The plasmids were sequenced using pJet1.2F and pJET1.2R primers. From this putative reductase domain specific primers were designed and the conditions were optimized for RedEt3F/RedEt3R (Table 4-2) to produce a target of 424 bp. These PCR conditions were as follows; initial denaturation at 95$^{o}$C for 5 min followed by six cycles of touchdown PCR at 94$^{o}$C for 30 sec, 60$^{o}$C for 30 sec and decreasing 1$^{o}$C every cycle, and 72$^{o}$C for 30 sec. This was followed by 30 cycles of 94$^{o}$C for 30 sec, 54$^{o}$C for 30 sec, and 72$^{o}$C for 30 sec and a final extension at 72$^{o}$C for 10 min.

In addition, several attempts were carried out to amplify these DNA samples under gradient annealing temperatures using A3F/KWr and YTSGf/KWr (Table 4-2; Figure 4-5), where A3F and YTSGf are degenerate forward primers for the adenylation domain, to prove that the reductase domain is in fact linked to an adenylation domain. The PCR products were purified using Qiaquick PCR purification kit and cloned into pJET vector (fermentas) and transformed into competent cells. Cells produced were screened using PCR and positive cells were subjected to plasmid purification. Plasmids were sequenced

and investigated using NCBI blast. The sequenced data did not show recognizable domains.

In order to check if the samples collected and extracted in this study contains the thiolation-reductase domain present in EtuA2 reported previously (Rath et al., 2011), two primers E3ART-F and E3ART-R (Figure 4-5; Table 4-2) were used. A touchdown PCR cycle was performed as follows; initial denaturation at 98$^{o}$C for 2 min followed by nine cycles of touchdown PCR at 98$^{o}$C for 15 sec, 65$^{o}$C for 15 sec and decreasing 1$^{o}$C every cycle, and 72$^{o}$C for 45 sec. This was followed by 30 cycles at 98$^{o}$C for 15 sec, 56$^{o}$C for 15 sec, and 72$^{o}$C for 45 sec and a final extension at 72$^{o}$C for 5 min. PCR products were purified using Qiaquick PCR purification kit (Qiagen) and sequenced using amplification primers.

### 4.2.6. Construction of fosmid libraries from *Ecteinascidia*

Fosmid libraries were constructed using CopyControl Fosmid Library Production Kit (Epicentre Biotechnologies) according to the manufacturer's protocol with slight modifications. Briefly, DNA was end repaired and applied onto low melting point agarose. DNA of size 30-45 kbp was separated using a CHEF mapper (Biorad, CA) with conditions: FIGE mode 180$^{o}$ angle, 1× TAE at 14$^{o}$C, 9 V/cm forward, 6 V/cm reverse, switch time 200-800 msec, forward time = reverse time = 18 hours. The DNA was retrieved from the gel using Agarase enzyme (NEB) followed by ligation into the CopyControl pCC1FOS vector using ligase enzyme (NEB) at 16$^{o}$C overnight. The ligated vector was packaged for two hours followed by two more hours. Infection occurred in the recommended medium with the addition of 0.2% maltose. Infection took place at 30$^{o}$C for a total of 90 min. Screening of the library occurred by two methods. Colonies from

plates were pooled together. The plasmids of the pooled colonies were extracted using miniprep and the pool was screened by PCR using specific RedEt3F/R and QAf/KWr primers (Figure 4-5; Table 4-2). In addition, a different method was used that involves liquid gel pool (Hrvatin and Piel, 2007) based on the use of ultra low melting point agarose (Elsaesser and Paysan, 2004).

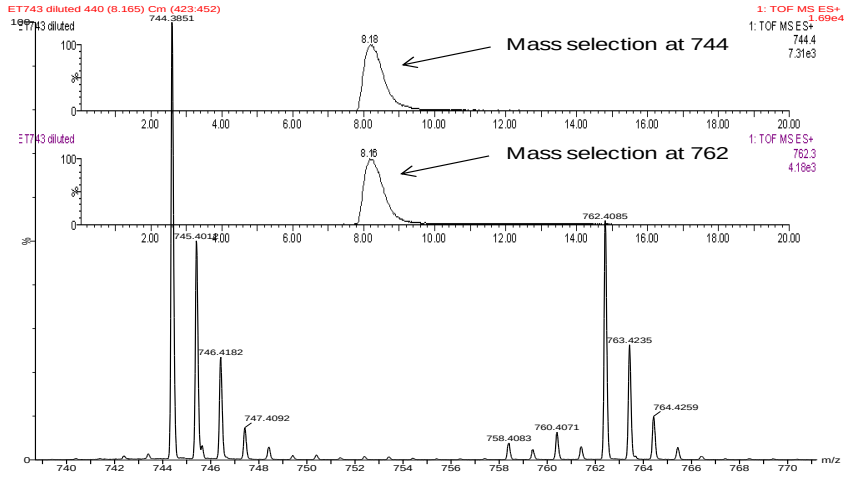
## 4.3. Results

### 4.3.1. Verification of Et743 content

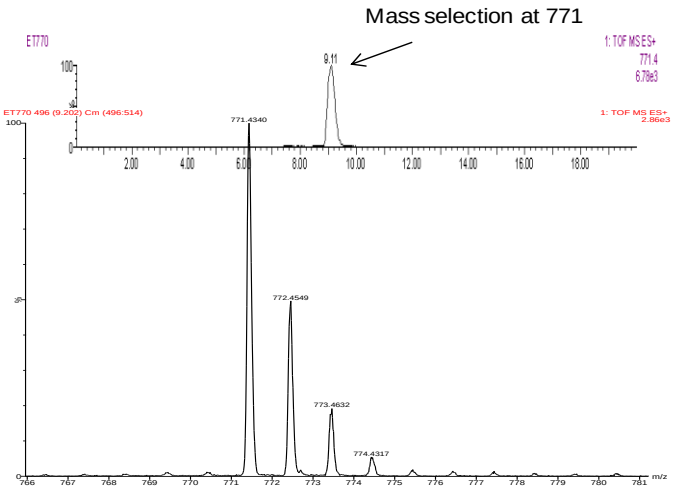
Ecteinascidins are reactive molecules through their carbinolamine group. It is possible that irreversible reactions might occur with Et743 and other proteins or compounds leading to absence of the characteristic peaks, especially with tissues that contain trace amounts of ecteinascidins. Thus, samples that showed no evidence of the presence of Et743 were subjected to the more sensitive cyanolation method. This method depends on the use of potassium cyanide to react with the highly reactive carbinolamine group and convert it into the stable cyanide derivative. A pure standard of Et743 was used as a positive control.

The chromatograms of the crude extract from the ascidian, including diode array, total ion chromatogram (TIC), and filtered chromatograms for *m/z* = 744 (Et743 minus water), 762 (Et743 parent compound), and 771 (the CN adduct) are shown in Figure 4-3. HR-MS and LC/MS showed that Et743 was present in both locations from Florida (Table 4-3). In addition, *Ecteinascidia* in two out of three locations of the Red Sea (SS and WR but not F) showed Et743. The Fayed sample did not show any evidence for the presence of Et743 even after using the more sensitive cyanolation method (Table 4-3).

A.




B.




C.

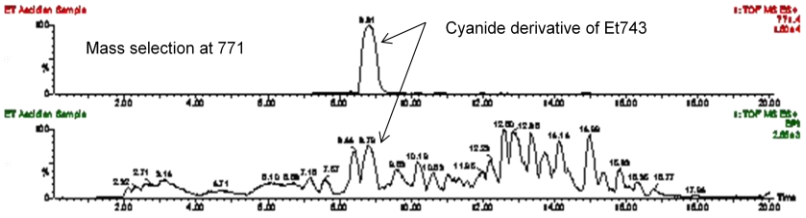
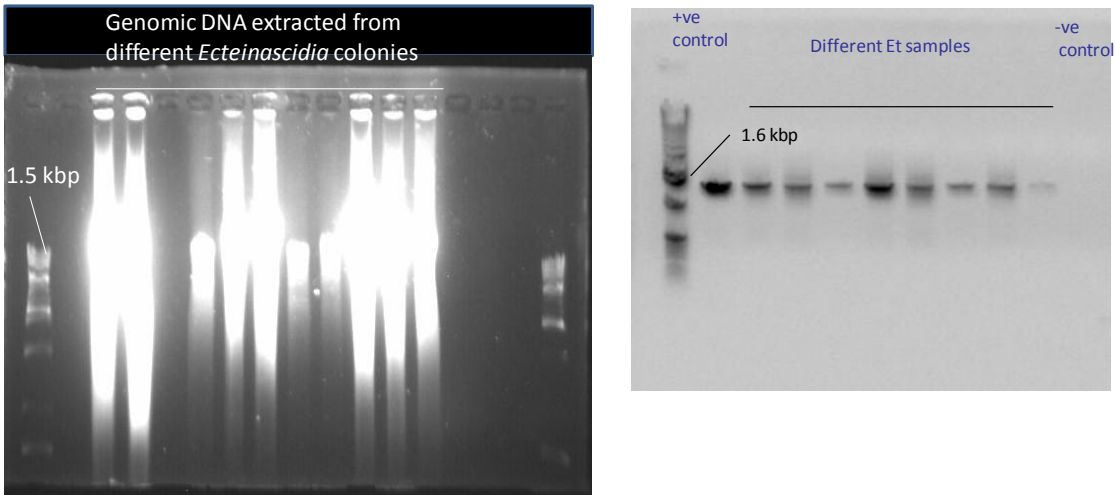



**Figure 4-3. Representative HPLC/MS spectra of crude extract and Et743 standard.**
LC/MS specra of A) Et743 and B) its cyanide derivative. C) HPLC chromatograms of the crude extract showing evidence for the presence of Et743 selected for 771 to detect the cyanide derivative (top) and total chromatogram with no mass selection (bottom).

### 4.3.2. DNA extraction and purification

It is clear that the modified CTAB DNA extraction method yields high molecular weight DNA sufficient to construct genomic libraries (Figure 4-4). In addition, the DNA is free of PCR inhibitors as evidenced by the amplification with universal bacterial primers (Figure 4-4). In spite of the high molecular weight of the DNA from our sample and its apparent purity, we did not get high titer numbers from our genomic libraries. We hypothesized that inhibitory pigments or proteins might still be present that interfere with some enzymatic reactions such as ligation or end repair of the DNA. Using sucrose gradient method (Figure 4-4) or formamide denaturant buffer increased the purity of our DNA sample but in case of the latter, DNA was sheared slightly less than that required for cosmid size libraries (Figure 4-4). Only the sucrose gradient method improved the titer number.

A. B.




C. D.

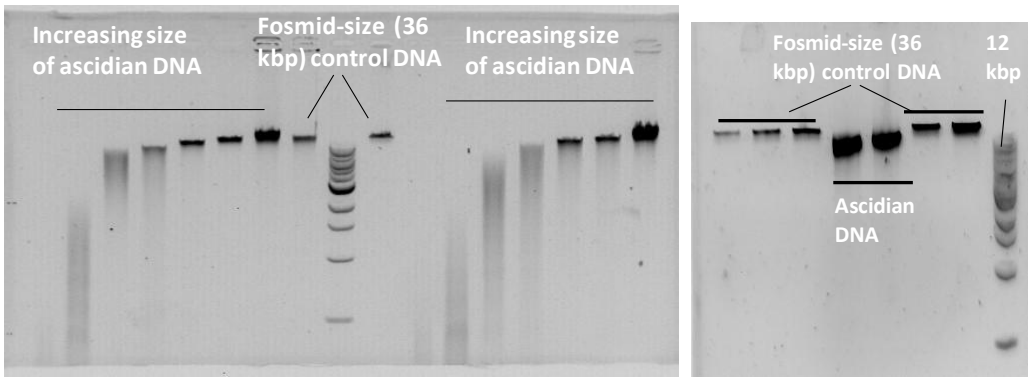



**Figure 4-4. Gel electrophoresis of *E. turbinata* metagenomic DNA and different purification methods demonstrating high molecular weight.**
A) Gel photo showing high molecular weight DNA from different CTAB phenol chloroform extraction method and B) successful amplification using universal primers. Results of C) sucrose gradient method showing different sizes of DNA and D) formamide denaturant buffer method showing decreased size of DNA after purification.

### 4.3.3. Amplification of a candidate gene homologous to reductase domain

Our strategy for recovering Et743 biosynthesis genes is based on a prediction of similarity between the Et743 biosynthesis pathway and that of the three model

compounds; safracin B (Velasco et al., 2005), saframycin A (Li et al., 2008), and saframycin MX1 (Pospiech et al., 1995) that has been published previously (Figure 4-5). The similarity is in the reductase domain present at the C-terminus of the last NRPS module. This module is expected to contain a reductase domain that causes cyclization of tetrahydroisoquinoline compounds. A nested PCR approach was performed using primers designed to capture the unusual reductase domain (Table 4-2) on a range of different *Ecteinascidia* populations from two locations from Florida including adults and larvae in addition to one tiger flatworm expected to contain Et743 (Table 4-3). A 450 bp gene fragment, that we called *ectC* was obtained whose sequence is similar to the corresponding gene in gene clusters of structurally related compounds (Figure 4-6). This dominant sequence was found to be 38-42% identical to the reductase domains of the three model pathways on the amino acid level (Table 4-4; Figure 4-6). The same sequence was dominant among all adult, larvae, and flatworm samples obtained from the two locations of Florida. Specific primers RedEt3F/R (Table 4-2) were designed based on sequences from *E. turbinata* Florida populations to screen the metagenomic library constructed from adult *E. turbinata* samples (see below).

However, when the same nested PCR approach was used on the Red Sea samples, only two locations, Sharm El-Sheikh (SES) and Wadi El-Ra'da (WR) showed amplification while three different collections from Fayed did not give an amplicon of the expected size. Sequencing of the amplified products from SES and WR revealed products that are closely related to *ectC* from Florida (Figure 4-7: Table 4-4). Moreover, in spite of difference in sequences, they all had the characteristic triad system (serine, tyrosine, and lysine) of the reductase domain. The sequences were different, yet similar to the

reductase sequence obtained from the Florida samples and the corresponding sequences from saframycins pathway (Figure 4-7). These results suggested that the gene fragment is a part of the biosynthesis of tetrahydroisoquinoline compounds.

Results were consistent with chemical analysis of Et743, i.e. the presence of both the compound and the *ectC* gene were correlated (Table 4-3) suggesting that *ectC* might be involved in the biosynthesis of these types of compounds. In addition to providing evidence that these two samples are capable of producing Et743, these samples could be used to construct fosmid libraries and screen the clones.

In addition, the *Ecteinascidia* samples were also screened using primers to amplify the thiolation-reductase domain in *etuA2* that was recently reported (Rath et al., 2011). Two primers E3ART-F and E3ART-R (Table 4-2; Figure 4-5) were used to target the 1.25 kbp region. Samples from the two Florida locations amplified and gave the expected product (Figure 4-8; Figure 4-9). Surprisingly, the same amplicon was produced when *Ecteinascidia* population from all sites in the Red Sea were tested including the Fayed population (Table 4-3).

A. B.

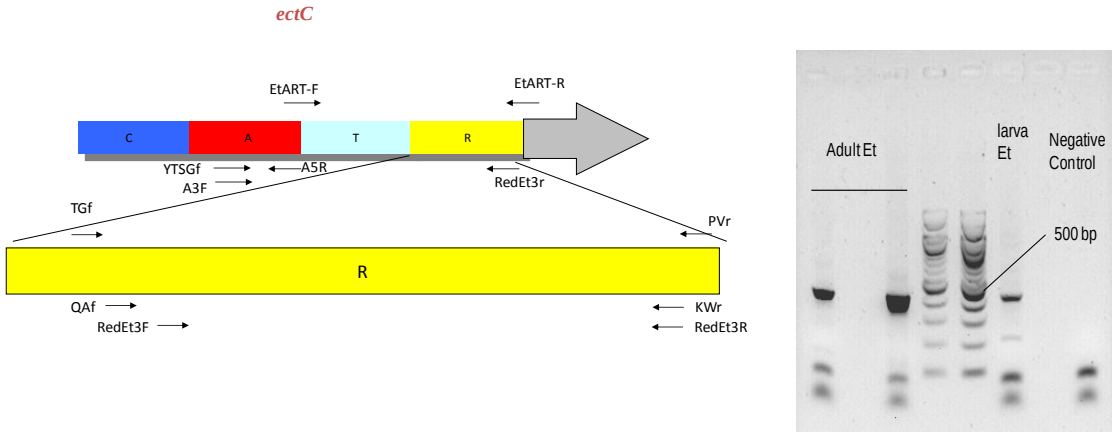



**Figure 4-5. Amplification strategy for reductase domain.**

A) Schematic diagram for primer locations. B) Amplification of putative Et reductase with specific primers RedEt3F and RedEt3R from adult and larval metagenomic DNA.

**Table 4-2. Sequence of primers used in the *Ecteinascidia* survey.**

| Primer name | Primer sequence | Source |
|---|---|---|
| TGf | 5’-CGG CGC CAC CGG NTW YBT NGG-3’ | Schmidt et al., unpub data |
| PVr | 5’-TGG CCT TGG CCA CGT RRT CAN CNG G-3’ | Schmidt et a.l, unpub data |
| QAf | 5’-CCA GGC CAG GGT GTA CTG YYT NGT NMG -3’ | Schmidt et al., unpub data |
| KWr | 5’-TCA CCA GCT GCT CGG CNR CCC AYT T -3’ | Schmidt et al., unpub data |
| YTSGf | 5’-CCT ACG TGA TCT ACA CCT CCG GNW SNA CGG -3’ | Schmidt et al., unpub data |
| A3F | 5’-TAY ACB TCH GGI WCI AAR GC -3’ | (Paungmoung et al., 2007) |
| A5R | 5’-YTC BGT IGG BCC RTA KGC -3’ | (Paungmoung et al., 2007) |
| RedEt3F | 5'-TGT GCG AGC CGA AGA TCT GGA-3’ | This study |
| RedEt3R | 5'-GGC TGC CCA TTT GGT TTG CGA ATA -3' | This study |
| QVESf | 5’-CCA GGT GGA GTT CGT CTA CCA YAA YGG NGC -3’ | This study |
| KWAAr | 5’-CAC CAC CTG CTC GGC NRC CCA YTT -3’ | This study |
| E3ART-F | 5’-ACC TTG CAA AAA GAA GGA ATT G-3’ | This study |
| E3ART-R | 5’-TTA TAT TTT TTT CGG ATG AGG AAA G-3’ | This study |

**Table 4-3. Prevelance of Et743 among different *Ecteinascidia* population.**
*Ecteinascidia* samples, time of collection, and description. The sign + or – indicates presence or absence, respectively of the expected amplicon or Et743; NT, not tested.

| Location | Time of Collection | Life Stage | Habitat | Detection of Et743 | | |
|---|---|---|---|---|---|---|
| | | | | MS | QAf/KWr | E3RAT-F/R |
| **Keys, Fl** | June 2008 | adult | mangroves | **+** | **+** | **+** |
| **Keys, Fl** | Feb 2007 | adult | mangroves | **+** | **+** | **+** |
| **HBOI, Fl** | October 2006 | adult | docks | **+** | **+** | **+** |
| **HBOI, Fl** | June 2007 | adult | docks | **+** | **+** | **+** |
| **Keys, Fl** | June 2008 | flatworm | mangroves | **NT** | **+** | **+** |
| **Keys, Fl** | June 2008 | larvae | mangroves | **NT** | **+** | **NT** |
| **Keys, Fl** | June 2007 | larvae | mangroves | **NT** | **+** | **NT** |
| **Fayed (F)** | May 2005 | adult | metal pilings | **-** | **-** | **NT** |
| **Fayed (F)** | Sept 2007 | adult | metal pilings | **-** | **-** | **+** |
| **SharmElsheikh (SES)** | June 2005 | adult | mangroves | **+** | **+** | **+** |
| **Fayed (F)** | June 2005 | adult | stolon | **NT** | **-** | **+** |
| **Wadi El-Ra'da (WR)** | 2006 | adult | mangroves | **+** | **+** | **+** |

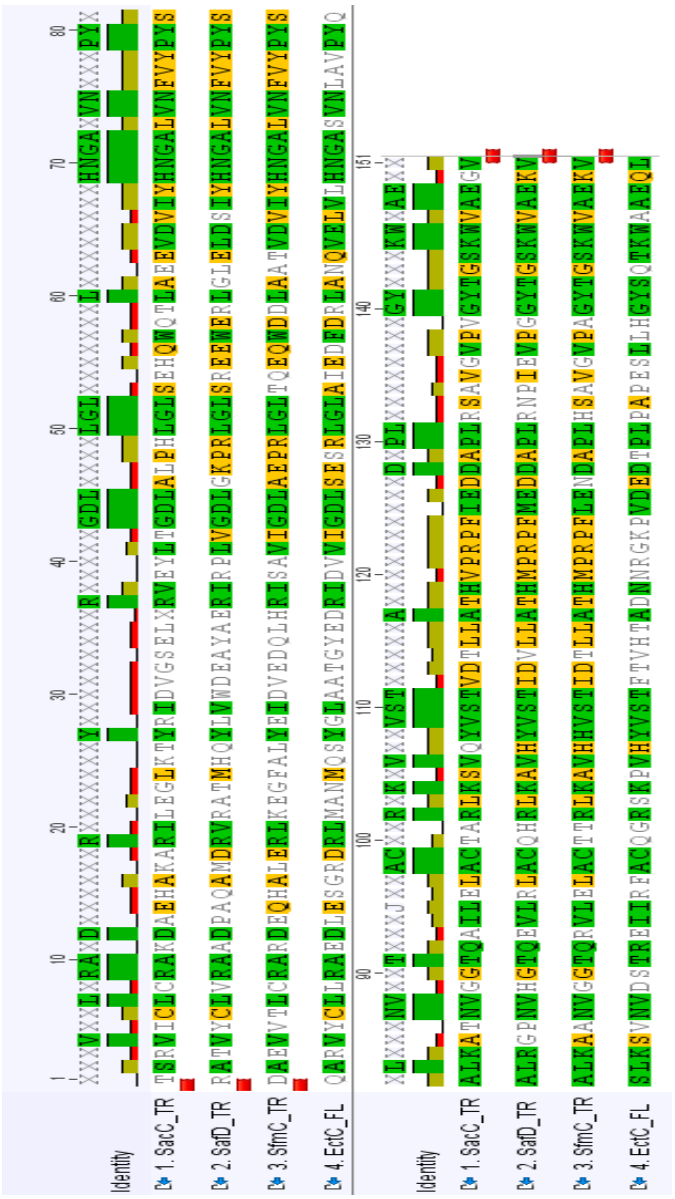



**Figure 4-6. Multiple sequence alignment of EctC_FL with SafD, SfmC, and SacC.**
FL, Florida. The figure shows highly conserved (green bar), conserved (brown bar) and less conserved (red bar) amino acids.

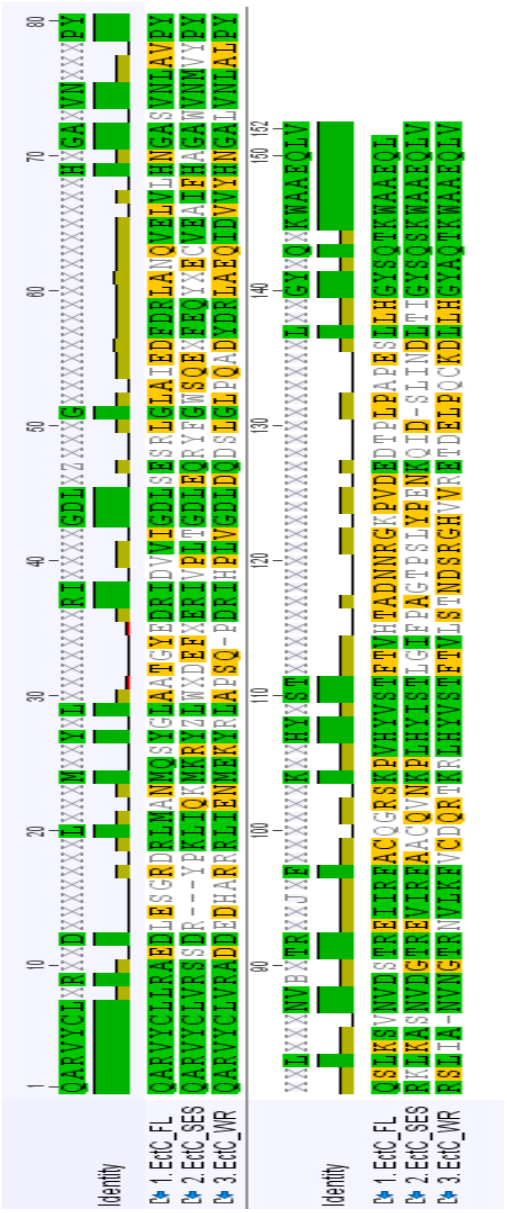



**Figure 4-7. Multiple sequence alignment of EctC_FL, EctC_SES, and EctC_WR.**
FL, Florida; SES, Sharm Elsheikh; WR, Wadi El-Ra'da The figure shows highly conserved (green bar), conserved (brown bar) and less conserved (red bar) amino acids.

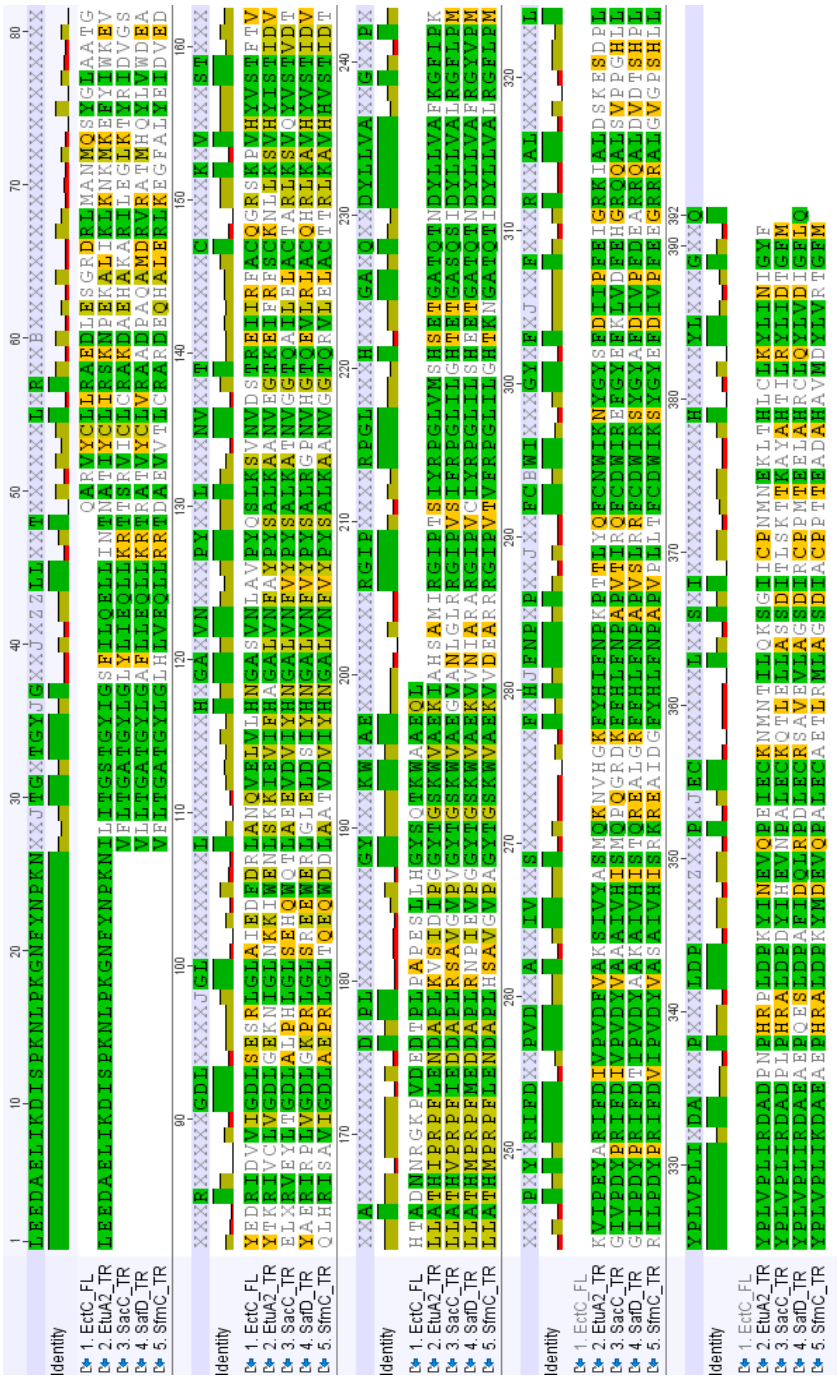



**Figure 4-8. Multiple sequence alignment of EctC-FL with EtuA2, SafD, SfmC, and SacC.**
The figure shows highly conserved (green bar), conserved (brown bar) and less conserved (red bar) amino acids.


**Figure 4-9. Alignment of the thiolation-reductase domain in EtuA2.**

The figure shows highly conserved (green bar), conserved (brown bar) and less conserved (red bar) amino acids. EtuA2_TR is published elsewhere (Rath et al., 2011) while EtuA2_HaygoodRed is the sequence obtained in our samples using specific primers to *etuA2*.

**Table 4-4. Identity/similarity percentage of amino acid sequences of the putative reductase domains from *Ecteinascidia* to those of the model compounds.**

| | SafD_TR | SfmC_TR | SacC_TR | EctC-FL | EctC-SES | EctC-WR | EtuA2_TR |
|---|---|---|---|---|---|---|---|
| SafD | X | X | X | X | X | X | X |
| SfmC | 64/77 % | X | X | X | X | X | X |
| SacC | 60/73 % | 68/81 % | X | X | X | X | X |
| EctC-FL | 42/62 % | 39/54 % | 36/57 % | X | X | X | X |
| EctC-SES | 39/59 % | 34/51 % | 34/52 % | 41/58 % | X | X | X |
| EctC-WR | 41/58 % | 44/57 % | 42/60 % | 54/73 % | 42/59 % | X | X |
| EtuA2 | 55/79 % | 51/71 % | 50/70 % | 36/59% | 40/62 % | 34/56 % | X |

### 4.3.4. Construction and screening of *Ecteinascidia* metagenomic library

The size of a library (number of clones) from a single genome required to ensure representation of a desired gene with high probability can be calculated accurately from the genome size of the organism and the average insert size in the library using the formula N = ln (1-P)/ ln (1-f), where N = number of colonies that need to be screened, P is the probability in decimal percentage, and f is fraction of the genome contained in a single average insert. In the case of a metagenomic sample with a eukaryote host of unknown genome size, containing numerous associated bacteria of unknown abundance and genome size, this accurate calculation is not possible. However, published successful metagenomic studies especially in sponges and soil typically screen several hundred thousand to a million fosmid clones (Chang and Brady, 2011; Fisch et al., 2009). In the case of *E. turbinata*, adult colonies are expected to have a complex metagenome with contributions from microbes associated with the food filtration apparatus and digestive tract.

Nevertheless, a rough estimate of the required size of the library was calculated. In order to get a 99% coverage of this metagenomic sample, and taking into account that the ascidian genome is around 160 Mbp based on the genome of the only ascidian genome sequenced; Ciona, then screening 19,000 clones should be enough to cover the ascidian genome. To account for the microbial community in the library and based on previous analysis of the microbial community of these populations (Elbaz, 2009), bacteria population was estimated. The samples do not seem to be rich with microorganisms as in the case with sponges. In fact, it was estimated that 20 different OTUs are present. Assuming that each microorganism has the same genome as *E. coli* genome which is

covered by 461 clones then screening about $25 \times 10^3$ clones should cover the metagenome sample with a 99% chance of coverage. This calculation is rough and might be underestimating the actual number of clones required to be screened as it is based on the genome of Ciona and the assumption that all OTUs are represented equally. A total of 40,000 clones were produced from four genomic libraries. The clones were screened by specific primers targeting *ectC* using liquid gel pool method (Hrvatin and Piel, 2007) and plate pool method (Piel, 2002). The first depends on using an ultra low melting point agarose that solidifies at low temperature. The nature of this solidified gel allows suspending and formation of mirocolonies to produce an unbiased library. This gel was previously used to amplify a cDNA library and successfully screen a metagenome library (Fisch et al., 2009).

One hit was obtained; SE-01. The fosmid was purified with copy induction solution. PCR amplification using primers targeting the A-T-R region produced a faint amplicon around the expected size, 2 kbp. Sequencing the ends of this fosmid was problematic. Only one end would give a sequence (the forward primer) but not the reverse. Efforts to sequence this fosmid using insert primers RedEt3F/R (primers specific for *ectC* sequence) failed. Restriction digestion analysis using the rare cutter NotI gave only one band instead of two; 7.5 kbp vector and 30-40 kbp insert (Figure 4-10). Restriction digestion using a more frequent cutter, BamH1 was expected to give an amplicon with 8.1 kbp in addition to other bands from the insert. However this was not the case, which suggested that there might be some kind of rearrangements involving the reverse end of the insert. Growing the plasmid without the addition of the copy induction solution also did not improve digestion suggesting that this rearrangement if true still exists.

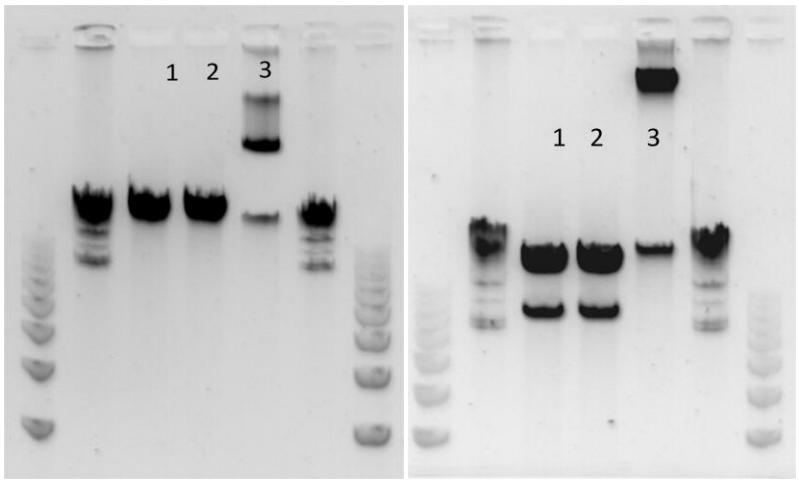



**Figure 4-10. Restriction digestion analysis of SE01.**
Digestion was performed using (a) NotI and (b) BamHI enzymes. 1 refers to incubation for 1 hour with the enzyme, 2 for 3 hours with the enzyme, 3 in the absence of enzymes. 1 kb and λ DNA standards were used to estimate the size of digested bands.

## 4.4. Discussion

Ecteinascidins seem to be structurally related to other tetrahydroisoquinoline compounds such as renieramycins and jorumycin originally isolated from different sponges (Amnuoypol et al., 2004; Suwanborirux et al., 2003; Frincke and Faulkner, 1982) and the nudibranch *Jorunna funebris* (Fontana et al., 2000), respectively (Figure 4-1). In addition, ecteinascidins seem to be structurally related to three other bacterial metabolites; saframycin A, produced by the actinobacterium *Streptomyces lavendulae* (Arai et al., 1980), saframycin MX1 isolated from the deltaproteobacterium *Myxococcus Xanthus* (Irschik et al., 1988), and safracin B produced by the gammproteobacterium *Pseudomonas fluorescens* (Ikeda et al., 1983). In addition, the gene clusters responsible for the biosynthesis of safracin (Velasco et al., 2005), saframycin A (Li et al., 2008), and saframycin MX1 (Pospiech et al., 1995) are also reported. The structural resemblance of compounds isolated originally from marine eukaryotes and bacterial metabolites strongly suggests that the actual producers are microorganisms living in a symbiotic relationship in these marine animals. Recent publications have provided evidence for this hypothesis. Bryostatins were originally isolated from the bryozoan *B. neritinia* but were later found to be produced by the gammaproteobacterium *Candidatus* Endobugula sertula (Davidson and Haygood, 1999; Davidson et al., 2001). The macrocyclic ribosomal peptides, patellamides isolated from *L. patella* were found to be produced by the cyanobacterium symbiont *P. didemni* using two different approaches (Long et al., 2005; Schmidt et al., 2005). The possibility that the tetrahydroisoquinolines and specifically the ecteinascidins are produced by bacteria and that the biosynthetic gene cluster might be present in a bacterial genome increases the possibility towards identification of the whole pathway.

This is attributed to the fact that genes performing a certain function tend to be clustered together in microbial genomes. Most of these symbionts have yet to be cultured. However, it might be possible to identify the biosynthetic gene cluster responsible for the production of these compounds through identifying unique genes within the pathway. Identification of the biosynthetic pathway for ecteinascidins and specifically Et743 might allow the expression of this pathway in a heterologous host. This will create a cost-effective and a stable source for this drug. In addition, the understanding of the biosynthetic machinery will facilitate the production of combinatorial libraries of derivatives of this compound with stronger activity and less toxicity that target other types of cancer.

By studying the biosynthetic gene clusters and mechanisms of the bacterial tetrahydoisoquionline metabolites, it might be possible to predict the gene cluster architecture of Et743 and the mechanism of its biosynthesis (Figure 4-11). The structure of ecteinascidins seems to be formed out of four amino acids; alanine, glycine and two tyrosine derivatives. This would typically require four NRPS modules. However, only three modules are present in the biosynthetic gene cluster of the three compounds. There is a dispute about which modules load which amino acids in the safracin (*sac*) cluster (Velasco et al., 2005) and saframycin A (*sfm*) cluster (Li et al., 2008). Bioinformatic analysis of the core residues of the adenylation domains (Stachelhaus et al., 1999; Challis et al., 2000) of SacA, SafB, and SfmA suggests that the first module loads alanine not glycine (Table 4-5). Thus, in all three clusters there is only one tyrosine class module where two would be expected. This suggests that the third module in all three cases acts through a stuttering mechanism to condense the tyrosine derivative twice (Figure 4-11).

The three gene clusters *sac*, *sfm*, and *saf* contain accessory genes that code for enzymes that produce compounds bearing distinctly different tyrosine derivatives, 3-hydroxy-5-methyl-O-methyltyrosine (4-O-methyl-5-methyl-L-dopa) in the case of safracin and saframycin A, and 4-O-methyl-L-dopa in the case of saframycin MX1 (Table 4-5). It was previously shown also that L-dopa is converted to a 4-*O*-methyl derivative prior to loading by the NRPS by SafC (Nelson et al., 2007). Since there is one module that acts by a stuttering mechanism to load the amino acids 3 and 4, it is assumed that the same derivative is loaded. In all model pathways, the NRPS is followed by an ORF linked to biosynthesis of the amino acid loaded as amino acids 3 and 4, SafC, SfmD and SacD.

Our assumption for the biosynthesis of ecteinascidins suggested that the gene cluster would be formed of three modules of NRPS that we designated *ectA*, *ectB*, and *ectC* each containing the basic three domains; condensation, adenylation and thiolation domains. Alanine would be loaded by EctA, glycine by EctB, and a tyrosine derivative by EctC. It is possible that 4-O-methyl-5-methyl-L-dopa, 4-O-methyl-L-dopa, or even another tyrosine derivative could be loaded by EctC in the ecteinascidin pathway. A characteristic feature of the three pathways of saframycin A, saframycin MX1, and safracin is the presence of a reductase domain R (Figure 4-5) that is responsible for undergoing the Pictet-Spingler cyclization before the release of these compounds. Thus, it is expected that the *ect* pathway contains the same domain. The Sherman group hypothesized a different mechanism based on the gene cluster they identified (Rath et al., 2011). The first NRPS module EtuA3 would load cysteine based on the amino acid specificity motif (Table 4-5). EtuA1 was hypothesized that it loads glycolic acid while EtuA2, based on

the specificity motif of the A domain would load a tyrosine derivative in an iterative manner.

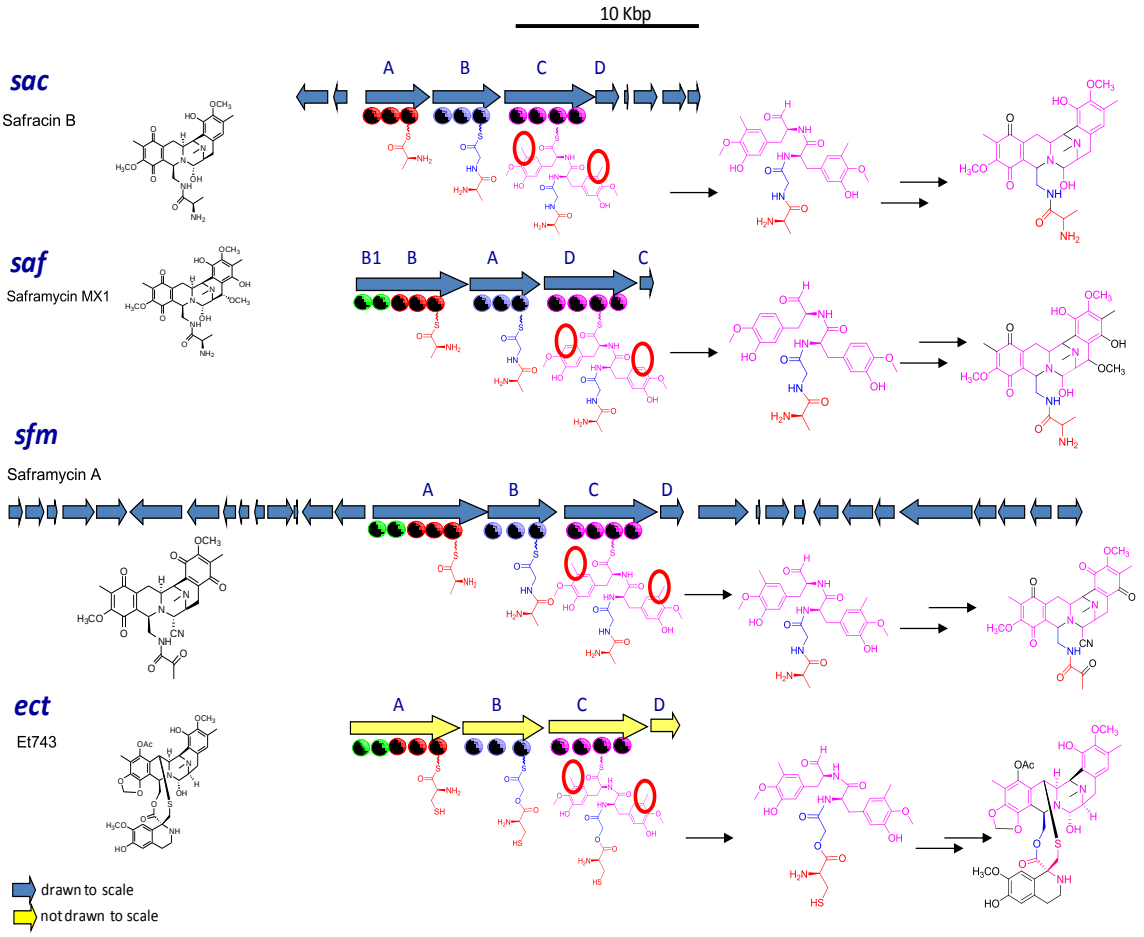



**Figure 4-11. Gene clusters involved in the biosynthesis of the ecteinascidin family compounds and intermediates formed and predicted.**
Black indicates modifications by tailoring enzymes.

**Table 4-5. Conserved residues and predicted loading substrates of the modules of the ecteinascidins-similar compounds.**

| Protein | Compound/ Organism | Conserved residues | Domains | Substrate |
|---|---|---|---|---|
| **SafD** | Saframycin MX1 *M. xanthus* | D P W G L G L I D K | C-A-T-R | 3-hydroxy,4-methyl-Phe/SafD, MX1 |
| **SacC** | Safracin B *P. fluorescens* | D P W G L G L I D K | C-A-T-R | 3-hydroxy,4-methyl-Phe/SafD, MX1 |
| **SfmC** | Saframycin A *S. lavendulae* | D P W G L G L I D K | C-A-T-R | 3-hydroxy,4-methyl-Phe/SafD, MX1 |
| **EtuA3** | Et743 *E. turbinata* | D L Y N L S L I W K | C-A-T | Cysteine |
| **EtuA1** | Et743 *E. turbinata* | ------- | C-A-T | glycolic acid |
| **EtuA2** | Et743 *E. turbinata* | D P W G L G L I D K | C-A-T-R | 3-hydroxy,4-O-methyl-5-methyl-tyrosine |

Targeting this predicted reductase domain using a nested PCR approach increased the chance that the amplicon is a specific reductase NRPS product. Although the reductase domain in *M. Xanthus* has a 70% GC content, the fragment specifying the putative Et reductase domain, *ectC* is 55.5%, while *etuA2* is 27.4%. The GC percentage of *ectC* is distinct from genes that function in the pathways of the model compounds and more typical of most gammaproteobacteria, consistent with a symbiont such as *Candidatus* E. frumentensis. The GC content of the *ectC* fragment is likely to be similar to that of the symbiont genome as a whole.

The fact that EctC is homologous to the three model pathways was found present in all our Florida *Ecteinascida* populations in addition to two locations of the Red Sea strongly suggested that this domain is part of the actual biosynthetic pathway of tetrahydroisoquinolines and possibly Et743. The high frequency with which this *ectC* was

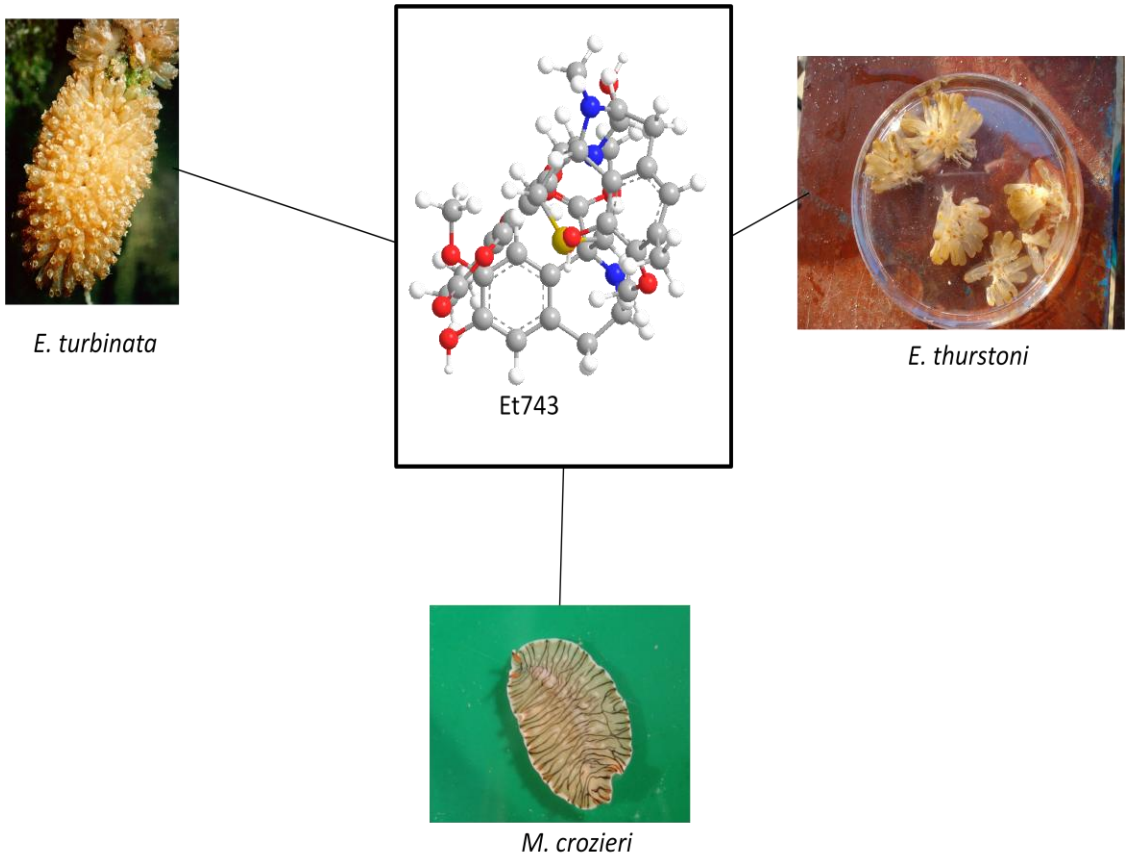
recovered, and its presence in all Florida sample types from adults, larvae and the predator flatworm (Figure 4-12), suggested that it does not originate in a casually associated environmental bacterium, but a tightly linked symbiont.

The thiolation-reductase domain in EtuA2 reported recently (Rath et al., 2011) was found in all *Ecteinascidia* populations tested (Table 4-3). This sequence showed higher similarity to each of SafD, SfmC, and SacC than EctC obtained in this study (Figure 4-8; Figure 4-9; Table 4-4). In addition, the compound Et743 was detected in all samples from the Florida collections as evidenced by the MS analysis. In addition, two out of three locations in the Red Sea showed the presence of this compound as detected by MS and the presence of a closely related reductase gene fragment *ectC*. However, the Fayed population did not show the presence of the characteristic MS peaks of Et743. Since *etuA2* was present in *Ecteinascidia* samples from the five locations tested in this study and showed higher similarity to model pathways, there is a higher possibility that EtuA2 is actually involved in the biosynthesis of ecteinascidins. It is possible that the production of Et743 in the Fayed population is not abundant as that in the other sites leading to low amounts that are not detected by MS. It is also possible that the production of Et743 in this area is affected by complex environmental conditions. Another possibility is that Et743 is not produced in every colony in the Fayed population. This does not completely exclude the fact, although less likely, that EctC is involved in the biosynthesis of Et743 not EtuA2. In this case, it might be possible that the Fayed population lack the symbiont as in the case with some aposymbiotic populations of *B. neritina* (McGovern and Hellberg, 2003).

The presence of both *ectC* and *etuA2*, in addition to Et743 in the tiger flatworm was surprising. No reports regarding the presence of ecteinascidins in *Maritigrella crozier* exist. The presence of the compound in these predators might be due to the direct feeding on the tunicate. Another possibility is that the flatworm harbors the symbiont that biosynthesizes ecteinascidins.

Chemical variability has been reported before and is considered one of the main problems in marine natural products (Schmidt et al., 2012). The two terpenoids, pukalide and 11β-acetocypukalide that have been shown to act as signals for larval settlement and gametogenesis (Slattery et al., 1999) have been reported to show temporal and spatial differences in the same species (Slattery et al., 2001). The yet uncultured cyanobacterium *P. didemni* is a symbiont of many marine ascidians and is considered the source of several cyanobactins such as patellamides and ulithiacyclamide (Schmidt et al., 2005). Different strains of *P. didemni* are capable of producing different compounds. The fact that Et743 was absent in the Fayed populations might suggest different ecological conditions, different predation pressure, or temporal variation. Since the biosynthesis of many of the natural products have a bioenergetic cost, it is expected that release of the predation pressure could lead to loss of the genes or the producer symbiont.

A.




B.

**Figure 4-12. Presence of Et743 in *Ecteinascidia* spp. and the flatworm.**
A) Et743 can be found in different samples including *E. turbinata*, *E. thurstoni*, and the tiger flatworm. B) The tiger flatworm *M. crozieri* feeding on Et743 inspite of the presence of feeding deterrent compounds.

The non filter feeding larvae are lecithotrophic and are expected to be less complex than the filter feeding colonies. Tiny amounts of larvae were available hindering genomic library construction from them. However, since these larvae do not filter feed, the metagenome DNA extracted from these samples are less complex than the tunicate adult. Having a candidate gene that possibly belongs to the pathway encouraged us to construct and screen a metagenomic library in order to identify the whole Et743 biosynthetic pathway. The challenge was to obtain adequate numbers of clones. Optimization of DNA extraction is an important task, because there is usually a trade-off between greater purity (less inhibition of enzymes used in library construction) and greater size: more manipulation usually leads to more shearing and reduced size. DNA from *E. turbinata* was typically refractory to enzymatic reactions, possibly due to the presence of Et743 itself, or other contaminants. We succeeded in obtaining high molecular weight DNA of high purity suitable for downstream applications. Screening of the library using primers specific for *ectC* led to a potential hit SE01. However, various efforts to sequence the fosmid ends, in addition to restriction analysis failed. This could be due to rearrangements reactions that are occurring if the sequence is unstable for cloning. Another possibility is that this is a false positive and does not represent any portion of the pathway.

By comparing the reductase gene obtained in our lab with that obtained in the Sherman group, it was found to be similar but not identical (Figure 4-8; Table 4-4). This group provided evidence including a similar relative synonymous codon usage (RSCU) between the contig carrying the 16S rRNA of *Candidatus* E. frumentensis and that with the pathway, metaproteome analysis of the enzymes predicted to be present from the

pathway, in addition to the expression of the reductase domain and confirming that it cyclizes expected substrates. The GC percentage of the pathway was found to be 27.4%. The approach taken in their lab was metagenome sequencing. However, the approach taken in our study was designation of degenerate primers and amplifying using PCR in different conditions, in addition to cloning. By analyzing the degenerate primers designed in this and the sequence of *etuA2*, it is clear that these primers would not amplify this fragment. In addition, it is possible that the high AT% would cause rearrangements.

In general, using chemical and genetic evidence, we were able to show that the anticancer compound Et743 is present in the Carribean ascidian *E. turbinata* in Florida, its larvae, in addition to its predator tiger flatworm (Figure 4-12). In addition, the compound was also found in the Red Sea ascidian *E. thurstoni* in two locations but was not confirmed from the Fayed populations. We also identified a potential gene fragment *ectC*, which might be involved in the biosynthesis of tetrahydroisoquinolines. In addition, we surveyed the same samples with primers specific to *etuA2* reported previously. Primers specific to any of these gene fragments could be used to probe the metagenomic library constructed in this study.

## 5. Summary and Conclusion

Marine natural products possess interesting chemical structures and have promising pharmacological activities. These secondary metabolites often play a major role in the chemical defense of the host or sometimes act as signal molecules within the host-microbiome community. Many compounds that were originally thought to be produced by marine macroorganisms are actually produced by bacterial endosymbionts. These compounds are biosynthesized by proteins encoded by genes that are clustered closely together in the bacterial genome. Identification and manipulation of the gene cluster could lead to expression of these compounds in higher yields. Studying the structure of a compound could predict the architecture of the genes involved. If a gene cluster is present with no known compound related to it (orphan cluster), it might be possible to predict the chemical structure of the compound or at the very least its class. Both approaches are important and in some cases proved successful.

Bioinformatic analysis of four bacterial genomes and one metagenome was performed for the presence of potential genes involved in the biogenesis of secondary metabolites (**chapter 2**). I investigated the genomes of four shipworm symbionts. The first was *T. turnerae* T7901 isolated from the shipworm *B. gouldi*, in addition to three other symbionts from *B. setacea*; BS02, BS08, and BS12. In addition, I analyzed the metagenome of the *B. setacea* shipworm. They all showed strong evidence to be rich sources for interesting compounds including polyketides and nonribosomal peptides.

I purified and isolated two tartrolons that turned out to be macrolides dimers formed by a polyketide synthase pathway (**chapter 3**). One of these compounds, tartrolon E possesses antibacterial activity against different pathogens including *P.* aeruginosa, *S. aureus*, and

methicillin resistant *S. aureus*. I was able to identify the biosynthetic gene cluster responsible for their formation. The modules and domains in this gene cluster were analyzed in depth and the biosynthetic scheme was predicted. This paves the road for construction of combinatorial biosynthetic libraries from this active compound and other structurally related compounds. In addition, I detected these compounds in the host tissue, suggesting that the compounds are not only produced under laboratory conditions but also expressed *in vivo* suggesting an important role in the microbial-host symbiosis. These compounds were also detected in other *Teredinibacter turnerae* strains isolated from different host shipworms suggesting a conserved essential function. Besides acting as a bacterial inhibitor, tartrolon chemical structure suggests an interesting mechanism to regulate the levels of boron in the shipworm-microbial community. Tartrolon might also be involved in providing boron in order to increase the strength of vesicles where nitrogen fixation occurs providing the anaerobic environment required by nitrogenase.

I surveyed different populations of the ascidian *E. turbinata* and *E. thurstoni* for the presence of an anticancer compound Et743 (**chapter 4**) thought to be produced by a gammaproteobacterium. In addition, I predicted a biosynthetic scheme for these compounds based on that from other bacterial structurally related compounds. I was able to obtain a sequence that is similar to reductase domains that might be involved in the biosynthesis of tetrahydroisoquinoline compounds. In addition, I screened different *Ecteinascidia* samples for a fragment recently reported by another group. Both fragments are present in the majority of the samples. I also constructed a genomic library which can be probed using specific sequences.

The genomic and chemical data in this dissertation suggest that gammaproteobacterial symbionts of marine invertebrates can present a rich source of bioactive secondary metabolites shifting the long-time attention from actinobacteria to this class of microorganisms.

# A1. Bioinformatic Analysis of the ATP-Binding Cassette (ABC) Transporters in the Metagenomes of *Prochloron* Containing Ascidians

## A1.1. Introduction

Many obligate symbionts associated with plants and animals are thought to perform essential primary and secondary metabolic functions for the host. One of these symbionts is the cyanobacterium *Prochloron didemni*, an as yet uncultivated obligate symbiont of marine ascidians of the family Didemnidae. Zooids of didemnid ascidians are filter feeders. *P. didemni* is different from other cyanobacteria and shares a few features with plants. First, it uses both chlorophyll a and b in photosynthesis. This allows it to capture as much energy as possible; which is ideal for the shallow water habitat of its host. Another aspect is that it contains thylakoids, and finally it lacks the light capturing molecules, phycobilins. It is estimated that *P. didemni* cells provide their host with most of the carbon assimilated by the host and participates in nitrogen recycling (Isao et al., 1996; Koike et al., 1993). In addition, *P. didemni* was shown to produce an array of cyanobactins (Schmidt et al., 2005) that might be involved in protection of the soft bodied hosts against predators. Thus, *P. didemni* is considered a major contributor in the ascidian symbiosis.

Studying the transport system for *P. didemni* might shed some light on the type of compounds and metals transported across the cell membrane and thus, the chemical interaction between the host and symbionts, in addition to that among symbionts. Transport systems in organisms play an important role as translocation machinery. They are involved in the transport of nutrients, toxins, and wastes in and out of the cell. One of the most important transporter classes is the energy coupled ATP-binding cassette (ABC)

transporters. These proteins couple the hydrolysis of ATP to the transportation process. They represent one of the largest families of proteins in the cell.

I decided to study the distribution of ABC transporters system in the three genomes of *P. didemni* since it represents one of the largest superfamilies of proteins and is involved in an important process. ABC systems of various prokaryotes (Tomii and Kanehisa, 1998; Linton and Higgins, 1998; Quentin et al., 1999; Braibant et al., 2000; Harland et al., 2007) and eukaryotes (Ardelli et al., 2010; Jasinski et al., 2003; Morris and Phuntumart, 2009; Zhao et al., 2007) have been analyzed. The nucleotide binding domain (NBD) of the ABC transporters contains essential motifs that are conserved among the three domains of life (Davidson and Chen, 2004; Davidson et al., 2008; Higgins, 1992; Saurin et al., 1999) facilitating their detection.

## A1.2. Materials and Methods

### A1.2.1. Sample collection and sequencing

The three *L. patella* ascidians samples, L1, L2, and L3 were collected by the Schmidt group as follows; L1: from Palau at coordinates N 7º15' E 134º15'; L2: from Fiji at coordinates S 17º55' E 177º16'; L3: from Solomon at coordinates S 8°57.35' E 159°59.12' (Figure A1-1). All details regarding the sequencing and annotation of the three genomes were reported elsewhere (Donia et al., 2011). Genomes obtained from L1, L2, and L3 were designated P1, P2, and P3, respectively. P1 was sequenced using Sanger method while P2 and P3 were sequenced with 454 and using P1 as a reference.

### A1.2.2. Bioinformatic analysis

Open reading frames (ORFs) encoding ATP binding domain sequences that were automatically annotated in the three genome database were collected using suitable keywords. Any protein that requires ATP will have the signatures Walker A and Walker B while the ABC superfamily will have the conserved ABC signature motif. Thus, manual investigation of each sequence was performed to confirm the presence of conserved motifs: Walker A (GxxGxGKST), ABC signature motif (LSGGQxQR), and Walker B (hhhhDE) (Figure A1-2). Adjacent genes were analyzed for possible membrane spanning domains and substrate binding domains. Each sequence was subjected to local BLAST analysis against the P1 database to retrieve any additional sequences. Each sequence identified was inserted into a local database using Geneious (Drummond et al., 2009) and subjected to a local BLAST analysis against the other genomes database to retrieve any additional sequences. In addition, each genome was searched for ABC characterized motifs and added to the inventory. Sequences not containing conserved

motifs within adequate distances were excluded. Each sequence was subjected to BLAST analysis on the protein database on the NCBI web server to confirm the presence of these motifs and to predict the possible substrate. In addition, each sequence was also analyzed by the transport classification database (Saier et al., 2006) on the UCSD server and the transport database (Ren et al., 2004) at the J. Craig Venter Institute (JCVI), (Currently at Macquarie University) to identify transmembrane regions and confirm substrate specificity and function. For the purpose of phylogenetic analysis, sequences of the nucleotide binding domains (NBDs) from other bacterial species were retrieved from the transport database and were also confirmed to contain the specific conserved motifs at the appropriate distance. All sequences were added to the inventory with the three *P. didemni* genomes. All sequences were aligned using ClustalW and phylogenetic analysis was performed using Mega4 (Tamura et al., 2007).

## A1.2. Results

The total number of genes involved in transport processes in the genome of *P. didemni* P1 was predicted to be a total of 282 representing about 3.8% of the total number of open reading frames.

In total, 41 NBDs that belong to the ABC transport systems were found in our analysis of the three genomes; these are predicted to form 40 different ABC transport systems (Table A1-1). Surprisingly, *P. didemni* had lower number of predicted ABC transporters when compared with the genomic distribution of these systems of other prokaryotes (Figure A1-4). Phylogenetic analysis (Figure A1-5) classified these NBDs according to their substrate specificity and transporter function. In general, domains with similar predicted specificity were clustered together.

Studying common transporters could illuminate the symbiotic relationship between the symbiont and the host. Studying the different transporters could reveal some information about the different habitats. Comparative analysis of the three genomes revealed that P1 had 37 ABC transport systems; P2 had 35, while P3 had 36. Most ABC transporters were conserved among the three genomes with percentage amino acid similarity higher than 97%. Most sequences were identical, however at least three transporters present in P1 were absent in either P2 or P3. P1 had three systems predicted to be related to drug resistance, one of which might be involved in the transport of a PKS/NRPS compound but absent in either P2 or P3. In addition one system involved in the uptake of iron-siderophore complex was present in P2 but not in the two other genomes. Finally, another system also related to drug resistance was present only in P3.

The abundance of transporters in the three genomes was analyzed and compared to other obligate and free living symbionts (Figure A1-3). The total number of all types of transporters in *P. didemni* is calculated as 46.2 per Mbp of the *P. didemni* genome. This is a higher number than other symbionts as Wolbachia (26.7 per Mbp genome) the symbiont of *Brugia malayi*, and lower than other free living bacteria such as *Bacillus subtilis* (69.05) *Aquifex aeolicus* (39.3), *E. coli* (76.3), *Nostoc* sp. (43.2), *Synechococcus* sp. (36.6), *P. marinus* MIT9313 (37.1), *P. putida* (62.3), and *W. brevipalpis* P-endosymbiont (35.7).

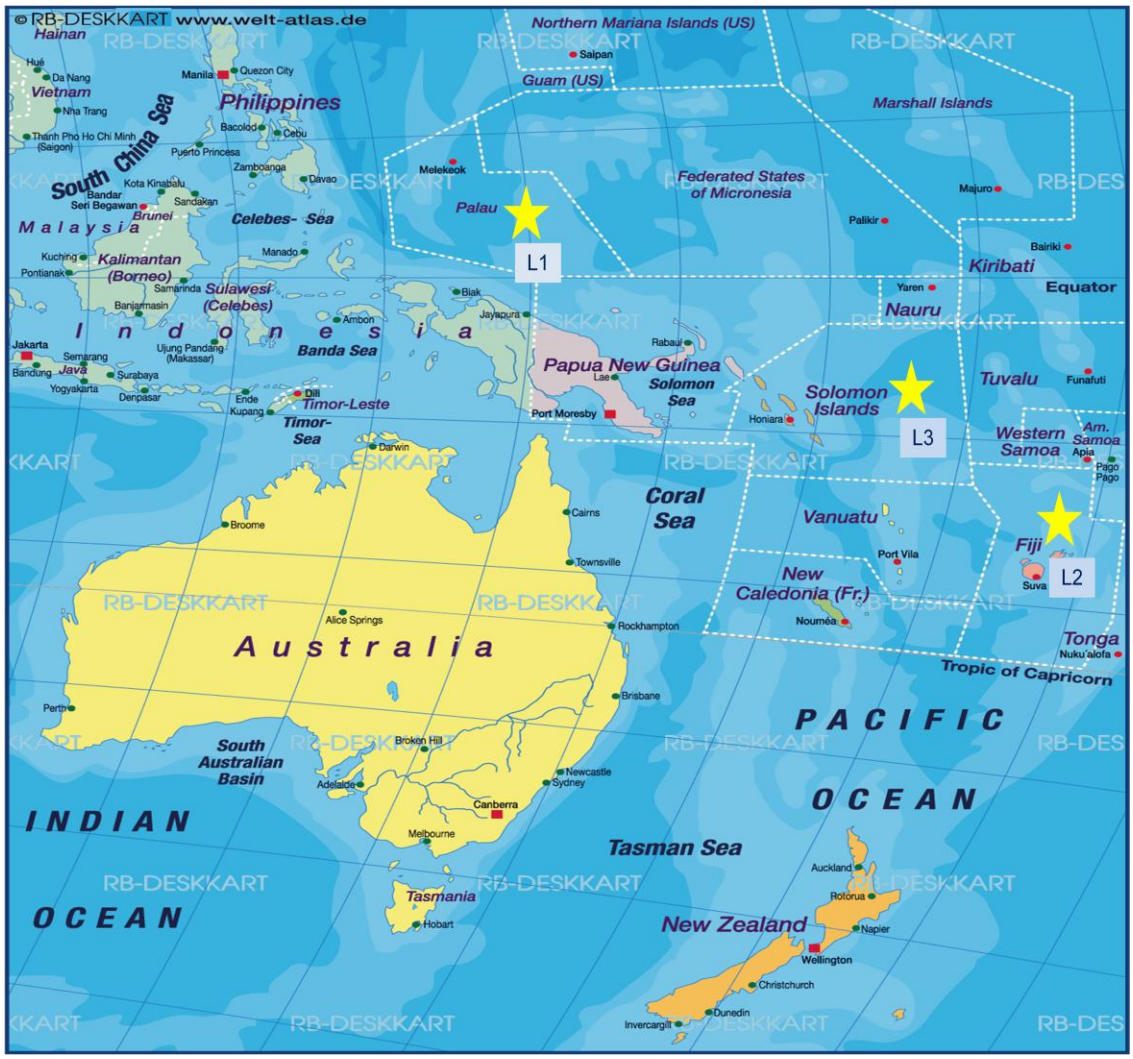



**Figure A1-1. A map showing locations of samples sequenced.**
Yellow stars indicate collection sites.

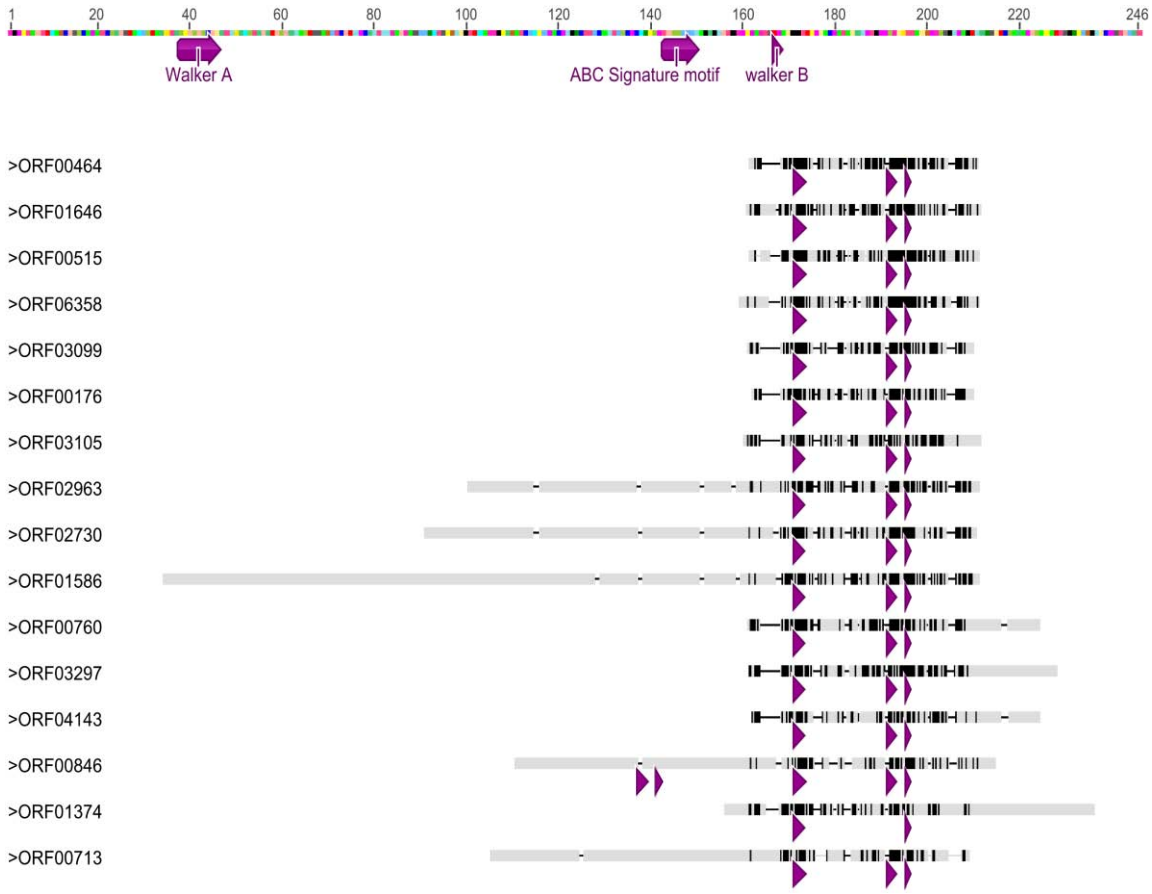



**Figure A1-2. Conserved motifs in ATP binding cassette transporters.**
A representative sequence of an ATP binding domain from P1 shows conserved motifs (top) and a representative alignment of sequences from *P. didemni* genomes showing aligned conserved motifs (bottom).

**Table A1-1. Summary of the predicted NBDs of ABC transporters found in the three sequenced *Prochloron* genomes.**

Highlighted rows indicate differences.

| ORF | aa | P1 | P2 | P3 | Predicted Substrate/Function |
|---|---|---|---|---|---|
| 6561_P2 | 870 | Absent | Present | Absent | Drug export/MDR |
| 8946_P2 | 261 | Absent | Present | Present | iron-siderophore complex |
| 4216_P3 | 537 | Absent | Absent | Present | Cobalt / Drug resistance |
| ORF00176 | 233 | Present | Present | Present | Urea |
| ORF00310 | 260 | Present | Present | Present | ferric/ iron siderophore complex |
| ORF00416 | 249 | Present | Present | Present | Drug export/MDR |
| ORF00464 | 246 | Present | Present | Present | Polar amino acid transport |
| ORF00515 | 267 | Present | Present | Present | Phosphate/Phosphonate Uptake |
| ORF0713 | 567 | Present | Present | Present | Drug export/MDR |
| ORF00760 | 316 | Present | Present | Present | Drug export/MDR |
| ORF00846 | 566 | Present | Present | Present | Cobalt /Drug resistance |
| ORF01361 | 245 | Present | Present | Present | Manganese |
| ORF01374 | 427 | Present | Absent | Present | polysaccharide export/bacitracin resistance |
| ORF01395 | 250 | Present | Present | Present | hydrophobic amino acid |
| ORF01416 | 231 | Present | Present | Present | DevA (heterocyst formation) Glycolipid exporter |
| ORF01578 | 213 | Present | Present | Present | Cobalt or drug transport |
| ORF01586 | 1011 | Present | Present | Present | Drug export/MDR |
| ORF01639 | 265 | Present | Present | Present | Phosphonate import |
| ORF01646 | 275 | Present | Present | Present | Glycine betaine uptake |
| ORF01746 | 360 | Present | Present | Present | Fe (III)/import |
| ORF02159 | 796 | Present | Present | Present | Drug export/MDR |
| ORF02681 | 653 | Present | Present | Present | Molybdate |
| ORF02692 | 322 | Present | Present | Present | Drug export/MDR |
| ORF02730 | 667 | Present | Absent | Absent | Drug export/MDR |
| ORF02963 | 618 | Present | Present | Present | Drug export/MDR |
| ORF03099 | 225 | Present | Present | Present | Cobalt/Cobalamine |
| ORF03105 | 264 | Present | Present | Present | Metal uptake (Zn/Fe) |
| ORF03297 | 341 | Present | Present | Present | Drug export/MDR |
| ORF03566 | 259 | Present | Present | Present | Polar amino acid or cobalt transport |
| ORF03789 | 605 | Present | Present | Present | Drug export/MDR |
| ORF03871 | 909 | Present | Absent | Absent | Drug export/MDR |
| ORF04060 | 581 | Present | Present | Present | Drug export/MDR |
| ORF04143 | 306 | Present | Present | Present | Drug export/MDR |
| ORF06092 | 578 | Present | Present | Present | Drug export/MDR |
| ORF06222 | 255 | Present | Present | Present | Nitrate/sulfonate |
| ORF06358 | 280 | Present | Present | Present | Phosphate import |
| ORF06376 | 260 | Present | Present | Present | Organic solvent export |
| ORF06501 | 612 | Present | Present | Present | Drug export/MDR |
| ORF06858 | 974 | Present | Present | Present | Drug export/MDR |
| ORF06860 | 875 | Present | Absent | Absent | Drug export/MDR |

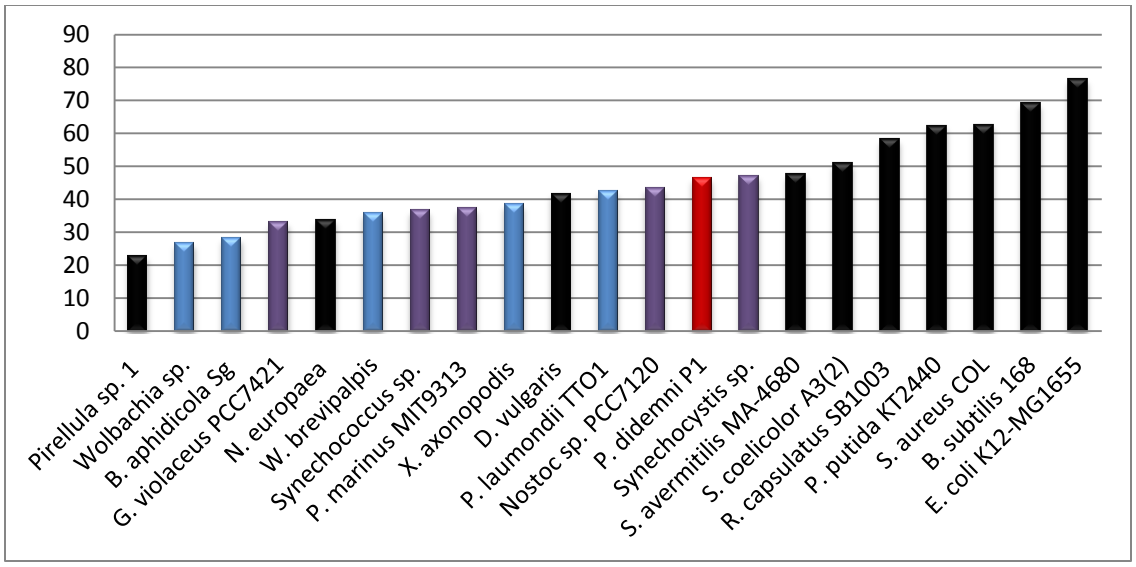



**Figure A1-3. Comparison of the number of transporters per Mbp genome of *P. didemni* and other organisms.**

Red, *P. didemni* (P1); blue, symbionts; purple, cyanobacteria; black: other bacteria.

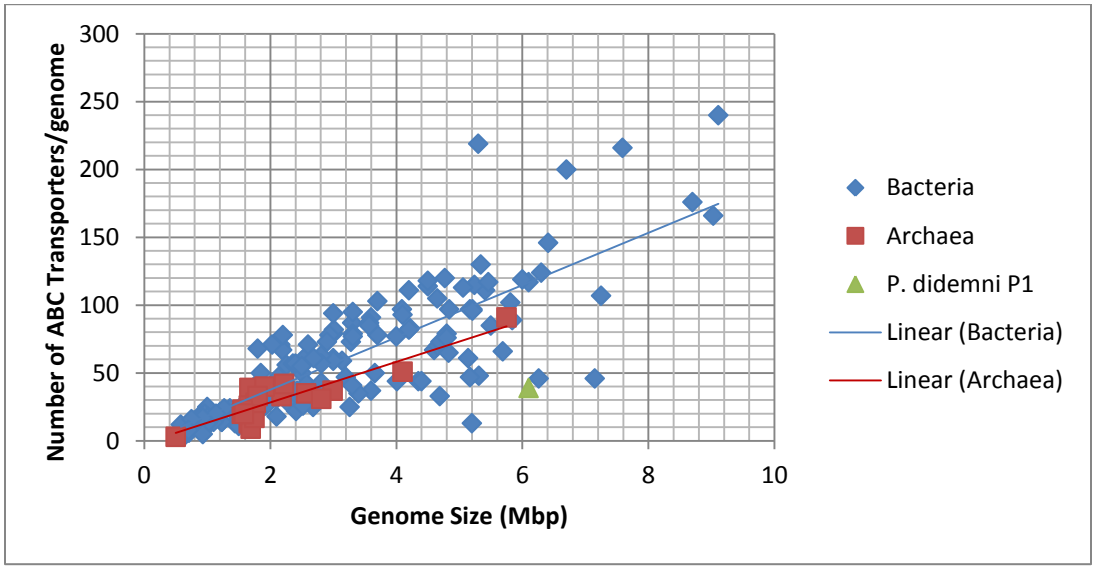



**Figure A1-4. ABC system distribution among prokaryotes.**

Green triangle, *P didemni*; blue diamonds, other bacteria; red squares, archaea. Data from all organisms except P1 was obtained from transport database at www.membranetransport.org.

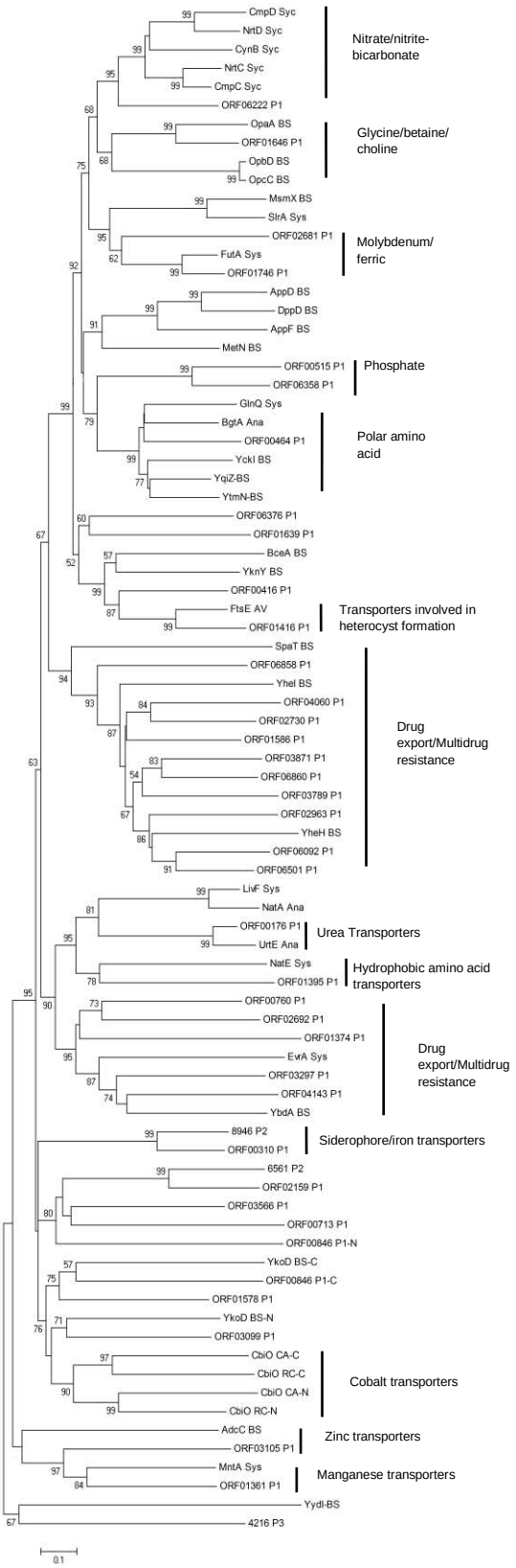



**Figure A1-5. Unrooted Neighbor-Joining phylogenetic tree of the NBDs of the ATP transporters from *Prochloron* and other organisms.**

The scale bar represents 0.1 fixed substitution per amino acid residue. The numbers on the branch represent the percentage of the bootstrap support of 1000 replicates that the interior branch length is greater than zero. Only bootstrap values higher than 50 are shown. Protein name is followed by the name of the organism (Ana, *Anabaena* species; AV, *Anabaena variabilis*; BS, *Bacillus subtilis*; CA, *Chloroflexus aurantiacus*; P1/P2/P3, *P. didemni* P1/P2/P2 respectively; RC, *Roseiflexus castenholzii*; Syc, *Synechococcus* sp.; Sys, *Synechocystis* sp). When the sequence contained two NBDs, an N and C was added to the name to indicate first and second domain, respectively.

## A1.3. Discussion

Bacterial cells use transport systems to communicate with their external environment. Obligate symbionts are subjected to reduction in their genome size and content over time and become dependent on their host (Perez-Brocal et al., 2006). Thus transport systems represent important machinery for the symbiotic relationship through the uptake of essential nutrients and ions and the secretion of chemical defense molecules. When the total number of transporters in *P. didemni* was compared to that of other bacteria, it was found to be higher than other symbionts and most cyanobacteria but lower than heterotrophs and other free living bacteria (Figure A1-3). This might reflect the flexible nature of *P. didemni* versus other obligate symbionts more adapted to a specific environmental niche enabling it to survive in different hosts and conditions. This is consistent with its relatively large genome and the presence of a complete set of metabolic genes in the *Prochloron* genome (Donia et al., 2011).

ORFs involved in the transport process were predicted to be 46.2/Mbp of genome, higher than most known symbionts and cyanobacteria. The number of predicted ABC transport systems seemed to deviate from the proportionality with genome size (Figure A1-4) seen with most other organisms (Davidson et al., 2008; Harland et al., 2005). Transport systems specific to essential minerals and amino acids were found across the three genomes (Figure A1-5; Table A1-1). ORFs predicted to encode urea transporters were also found as consistent with other cyanobacteria (Valladares et al., 2002). Differences detected in the ABC transporters among the three genomes were mainly dedicated to the export of toxins and biosynthesized metabolites such as nikkomycin-like compounds and the unknown nonribosomal peptide product of the *prn* cluster (Schmidt et al., 2004). The

fact that the differences among the genomes exist in the regions encoding secondary metabolites in addition to ABC transporters might be related to the differences in the competitive environments, predation pressure or specific to the different compounds produced by each. Strikingly, almost half of these systems were dedicated to multidrug resistance or to the efflux of peptides or virulence factors. This might be related to the capability of *P. didemni* to biosynthesize different secondary metabolites. This could also reflect the high competition among different bacteria to survive. It is possible that metabolites produced by *P. didemni* could be used as a way to chemically protect the sessile hosts against predators.

The presence of ABC systems predicted to function in iron and iron-siderophore uptake, in addition to other siderophore receptors in spite of the absence of siderophore biosynthesis genes might reveal the dependence of *P. didemni* cells on a siderophore, produced by another organism in the community, to supply its requirements for iron. This is supported by the presence of a TonB family C-terminus protein and a TonB receptor in addition to other proteins that might be involved in the uptake of a siderophore-ferric complex. Urea transporters were also detected that could provide its needs for nitrogen. Other transport systems specific to the uptake of amino acids and essential minerals were found (Table A1-1). In addition, a glycine-betaine uptake system important for the osmoprotection was also detected; these are common in marine bacteria. Supplying Fe (II), urea, and betaine in the culture medium might be important for the cultivation of this as yet uncultivated obligate symbiont. Although analysis of the genomes of *P. didemni* did not reveal any genes related to nitrogen recycling (Donia et al., 2011), the Dev system responsible for heterocyst formation was found. This system is specialized in nitrogen

fixation in conditions of nitrogen starvation (Toyoshima et al., 2010). It is unclear the reason for the presence of this system.

Although most transporters were present in the three genomes, differences also existed mostly in the export of toxins and peptides. Similarity in transporters could explain the symbiotic relationship while differences could indicate habitats dissimilarities.

The bioinformatic analysis of three different samples of an obligate symbiont as *P. didemni* could possibly lead to the cultivating of this obligate symbiont by including essential substrates, iron chelating compounds, and betaine in the growth medium.

## A2. Preliminary Microbial Analysis of *Prochloron* Containing Ascidians

### A2.1. Introduction

Ascidians are considered the most diverse class of marine tunicates, Tunicata. They are distributed throughout the world especially in shallow water (Shenkar and Swalla, 2011) in different habitats ranging from soft sediments to rocks. Each colony is made out of individual zooids that are surrounded by a cellulosic tunic. These zooids are responsible for filter feeding allowing a diverse microbial community to live in close association with the ascidian, in symbiosis. It was suggested that some symbionts live in a tight symbiotic relationship with a host by providing compounds used in chemical defense and in return share the host's metabolic products (Schmidt, 2008). The cyanobacterium *Prochloron didemni* is a strong example in this field. It is a photosynthetic bacterium and is considered the main symbiont in ascidians such as *Lissoclinum* and *Didemna* and contributes significantly to the green color on the surface of these animals. *P. didemni* share some similarities with plants. First, not only does it use chlorophyll a as other cyanobacteria but also uses chlorophyll b as in plants in photosynthesis. This allows it to capture as much energy as possible making it beneficial for the shallow water habitat of their host. Another aspect is that it contains thylakoids, the membrane that contains chlorophyll in plant cells and finally it lacks the light capturing chromophores present in other cyanobacteria, phycobilins. It is estimated that *P. didemni* provides their host with most of the carbon assimilated by the host (Koike and Suzuki, 1996).

Ascidians are sessile with soft bodies. It is proposed that ascidians rely on noxious small molecules, especially peptides, to deter predation (Schmidt, 2008). A different array of peptides was shown to be produced by *P. didemni* (Schmidt et al., 2005) that might be

involved in protection of the soft bodied hosts against predators. Thus, *P. didemni* is considered a major contributor in the ascidian symbiosis from both the primary and secondary metabolic point of view. In addition to *P. didemni*, other bacteria exist in the host. It is expected that these organisms also contribute to the primary and secondary metabolism of the holobiont.

Various bioinformatic analyses of the *P. didemni* genome detected genes responsible for the biosynthesis of many of the compounds reported previously from *L. patella.* However, no evidence was found for other compounds known to be present in the ascidian (Schmidt et al., 2012). An example of compounds with undetected genes in the *P. didemni* genome is patellazoles. Patellazoles (Figure A2-1) are a group of macrolides isolated from the ascidian *L. patella.* They are presumably biosynthesized by hybrid polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) enzymes and were reported to possess strong cytotoxic activity against human colon tumor cells (Richardson et al., 2005). Analysis of the genome of *P. didemni* recently published did not show any evidence of patellazole pathways (Donia et al., 2011) suggesting that these compounds are synthesized by a different microorganism within the *L. patella* community.

Studying the microbial community of different *L. patella* might shed some light on the other non-prochloron fraction of the community within ascidians. In addition, it might provide some information on the true source of these types of compounds.

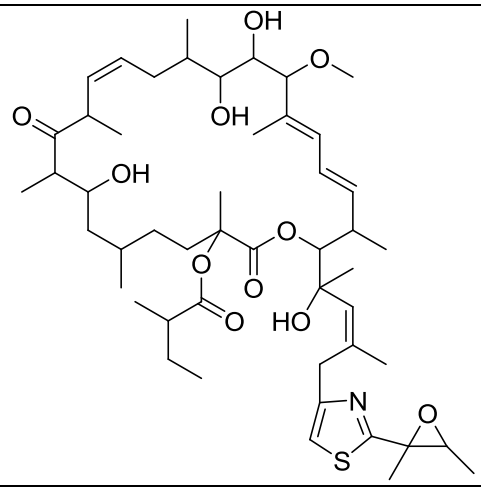



**Figure A2-1. Chemical structure of patellazole A.**

## A2.2. Materials and Methods

### A2.2.1. Sample Collection

Three ascidians samples were used in this analysis. 06-037 *L. patella* was collected from Fiji at coordinates S 17º55' E 177º16'; 06-027 *L. patella* from Solomon at coordinates S 8°57.35' E 159°59.12'; and 05-019 (*Lissolinum* sp.) collected at coordinates S 10º15.856' E 150º15.856'. Two of these samples contained patellazoles (06-037 and 06-027) while 05-019 did not.

### A2.2.2. DNA extraction and amplification

Ascidian tissue (~ 50 mg) was used for DNA extraction. Each ascidian tissue was extracted as previously described with slight modifications (Sambrook and Russell, 2001). Briefly, ascidian colonies were rinsed twice with filter-sterilized artificial seawater and homogenized. Lysis buffer (lysozyme 0.2 mg/ml, 100 mM Tris-HCl, 10 mM EDTA, 37 µM Triton 100×, 100 µg RNase A/ ml) was added to the tissue and the samples were incubated for 30 min at 37$^{o}$C. Proteinase K (0.5 mg/ml) was added and incubated for further 30 min at 65$^{o}$C with occasional shaking. The homogenate was incubated in CTAB (Cetyltrimethyl ammonium bromide) extraction buffer (1.4 N NaCl, 20 mM EDTA, 100 mM Tris-HCl, 3% CTAB, and 1% β-mercaptoethanol) for 30 min at 65$^{o}$C. Tissues were allowed to cool on ice for 5 min and centrifuged at 14,000 ×g for 5 min. The supernatant was extracted with 1× volume of phenol: chloroform: isoamyl alcohol (IAA). This was followed by centrifugation at 5,000 ×g for 5 min at 4$^{o}$C to separate phases. Chloroform: IAA was added to the aqueous layer and vortexed followed by centrifugation at 5,000 ×g for 5 min at 4$^{o}$C to obtain separate phases. DNA was precipitated by the addition of 0.1× sodium acetate and 2.5× ethanol and left at 4$^{o}$C for 2 hours. This was followed by

centrifugation at 13,000 ×g for 20 min at 4°C. The precipitated pellet was washed with 70% ethanol and left to air dry for ten minutes. The DNA was resuspended in $T_{10}E_1$ buffer in 4°C overnight. Samples where then purified using a Montage PCR purification kit (Millipore) to get rid of any remaining PCR inhibitors. DNA concentrations were measured using ND-1000 NanoDrop spectrophotometer and investigated by gel electrophoresis.

### A2.2.3. DNA amplification and sequencing

The 16S rRNA genes from the whole ascidian DNA were amplified by PCR using universal bacterial primers 27F and 1492R (Table A2-1). In addition, BSA at a final concentration of 0.2 µg/µl was included with a HF-Taq polymerase (Invitrogen). Cycling conditions were performed in a Gene Amp PCR System 9700 (Applied Biosystems) thermocycler using a 5 min hot start at 95°C; 20 cycles of 30 sec at 94°C, 2 min at 50°C, and 1.5 min at 72°C; and a final 10 min extension step at 72°C. The amplification success was confirmed using gel electrophoresis. Amplicons of the expected size (~1,500 bp) were gel extracted and purified using QIAEX® II Gel extraction kit (Qiagen). Purified DNA was cloned into pCR® 4 TOPO® Vector (invitrogen) and transformed into TOP10 chemically competent cells (invitrogen) as recommended by the manufacturer. Transformed cells were grown overnight and representatives of the transformed cells were analyzed for the presence of the insert by amplifying with the plasmid primers M13F and M13R designed by the manufacturer. Successfully transformed cells were inoculated in 96 well plates, and prepared for sequencing at the genome sequencing center at Washington University. The sequencing reaction was performed using M13F, 357F, 907R, and M13R (Table A2-1) with the BigDye TM terminator v3. Sequences

were obtained on an ABI Prism 3100 automated sequencer. Number of clones sequenced for each sample is indicated below the corresponding pie chart (Figure A2-2; Figure A2-3; Figure A2-4).

**Table A2-1. Sequence of primers used in microbial community analysis.**

| Primer name | Sequence |
|---|---|
| Universal 27F | 5'-AGA GTT TGA TCM TGG CTC AG-3' |
| Universal 1492R | 5'-TAC GGY TAC CTT GTT ACG ACT T-3' |
| 357F | 5'-CCT ACG GGA GGC AGC AG-3' |
| 907R | 5'-CCG TCA ATT CMT TTG AGT TT-3' |
| M13F | 5´-GTA AAA CGA CGG CCA G-3´ |
| M13R | 5´-CAG GAA ACA GCT ATG AC-3´ |
| 16S rRNA primers positions are according to *E. coli* | |

### A2.2.4. DNA annotation and phylogeny

All sequences were imported into the Geneious program (Drummond et al., 2009) and trimmed to remove plasmid sequences. Multiple sequences obtained from different primers of the same sample were assembled into one contig and checked for 27F and 1492R primers. Sequences were checked for plasmid contamination using the online tools VecScreen from NCBI. In addition, the sequences were analyzed by ChimeraCheck to ensure the absence of chimeras. Sequences were then classified using the RDP classifier (Wang et al., 2007), and analyzed using nucleotide BLAST search at National Center for Biotechnology Information (NCBI) Reference Sequences.

## A2.3. Results

Microbial analysis of 06-027 (P3) revealed that the ascidian mainly harbors alphaproteobacteria, constituting 80% of the total microbial population (Figure A2-3). In addition, cyanobacteria and gammaproteobacteria were present in lower abundance of approximately 8% of the total population. The main cyanobacterial sequence was *P. didemni* as expected. Actinobacteria were present but rare in the community.

The microbial community analysis also revealed that gammaproteobacteria are abundant in 06-037 (P2) representing more than 80% of the total microbial community in this ascidian sample (Figure A2-2). In addition, cyanobacteria were present in about 14% of the library where the major sequences found belonged to *P. didemni*. Actinobacteria and alphaproteobacteria were less abundant in the community. Nblast analysis suggested that the main actinobacteria is a *Streptomyces*.

Analysis of the non-patellazole sample 05-019 revealed that it is more diverse than the previous patellazole producing samples (Figure A2-4). Firmicutes and gammaproteobacteria together represented more than half of the microbial population. Alphaproteobacteria and cyanobacteria were also part of the community. 05-019 was the highest among the three samples in the abundance of actinobacteria, representing about 10% of the total population.

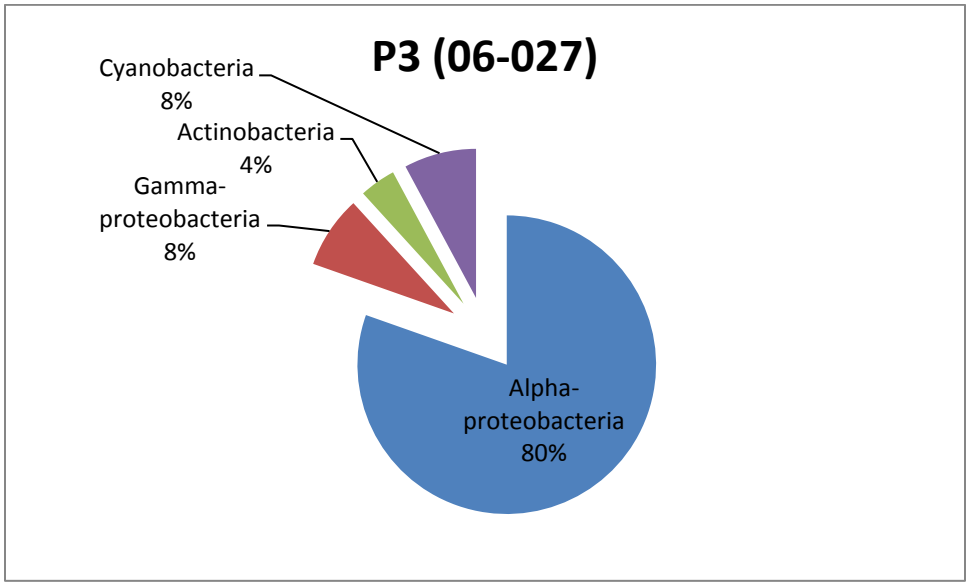



**Figure A2-2. Pie chart presenting the classification of the microbial community in patellazole positive *L. patella* 06-027 (P3).**
Number of sequences used, 51.

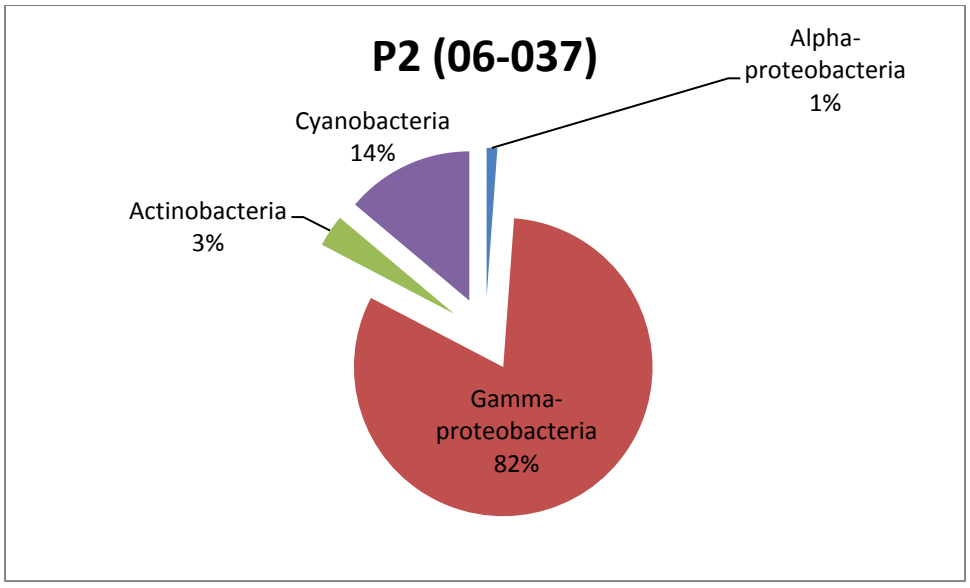



**Figure A2-3. Pie chart presenting the classification of the microbial community in patellazole positive *L. patella* 06-037 (P2).**
Number of sequences used, 173.

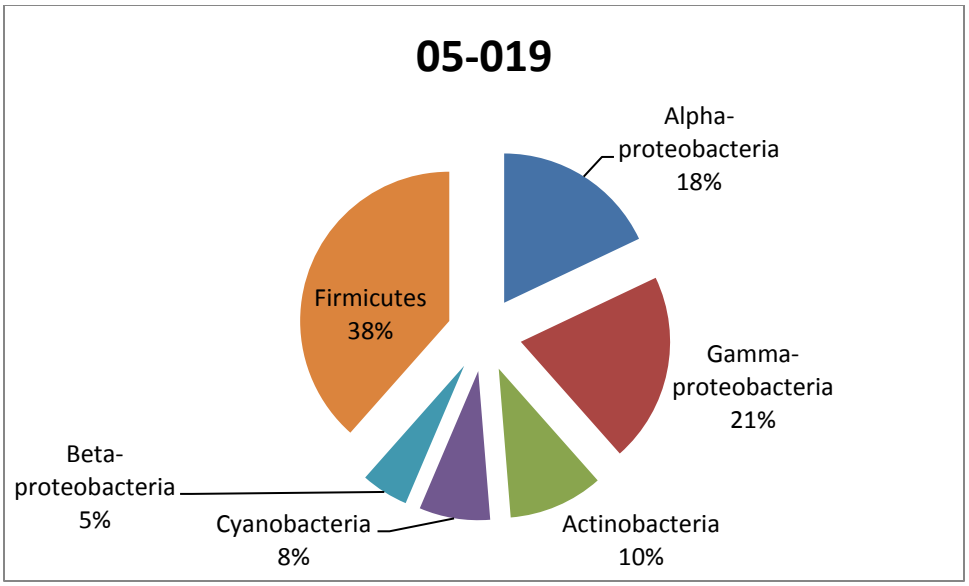



**Figure A2-4. Pie chart presenting the classification of the microbial community in patellazole negative *Lissoclinum* spp 05-019.**

Number of sequences used, 39.

## A2.4. Discussion

It is clear that the ascidian *Lissoclinum* harbors a rich community of microbial organisms that are thought to play a major role in the metabolism of the host. In addition, it is suggested that this rich source of microbial diversity contributes significantly to the secondary metabolites isolated from the tunic (Donia et al., 2011). Two *Lissoclinum* samples collected from Solomon (06-027) and Fiji (06-037) were found to contain the potent cytotoxic patellazole. It is more likely that they are produced by a microorganism. In order to help identify a possible symbiont in these samples that might be responsible for producing patellazoles, we aimed to compare and contrast the microbial communities of these samples using PCR and cloning techniques.

Besides *P. didemni* the main genus in the cyanobacteria class, another cyanobacterium was found to be the filamentous *Calothrix* present in the microbial community of P2 (06-037). An interesting feature is that actinobacteria, an interesting clade known to produce antibiotics and other secondary metabolites, represent about 3-10% of the microbial population in the three tunicates and might be responsible for production of the active compound. However, one cannot underestimate the ability of many gammaproteobacteria of producing interesting metabolites.

Cyanobacteria in this analysis were not as abundant as was shown in a different microbial analysis using 454 (Donia et al., 2011) in spite of the fact that no subtractive hybridization was done. This might be related to either the method of DNA extraction of the tissue used in our analysis, or the body part used. Other possibilities include the low number of PCR cycles (20-25) used or the cycle conditions. A bias of the DNA extraction or amplification using PCR primers in either of the studies might be contributing to this.

Nevertheless, it was clear that proteobacteria was a dominant class in the ascidian community. While proteobacterial are typically overlooked as sources of secondary metabolites, recent examples suggest they are worth of attention as in the case of *Xenorhabdus* (Bode, 2009) and *Teredinibacter turnerae* (Yang et al., 2009).

The Schmidt group has recently discovered potential *trans* AT and KS genes that are strongly suggested to belong to the patellazole pathway. They suggested a potential producer to be an alphaproteobacterium (Kwan, 2012). Current work is underway through the collaboration between the Haygood and the Schmidt labs to identify the real patellazole producer using bioinformatic, microscopy, and expression analyses.

## A3. Spectral and Biological Data

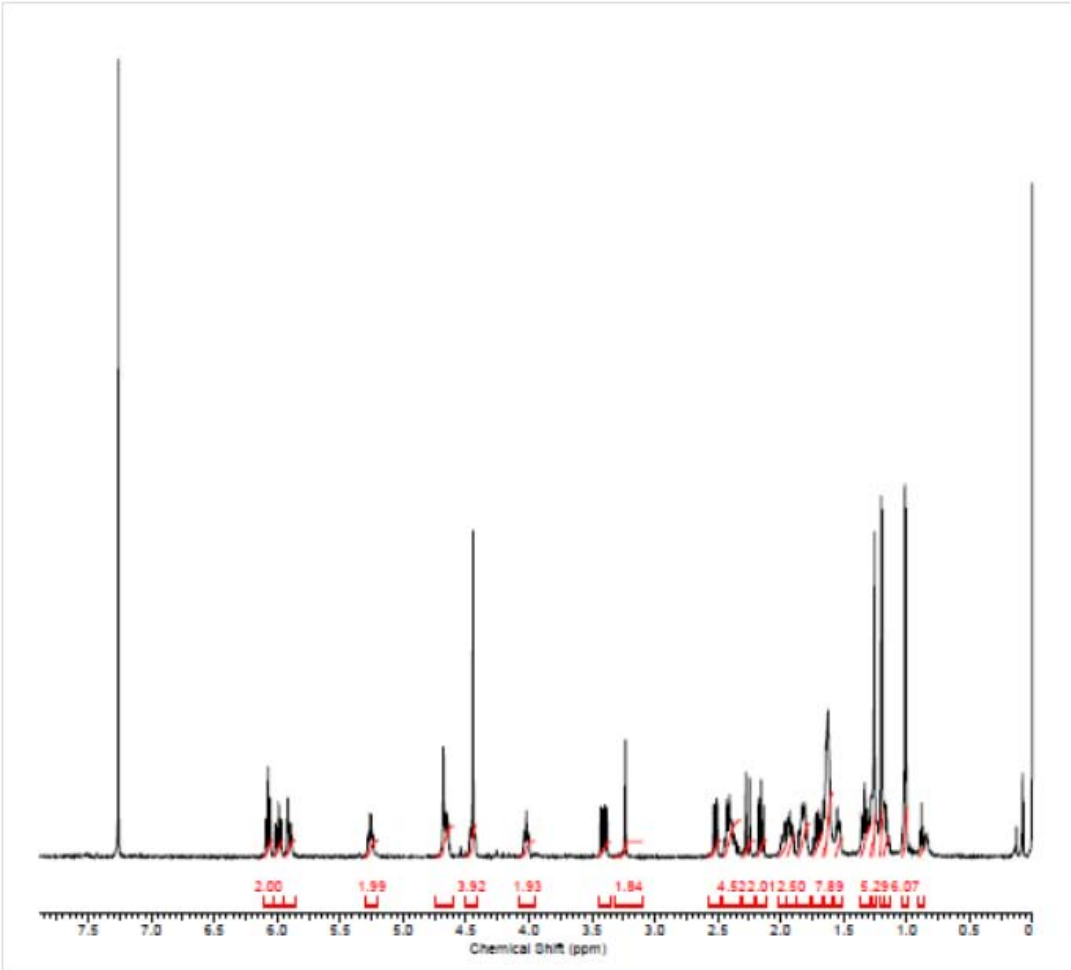



**Figure A3-1. $^{1}$H NMR spectrum of tartrolon E.**

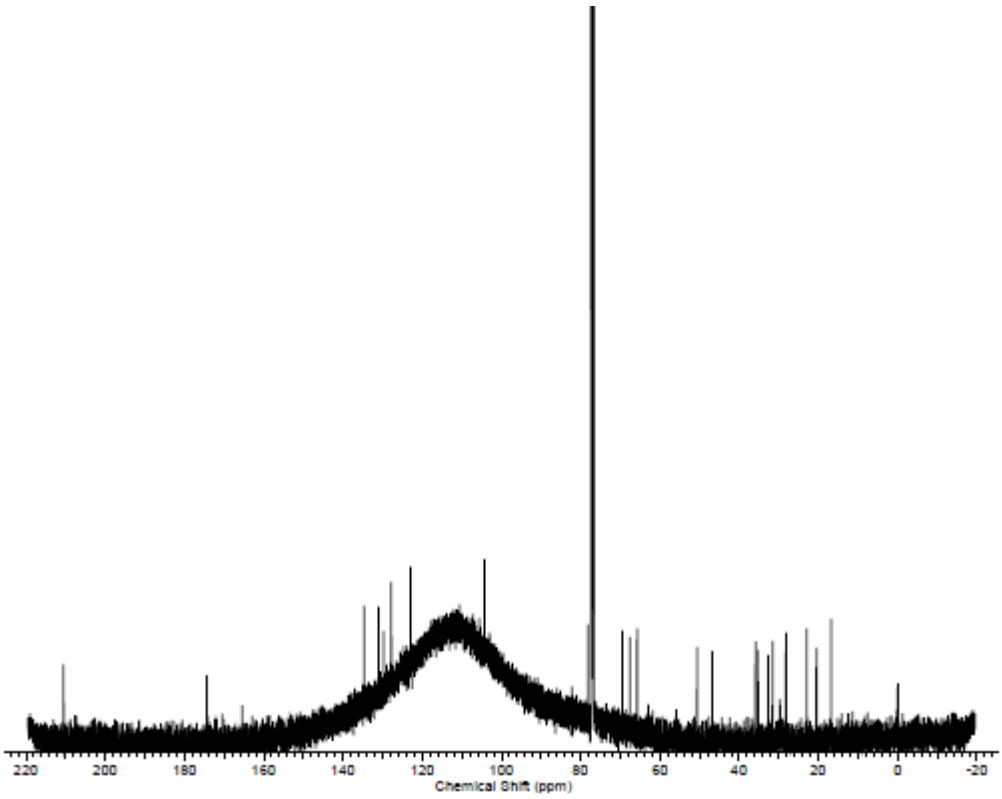



**Figure A3-2. $^{13}C$ NMR spectrum of tartrolon E.**

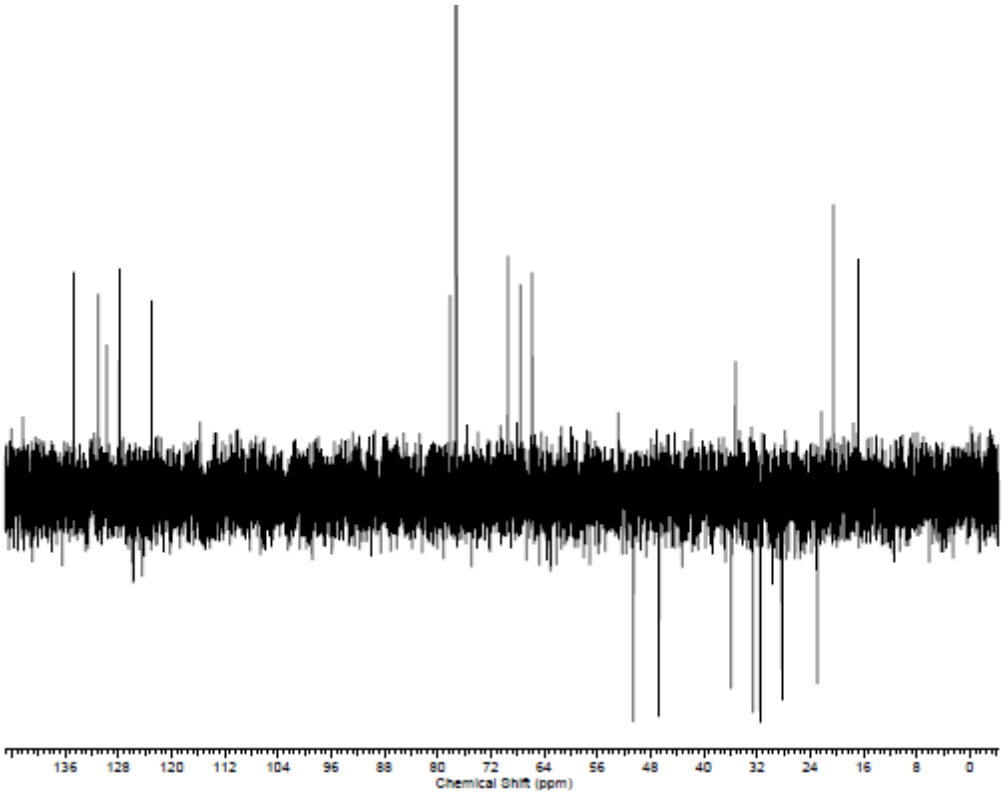



**Figure A3-3. $^{13}C$ Dept 135 NMR spectrum of tartrolon E.**

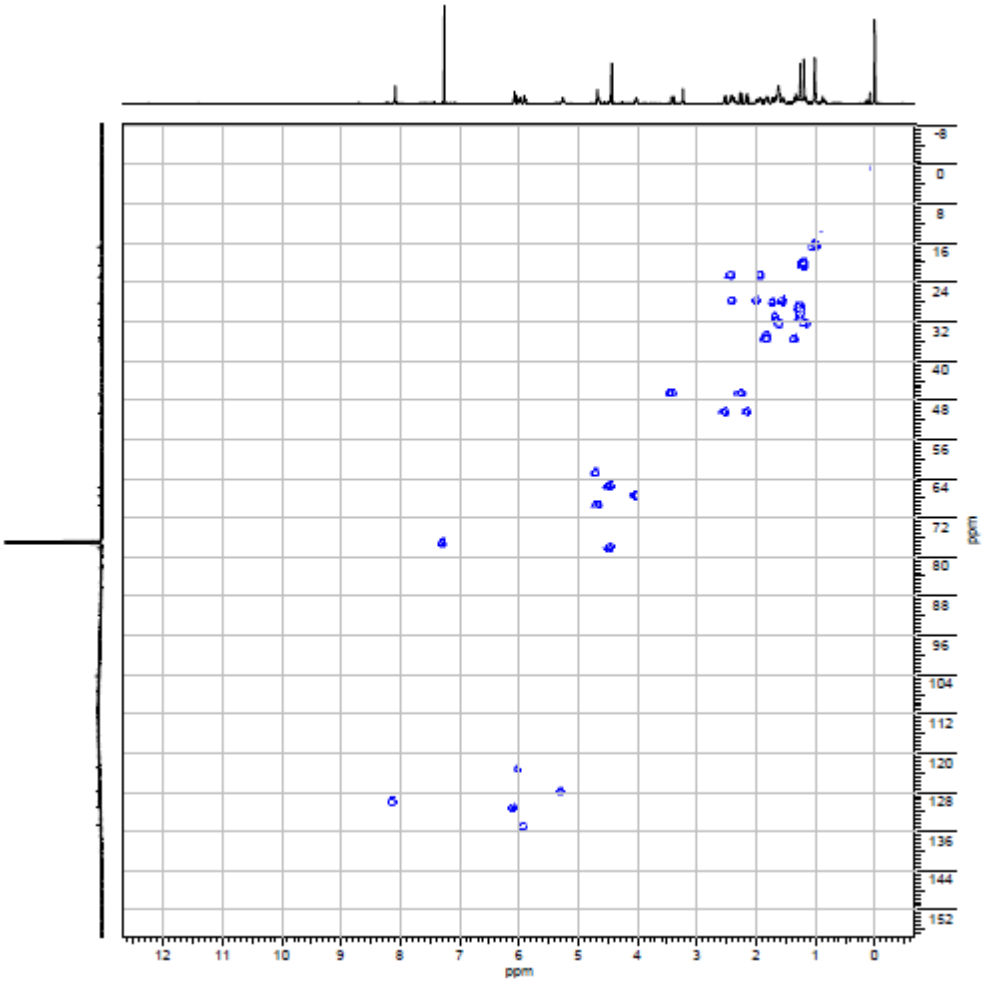



**Figure A3- 4. $^{1}H$-$^{13}C$ HSQC NMR spectrum of tartrolon E.**

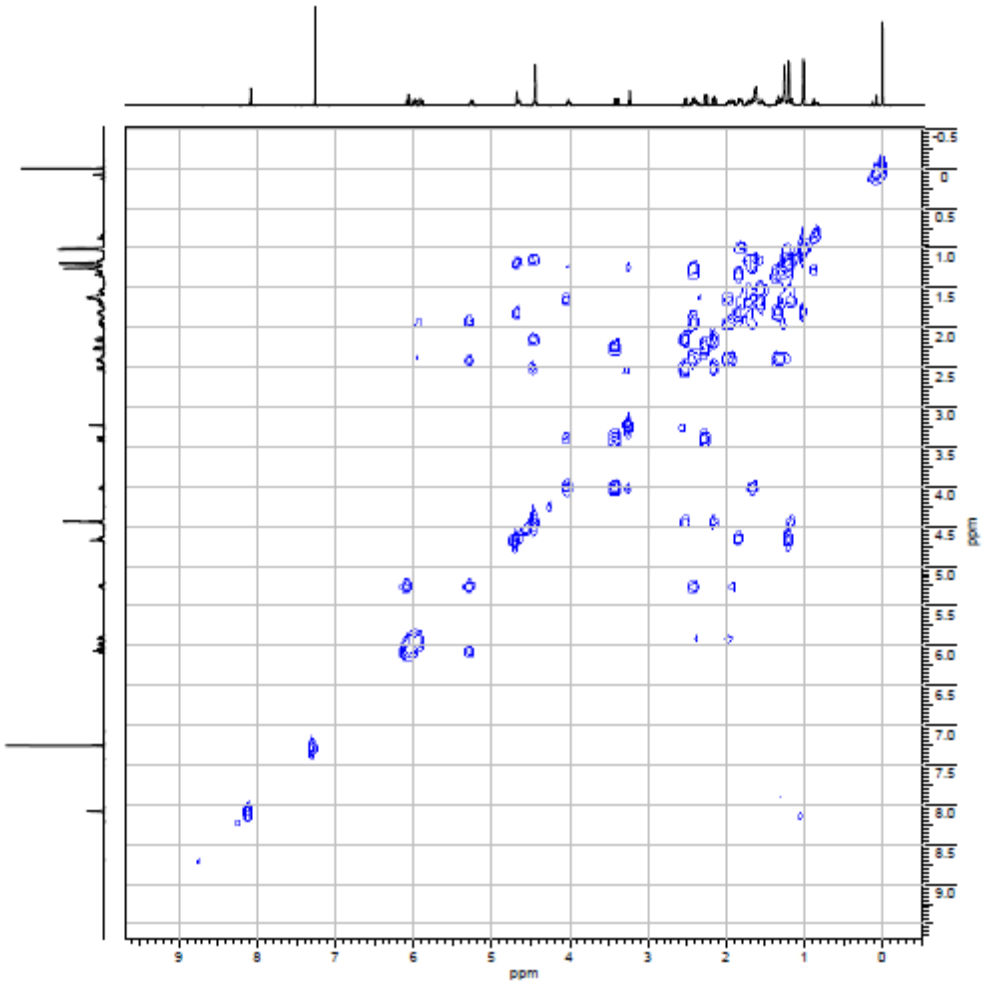



**Figure A3-5. $^{1}$H-$^{1}$H COSY NMR spectrum of tartrolon E.**

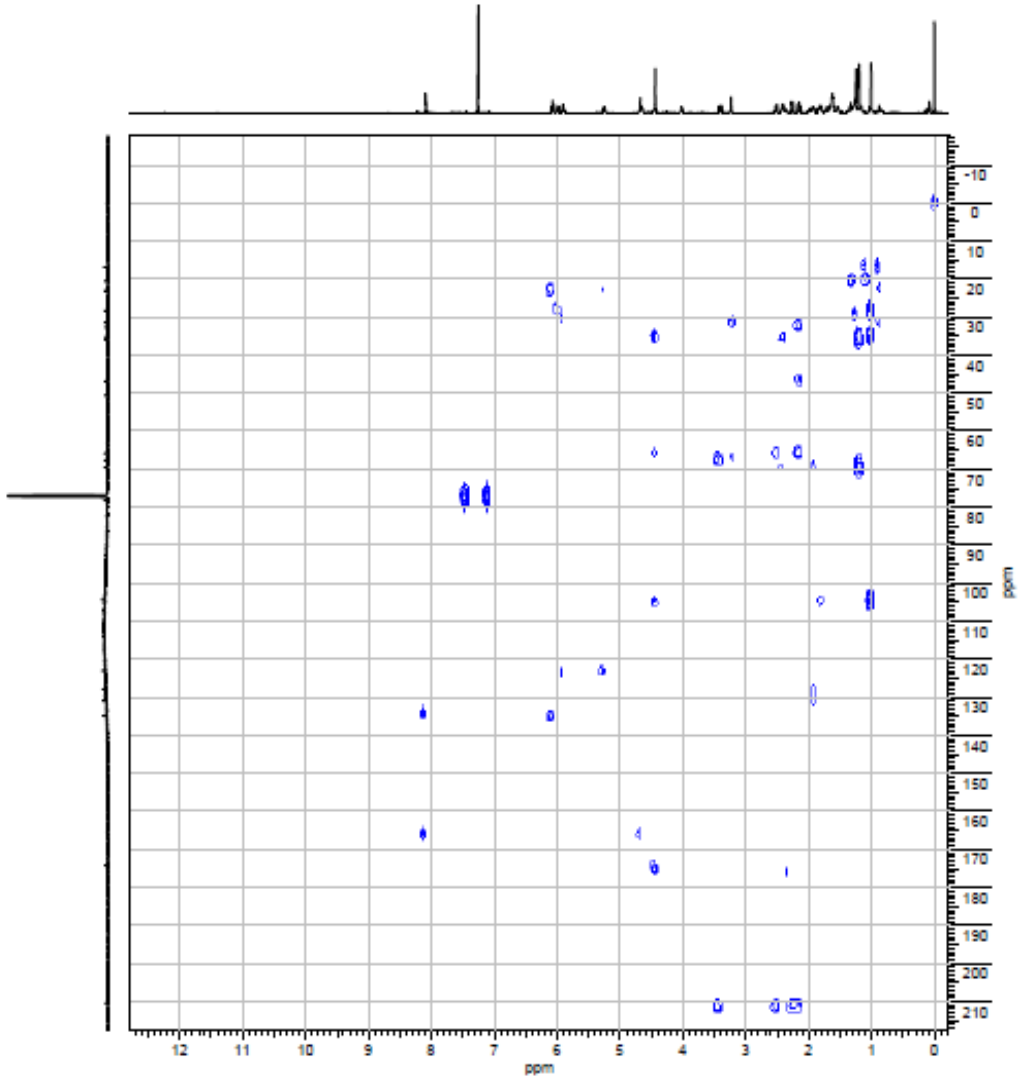



**Figure A3-6. $^{1}H$-$^{13}C$ HMBC NMR spectrum of tartrolon E.**

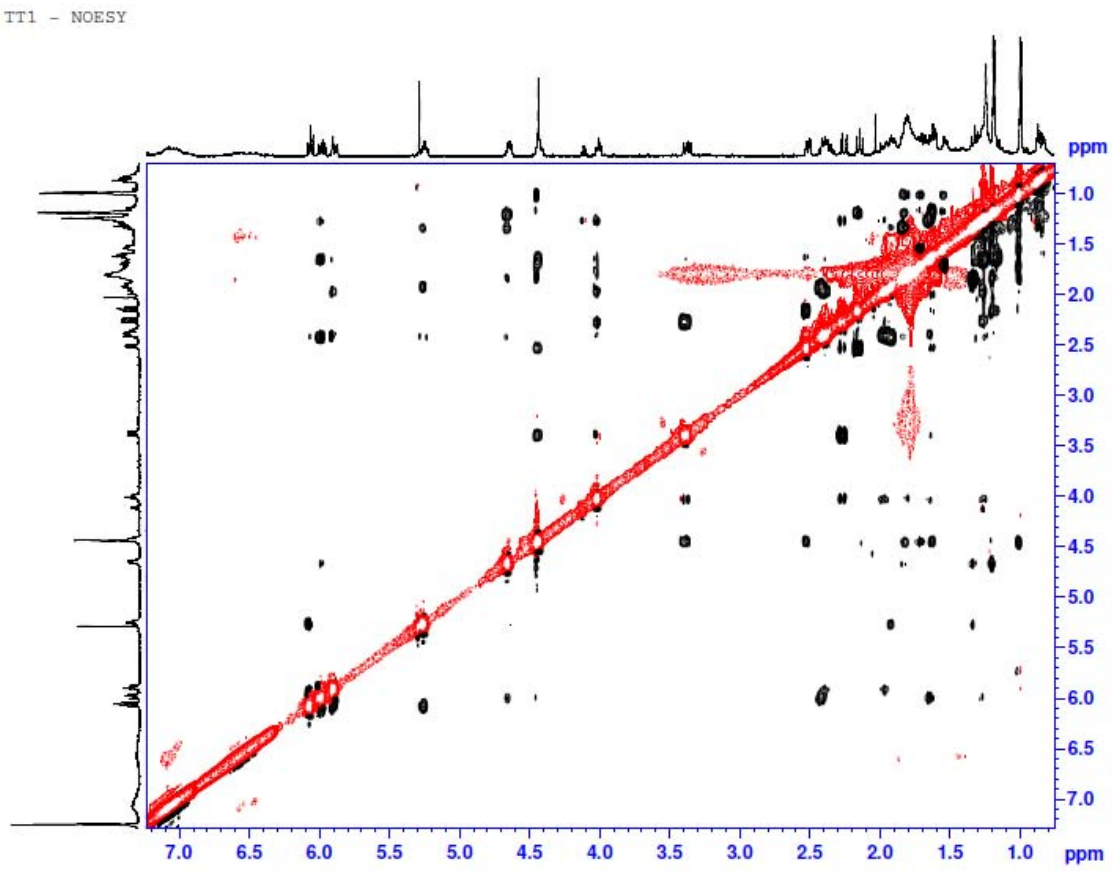



**Figure A3-7. NOESY NMR spectrum of tartrolon E**

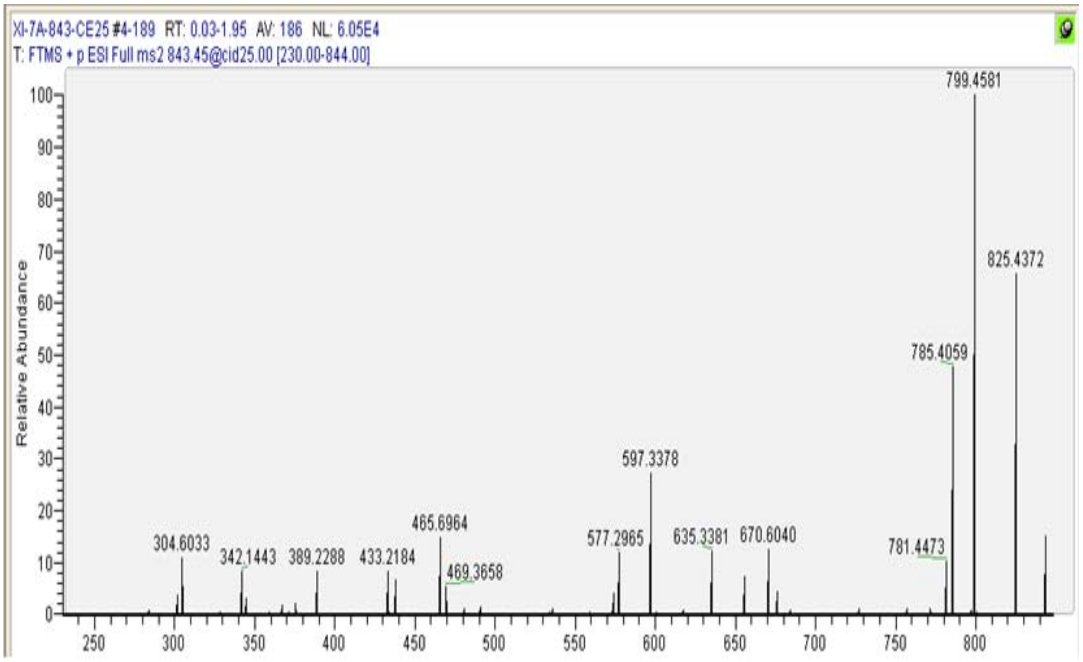



**Figure A3-8. MS/MS fragmentation pattern of tartrolon D.**

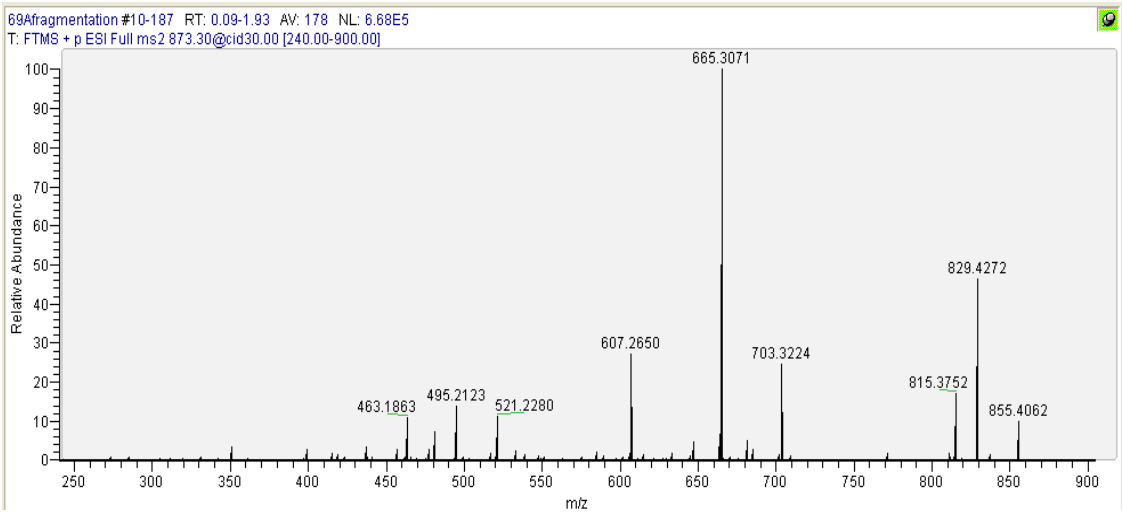



**Figure A3-9. MS/MS fragmentation pattern of tartrolon E.**

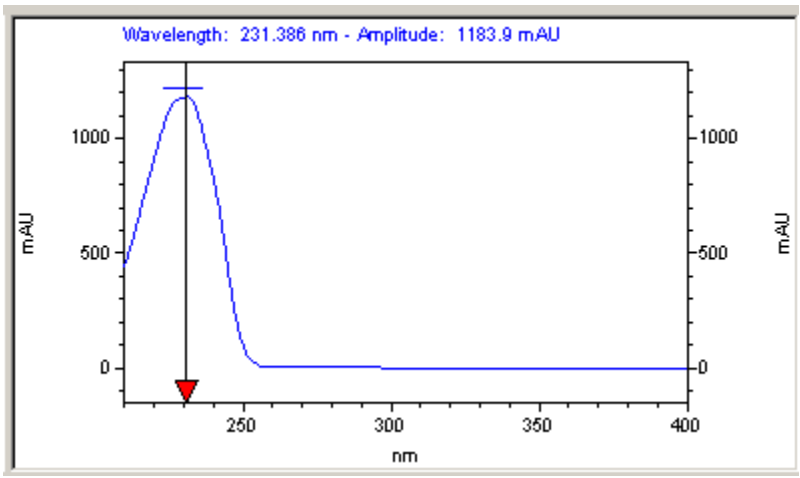



**Figure A3-10. UV spectrum of tartrolon E.**

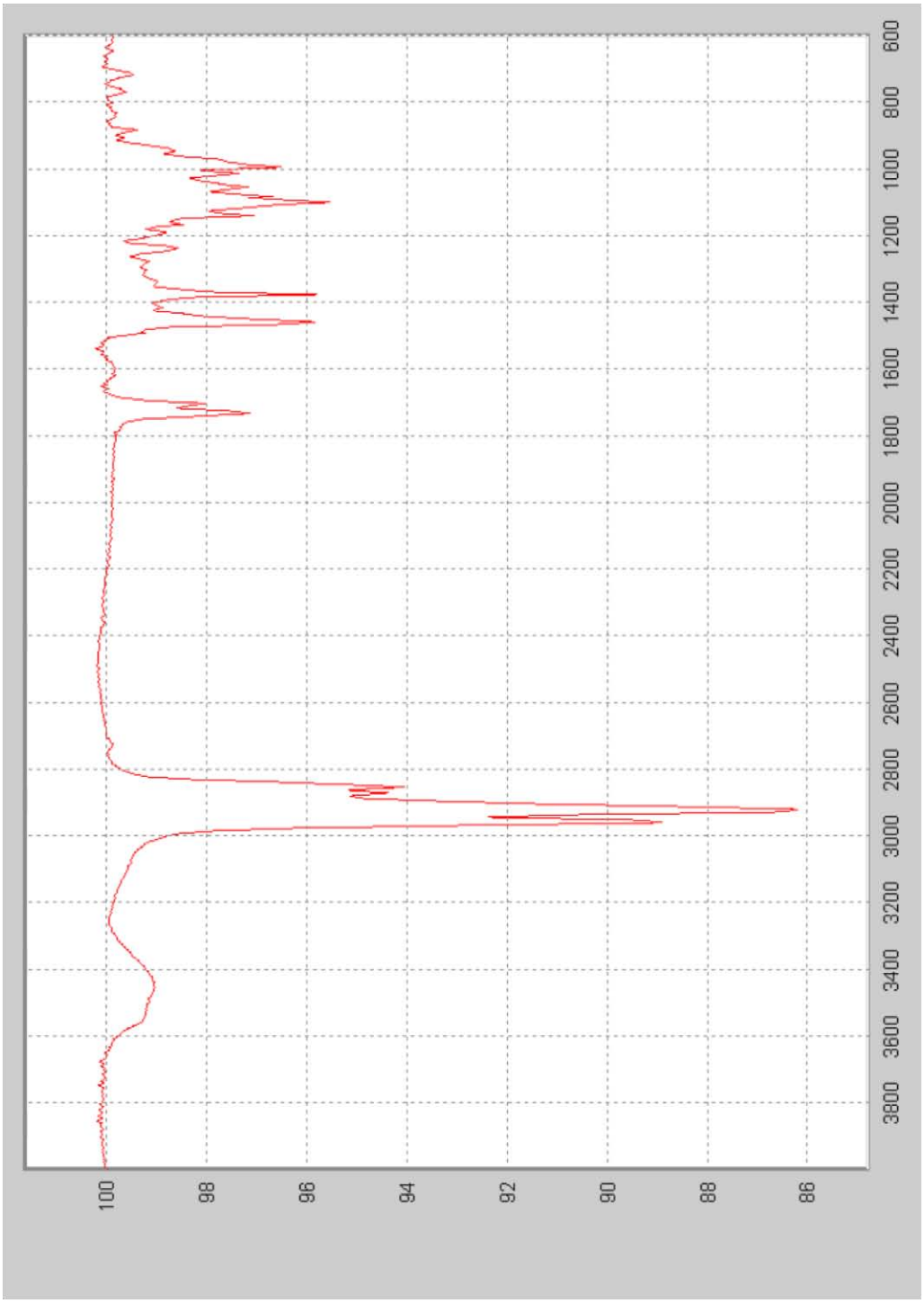



**Figure A3-11. Neat FT-IR spectrum of tartrolon E.**

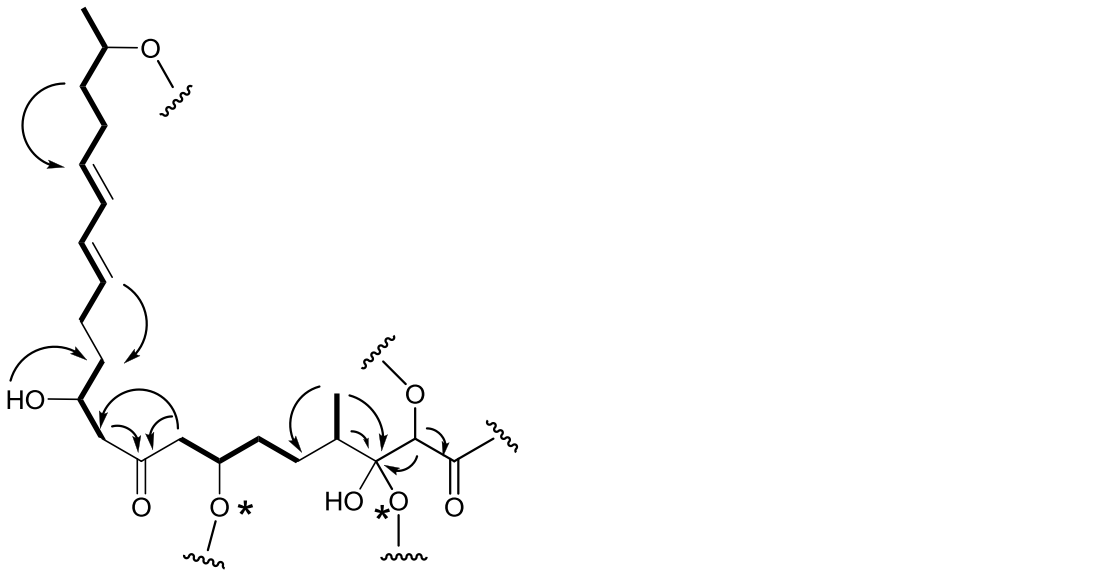



**Figure A3-12. Key $^{1}$H–$^{13}$C HMBC and $^{1}$H–$^{1}$H COSY correlations of partial structure of tartrolon E.**
*, similar oxygen connected through pyran ring.

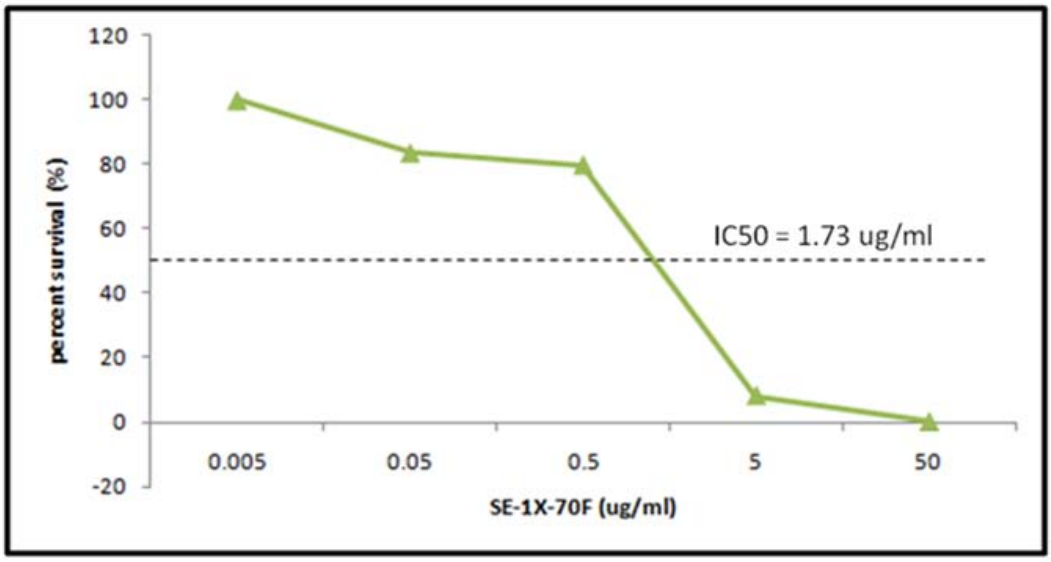



**Figure A3-13. Percent survival of MCF-7 cells as a function of tartrolon E concentration showing $IC_{50}$.**

# References


Amnuoypol, Surattana, Khanit Suwanborirux, Sunibhond Pummangura, Akinori Kubo, Chieko Tanaka, and Naoki Saito. 2004. “Chemistry of Renieramycins. Part 5. Structure Elucidation of Renieramycin-type Derivatives O, Q, R, and S from Thai Marine Sponge *Xestospongia* Species Pretreated with Potassium Cyanide.” *Journal of Natural Products* 67 (6) (June): 1023–1028. doi:10.1021/np030534o.

An, Tianying, Jamie Winshell, Gloria Scorzetti, Jack W Fell, and Kathleen S Rein. 2010. “Identification of Okadaic Acid Production in the Marine Dinoflagellate *Prorocentrum rhathymum* from Florida Bay.” *Toxicon: Official Journal of the International Society on Toxinology* 55 (2-3) (March): 653–657. doi:10.1016/j.toxicon.2009.08.018.

Andrews, Jennifer M. 2002. “Determination of Minimum Inhibitory Concentrations.” *Journal of Antimicrobial Chemotherapy* 49 (6) (June 1): 1049. doi:10.1093/jac/dkf083.

Antón, Nuria, Javier Santos-Aberturas, Marta V Mendes, Susana M Guerra, Juan F Martín, and Jesús F Aparicio. 2007. “PimM, a PAS Domain Positive Regulator of Pimaricin Biosynthesis in *Streptomyces natalensis*.” *Microbiology (Reading, England)* 153 (Pt 9) (September): 3174–3183. doi:10.1099/mic.0.2007/009126-0.

Arai, T, K Takahashi, S Nakahara, and A Kubo. 1980. “The Structure of a Novel Antitumor Antibiotic, Saframycin A.” *Experientia* 36 (9) (September 15): 1025–1027.

Ardelli, B. F., L. E. Stitt, and J. B. Tompkins. 2010. “Inventory and Analysis of ATP-binding Cassette (ABC) Systems in *Brugia malayi*.” *Parasitology* (March 17): 1–18.

Aune, Gregory J, Takahisa Furuta, and Yves Pommier. 2002. “Ecteinascidin 743: a Novel Anticancer Drug with a Unique Mechanism of Action.” *Anti-Cancer Drugs* 13 (6) (July): 545–555.

Bachmann, Brian O, and Jacques Ravel. 2009. “Chapter 8. Methods for in Silico Prediction of Microbial Polyketide and Nonribosomal Peptide Biosynthetic Pathways from DNA Sequence Data.” *Methods in Enzymology* 458: 181–217. doi:10.1016/S0076-6879(09)04808-3.

Baden, Daniel G., Andrea J. Bourdelais, Henry Jacocks, Sophie Michelliza, and Jerome Naar. 2005. “Natural and Derivative Brevetoxins: Historical Background,

Multiplicity, and Effects." *Environmental Health Perspectives* 113 (5) (May): 621–625. doi:10.1289/ehp.7499.

Banker, Ronny, and Shmuel Carmeli. 1998. "Tenuecyclamides A−D, Cyclic Hexapeptides from the Cyanobacterium *Nostoc spongiaeforme* Var. Tenue." *Journal of Natural Products* 61 (10): 1248–1251. doi:10.1021/np980138j.

Bennett, A., R. I. Rowe, N. Soch, and C. D. Eckhert. 1999. "Boron Stimulates Yeast (*Saccharomyces cerevisiae*) Growth." *The Journal of Nutrition* 129 (12) (December 1): 2236 –2238.

Berry, A M, O T Harriott, R A Moreau, S F Osman, D R Benson, and A D Jones. 1993. "Hopanoid Lipids Compose the *Frankia* Vesicle Envelope, Presumptive Barrier of Oxygen Diffusion to Nitrogenase." *Proceedings of the National Academy of Sciences* 90 (13) (July 1): 6091–6094.

Betcher, Meghan. 2011. "Insights into the Shipworm (Bivalvia: Teredinidae)–bacterial Symbiosis: Novel Isolates and Micrographic Bacterial Localization Studies." Master's Thesis, Oregon Health & Science University.

Blodgett, Joshua A. V., Dong-Chan Oh, Shugeng Cao, Cameron R. Currie, Roberto Kolter, and Jon Clardy. 2010. "Common Biosynthetic Origins for Polycyclic Tetramate Macrolactams from Phylogenetically Diverse Bacteria." *Proceedings of the National Academy of Sciences* 107 (26) (June 29): 11692–11697. doi:10.1073/pnas.1001513107.

Bode, Helge B. 2009. "Entomopathogenic Bacteria as a Source of Secondary Metabolites." *Current Opinion in Chemical Biology* 13 (2) (April): 224–230. doi:10.1016/j.cbpa.2009.02.037.

Bolaños, Luis, Miguel Redondo-Nieto, Ildefonso Bonilla, and Luis G Wall. 2002. "Boron Requirement in the *Discaria trinervis* (Rhamnaceae) and *Frankia* Symbiotic Relationship. Its Essentiality for *Frankia* BCU110501 Growth and Nitrogen Fixation." *Physiologia Plantarum* 115 (4) (July): 563–570.

Bonilla, I, M Garcia-González, and P Mateo. 1990. "Boron Requirement in Cyanobacteria : Its Possible Role in the Early Evolution of Photosynthetic Organisms." *Plant Physiology* 94 (4) (December): 1554–1560.

Bonner, D P, J O'Sullivan, S K Tanaka, J M Clark, and R R Whitney. 1988. "Lysobactin, a Novel Antibacterial Agent Produced by *Lysobacter* Sp. II. Biological Properties." *The Journal of Antibiotics* 41 (12) (December): 1745–1751.

Bowen, H. J. M. 1966. "*Trace Elements in Biochemistry.*" Academic Press: London.

Braibant, M., P. Gilot, and J. Content. 2000. “The ATP Binding Cassette (ABC) Transport Systems of *Mycobacterium tuberculosis*.” *FEMS Microbiology Reviews* 24 (4) (October): 449–467.

Brown, Derek F. J., David I. Edwards, Peter M. Hawkey, Donald Morrison, Geoffrey L. Ridgway, Kevin J. Towner, Michael W. D. Wren, and the Joint Working Party of the British Society for Antimicrobial Chemotherapy. 2005. “Guidelines for the Laboratory Diagnosis and Susceptibility Testing of Methicillin-resistant *Staphylococcus aureus* (MRSA).” *Journal of Antimicrobial Chemotherapy* 56 (6) (December): 1000–1018. doi:10.1093/jac/dki372.

Bumpus, Stefanie B, Bradley S Evans, Paul M Thomas, Ioanna Ntai, and Neil L Kelleher. 2009. “A Proteomics Approach to Discovering Natural Products and Their Biosynthetic Pathways.” *Nature Biotechnology* 27 (10) (October): 951–956. doi:10.1038/nbt.1565.

Bunet, Robert, Lijiang Song, Marta Vaz Mendes, Christophe Corre, Laurence Hotel, Nicolas Rouhier, Xavier Framboisier, Pierre Leblond, Gregory L. Challis, and Bertrand Aigle. 2011. “Characterization and Manipulation of the Pathway-specific Late Regulator AlpW Reveals *Streptomyces ambofaciens* as a New Producer of Kinamycins.” *Journal of Bacteriology* 193 (5) (March 1): 1142–1153. doi:10.1128/JB.01269-10.

Butcher, Rebecca A, Frank C Schroeder, Michael A Fischbach, Paul D Straight, Roberto Kolter, Christopher T Walsh, and Jon Clardy. 2007. “The Identification of Bacillaene, the Product of the PksX Megacomplex in *Bacillus subtilis*.” *Proceedings of the National Academy of Sciences of the United States of America* 104 (5) (January 30): 1506–1509. doi:10.1073/pnas.0610503104.

Cao, Shugeng, Joshua A. V. Blodgett, and Jon Clardy. 2010. “Targeted Discovery of Polycyclic Tetramate Macrolactams from an Environmental *Streptomyces* Strain.” *Organic Letters* 12 (20): 4652–4654. doi:10.1021/ol1020064.

Carballo, Jos Luis. 2000. “Larval Ecology of an Ascidian Tropical Population in a Mediterranean Enclosed Ecosystem.” *Marine Ecology Progress Series* 195 (March 31): 159–167. doi:10.3354/meps195159.

Carballo, José Luis, Santiago Naranjo, Bulent Kukurtzü, Fernando Calle, and Aida Hernández-Zanuy. 2000. “Production of *Ecteinascidia turbinata* (Ascidiacea:

Perophoridae) for Obtaining Anticancer Compounds." *Journal of the World Aquaculture Society* 31 (4) (December 1): 481–490. doi:10.1111/j.1749-7345.2000.tb00900.x.

Cardellina, Jh, Fj Marner, and Re Moore. 1979. "Seaweed Dermatitis: Structure of Lyngbyatoxin A." *Science* 204 (4389) (April 13): 193–195. doi:10.1126/science.107586.

Carrano, Carl J, Stephen Schellenberg, Shady A Amin, David H Green, and Frithjof C Küpper. 2009. "Boron and Marine Life: a New Look at an Enigmatic Bioelement." *Marine Biotechnology (New York, N.Y.)* 11 (4) (August): 431–440. doi:10.1007/s10126-009-9191-4.

Challis, G L, J Ravel, and C A Townsend. 2000. "Predictive, Structure-based Model of Amino Acid Recognition by Nonribosomal Peptide Synthetase Adenylation Domains." *Chemistry & Biology* 7 (3) (March): 211–224.

Chang, Fang-Yuan, and Sean F. Brady. 2011. "Cloning and Characterization of an eDNA Derived Gene Cluster That Encodes the Biosynthesis of the Antitumor Substance BE-54017." *Journal of the American Chemical Society* 133 (26) (June 6): 9996–9999. doi:10.1021/ja2022653.

Chen, T S, C J Chang, and H G Floss. 1980. "Biosynthesis of the Boron-containing Antibiotic Aplasmomycin. Nuclear Magnetic Resonance Analysis of Aplasmomycin and Desboroaplasmomycin." *The Journal of Antibiotics* 33 (11) (October): 1316–1322.

Chen, Tom S. S., Ching-Jer Chang, and Heinz G. Floss. 1981. "Biosynthesis of the Boron-containing Macrolide Antibiotic Aplasmomycin by *Streptomyces griseus*." *Journal of the American Chemical Society* 103 (15): 4565–4568. doi:10.1021/ja00405a044.

Chen, X H, A Koumoutsi, R Scholz, K Schneider, J Vater, R Süssmuth, J Piel, and R Borriss. 2009. "Genome Analysis of *Bacillus amyloliquefaciens* FZB42 Reveals Its Potential for Biocontrol of Plant Pathogens." *Journal of Biotechnology* 140 (1-2) (March 10): 27–37. doi:10.1016/j.jbiotec.2008.10.011.

Chen, Xiao-Hua, Joachim Vater, Jörn Piel, Peter Franke, Romy Scholz, Kathrin Schneider, Alexandra Koumoutsi, Gabriele Hitzeroth, Nicolas Grammel, Axel W. Strittmatter, Gerhard Gottschalk, Roderich D. Süssmuth, and Rainer Borriss. 2006. "Structural and Functional Characterization of Three Polyketide Synthase Gene Clusters in

*Bacillus amyloliquefaciens* FZB 42." *Journal of Bacteriology* 188 (11) (May 1): 4024–4036. doi:10.1128/JB.00052-06.

Cuevas, C., and A. Francesch. 2009. "Development of Yondelis (trabectedin, ET-743). A Semisynthetic Process Solves the Supply Problem." *Natural Product Reports* 26 (3) (February): 322–337.

Davidson, A. L., and J. Chen. 2004. "ATP-binding Cassette Transporters in Bacteria." *Annual Review of Biochemistry* 73: 241–268.

Davidson, A. L., E. Dassa, C. Orelle, and J. Chen. 2008. "Structure, Function, and Evolution of Bacterial ATP-binding Cassette Systems." *Microbiology and Molecular Biology Reviews: MMBR* 72 (2) (June): 317–64.

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*." *Applied and Environmental Microbiology* 67 (10) (August): 4531–4537.

Davidson, S. K., and M. G. Haygood. 1999. "Identification of Sibling Species of the Bryozoan *Bugula neritina* that Produce Different Anticancer Bryostatins and Harbor Distinct Strains of the Bacterial Symbiont' *Candidatus* Endobugula sertula'." *The Biological Bulletin* 196 (3): 273.

Davies, Julian, and Katherine S. Ryan. 2011. "Introducing the Parvome: Bioactive Compounds in the Microbial World." *ACS Chemical Biology* doi:10.1021/cb200337h. http://dx.doi.org/10.1021/cb200337h.

Dembitsky, Valery M, Abed Al Aziz Al Quntar, and Morris Srebnik. 2011. "Natural and Synthetic Small Boron-containing Molecules as Potential Inhibitors of Bacterial and Fungal Quorum Sensing." *Chemical Reviews* 111 (1) ( 12): 209–237. doi:10.1021/cr100093b.

Demydchuk, Yuliya, Yuhui Sun, Hui Hong, James Staunton, Jonathan B Spencer, and Peter F Leadlay. 2008. "Analysis of the Tetronomycin Gene Cluster: Insights into the Biosynthesis of a Polyether Tetronate Antibiotic." *Chembiochem: A European Journal of Chemical Biology* 9 (7) (May 5): 1136–1145. doi:10.1002/cbic.200700715.

Distel, D L. 2003. "The Biology of Marine Wood Boring Bivalves and Their Bacterial Endosymbionts." in *Wood Deterioration and Preservation*, ed. B. Goodell, D. Nicholas, & T. Schultz. 253–271. American Chemical Society Press, Washington.

Distel, Daniel L, David J Beaudoin, and Wendy Morrill. 2002. “Coexistence of Multiple Proteobacterial Endosymbionts in the Gills of the Wood-boring Bivalve *Lyrodus pedicellatus* (Bivalvia: Teredinidae).” *Applied and Environmental Microbiology* 68 (12) (December): 6292–6299.

Distel, Daniel L, Wendy Morrill, Noelle MacLaren-Toussaint, Dianna Franks, and John Waterbury. 2002. “*Teredinibacter turnerae* Gen. Nov., Sp. Nov., a Dinitrogen-fixing, Cellulolytic, Endosymbiotic Gamma-proteobacterium Isolated from the Gills of Wood-boring Molluscs (Bivalvia: Teredinidae).” *International Journal of Systematic and Evolutionary Microbiology* 52 (Pt 6) (November): 2261–2269.

Donia, Mohamed S, W Florian Fricke, Jacques Ravel, and Eric W Schmidt. 2011. “Variation in Tropical Reef Symbiont Metagenomes Defined by Secondary Metabolism.” *PloS One* 6 (3): e17897. doi:10.1371/journal.pone.0017897.

Donia, Mohamed S., W. Florian Fricke, Frédéric Partensky, James Cox, Sherif I. Elshahawi, James R. White, Adam M. Phillippy, Michael C. Schatz, Joern Piel, Margo G. Haygood, Jacques Ravel, and Eric W. Schmidt. 2011. “Complex Microbiome Underlying Secondary and Primary Metabolism in the Tunicate-*Prochloron* Symbiosis.” *Proceedings of the National Academy of Sciences* (November 28). doi:10.1073/pnas.1111712108. http://www.pnas.org/content/early/2011/11/22/1111712108.

Drummond, A. J., B. Ashton, M. Cheung, J. Heled, M. Kearse, R. Moir, S. Stones-Havas, T. Thierer, and A. Wilson. 2009. “Geneious V4. 7.” *Biomatters Ltd*.

Dunitz, J. D, D. M Hawley, D. Mikloš, D. N. J White, Yu Berlin, R. Marušić, and V. Prelog. 1971. “Structure of Boromycin.” *Helvetica Chimica Acta* 54 (6) (January 1): 1709–1713. doi:10.1002/hlca.19710540624.

Ekborg, Nathan A., Wendy Morrill, Adam M. Burgoyne, Li Li, and Daniel L. Distel. 2007. “CelAB, a Multifunctional Cellulase Encoded by *Teredinibacter turnerae* T7902T, a Culturable Symbiont Isolated from the Wood-boring Marine Bivalve *Lyrodus pedicellatus*.” *Applied and Environmental Microbiology* 73 (23) (December): 7785–7788. doi:10.1128/AEM.00876-07.

Elbaz, Mohamed Ahmed. 2009. “Bacterial Communiaties Associated with the Sea Squirts of the Genus *Ecteinascidia* from the Red Sea and Florida”. Master’s Thesis, Oregon Health & Science University.

Elsaesser, Rebecca, and Jacques Paysan. 2004. “Liquid Gel Amplification of Complex Plasmid Libraries.” *BioTechniques* 37 (2) (August): 200, 202.

Eys, S, D Schwartz, W Wohlleben, and E Schinko. 2008. “Three Thioesterases are Involved in the Biosynthesis of Phosphinothricin Tripeptide in *Streptomyces viridochromogenes* Tü494.” *Antimicrobial Agents and Chemotherapy* 52 (5) (May): 1686–1696. doi:10.1128/AAC.01053-07.

Fang, Jie, Yiping Zhang, Lijuan Huang, Xinying Jia, Qi Zhang, Xu Zhang, Gongli Tang, and Wen Liu. 2008. “Cloning and Characterization of the Tetrocarcin A Gene Cluster from *Micromonospora chalcea* NRRL 11289 Reveals a Highly Conserved Strategy for Tetronate Biosynthesis in Spirotetronate Antibiotics.” *Journal of Bacteriology* 190 (17) (September): 6014–6025. doi:10.1128/JB.00533-08.

Fisch, Katja M, Cristian Gurgui, Nina Heycke, Sonia A van der Sar, Sally A Anderson, Victoria L Webb, Stefan Taudien, Matthias Platzer, Brent K Rubio, Sarah J Robinson, Phillip Crews, and Jörn Piel. 2009. “Polyketide Assembly Lines of Uncultivated Sponge Symbionts from Structure-based Gene Targeting.” *Nature Chemical Biology* 5 (7) (May 17): 494–501. doi:10.1038/nchembio.176.

Fisch, Katja M., Walid Bakeer, Ahmed A. Yakasai, Zhongshu Song, Jennifer Pedrick, Zahida Wasil, Andrew M. Bailey, Colin M. Lazarus, Thomas J. Simpson, and Russell J. Cox. 2011. “Rational Domain Swaps Decipher Programming in Fungal Highly Reducing Polyketide Synthases and Resurrect an Extinct Metabolite.” *Journal of the American Chemical Society* 133 (41): 16635–16641. doi:10.1021/ja206914q.

Fishlock, Dan, and Robert M. Williams. 2008. “Synthetic Studies on Et-743. Assembly of the Pentacyclic Core and a Formal Total Synthesis.” *The Journal of Organic Chemistry* 73 (24) (December 19): 9594–9600. doi:10.1021/jo801159k.

Fontana, Angelo, Paolo Cavaliere, Solimabi Wahidulla, Chandrakant G. Naik, and Guido Cimino. 2000. “A New Antitumor Isoquinoline Alkaloid from the Marine Nudibranch *Jorunna funebris*.” *Tetrahedron* 56 (37) (September 8): 7305–7308. doi:10.1016/S0040-4020(00)00629-3.

Frincke, James M., and D. John Faulkner. 1982. “Antimicrobial Metabolites of the Sponge *Reniera* Sp.” *Journal of the American Chemical Society* 104 (1): 265–269. doi:10.1021/ja00365a048.

Gab-Alla, Ali. 2008. “Distribution of the Sea Squirt *Ecteinascidia thurstoni* Herdman, 1890 (Ascidiacea: Perophoridae) along Suez Canal and Egyptian Red Sea Coasts.” *Oceanologia* 50 (2): 239–253.

Garcia, Lynne S. 2010. *Clinical Microbiology Procedures Handbook*. 3rd ed. ASM Press.

Gross, Harald, Virginia O Stockwell, Marcella D Henkels, Brian Nowak-Thompson, Joyce E Loper, and William H Gerwick. 2007. “The Genomisotopic Approach: a Systematic Method to Isolate Products of Orphan Biosynthetic Gene Clusters.” *Chemistry & Biology* 14 (1) (January): 53–63. doi:10.1016/j.chembiol.2006.11.007.

Guerriero, Antonio, Alexandre Meinesz, Michele D’Ambrosio, and Francesco Pietra. 1992. “Isolation of Toxic and Potentially Toxic Sesqui- and Monoterpenes from the Tropical Green Seaweed *Caulerpa taxifolia* which Has Invaded the Region of Cap Martin and Monaco.” *Helvetica Chimica Acta* 75 (3) (May 6): 689–695. doi:10.1002/hlca.19920750303.

Han, Andrew W. 2011. “The Chemical Biology of Siderophores from the Bacterial Endosymbionts of Shipworms”. Ph.D Dissertation, Oregon Health & Science University.

Harland, D. N., E. Dassa, R. W. Titball, K. A. Brown, and H. S. Atkins. 2007. “ATP-binding Cassette Systems in *Burkholderia pseudomallei* and *Burkholderia mallei*.” *BMC Genomics* 8 (March 28): 83.

Harland, David N., Helen S. Garmory, Katherine A. Brown, and Richard W. Titball. 2005. “An Association Between ATP Binding Cassette Systems, Genome Sizes and Lifestyles of Bacteria.” *Research in Microbiology* 156 (3) (April): 434–442. doi:10.1016/j.resmic.2004.12.002.

Harris, Wesley R, Shady A Amin, Frithjof C Küpper, David H Green, and Carl J Carrano. 2007. “Borate Binding to Siderophores: Structure and Stability.” *Journal of the American Chemical Society* 129 (40) (September 10): 12263–12271. doi:10.1021/ja073788v.

Hemscheidt, Thomas, Melany P. Puglisi, Linda K. Larsen, Gregory M. L. Patterson, Richard E. Moore, Jorge L. Rios, and Jon Clardy. 1994. “Structure and Biosynthesis of Borophycin, a New Boeseken Complex of Boric Acid from a Marine Strain of the Blue-green Alga *Nostoc linckia*.”*The Journal of Organic Chemistry* 59 (12): 3467–3471. doi:10.1021/jo00091a042.

Higgins, C. F. 1992. “ABC Transporters: From Microorganisms to Man.” *Annual Review of Cell Biology* 8: 67–113.

Hildebrand, M., L. E. Waggoner, G. E. Lim, K. H. Sharp, C. P. Ridley, and M. G. Haygood. 2004. “Approaches to Identify, Clone, and Express Symbiont Bioactive Metabolite Genes.” *Natural Product Reports* 21 (1): 122–142.

Hildebrand, Mark, Laura E Waggoner, Haibin Liu, Sebastian Sudek, Scott Allen, Christine Anderson, David H Sherman, and Margo Haygood. 2004. “*bryA*: an Unusual Modular Polyketide Synthase Gene from the Uncultivated Bacterial Symbiont of the Marine Bryozoan *Bugula neritina*.” *Chemistry & Biology* 11 (11) (October): 1543–1552. doi:10.1016/j.chembiol.2004.08.018.

Hoover, Cindi A, Marc Slattery, and Adam G Marsh. 2007. “A Functional Approach to Transcriptome Profiling: Linking Gene Expression Patterns to Metabolites that Matter.” *Marine Biotechnology (New York, N.Y.)* 9 (4) (August): 411–419. doi:10.1007/s10126-007-9008-2.

Hotopp, Julie C. Dunning, Michael E. Clark, Deodoro C. S. G. Oliveira, Jeremy M. Foster, Peter Fischer, Mónica C. Muñoz Torres, Jonathan D. Giebel, Nikhil Kumar, Nadeeza Ishmael Shiliang Wang, Jessica Ingram, Rahul V. Nene, Jessica Shepard, Jeffrey Tomkins, Stephen Richards, David J. Spiro, Elodie Ghedin, Barton E. Slatko,Hervé Tettelin, John H. Werren 2007. “Widespread Lateral Gene Transfer from Intracellular Bacteria to Multicellular Eukaryotes.” *Science* 317 (5845): 1753 –1756. doi:10.1126/science.1142490.

Hou, Jie, Lars Robbel, and Mohamed A Marahiel. 2011. “Identification and Characterization of the Lysobactin Biosynthetic Gene Cluster Reveals Mechanistic Insights into an Unusual Termination Module Architecture.” *Chemistry & Biology* 18 (5) (May 27): 655–664. doi:10.1016/j.chembiol.2011.02.012.

Hrvatin, Sinisa, and Jörn Piel. 2007. “Rapid Isolation of Rare Clones from Highly Complex DNA Libraries by PCR Analysis of Liquid Gel Pools.” *Journal of Microbiological Methods* 68 (2) (February): 434–436. doi:10.1016/j.mimet.2006.09.009.

Hütter, R, W Keller-Schierlein, F Knüsel, V Prelog, G C Rodgers Jr, P Suter, G Vogel, W Voser, and H Zähner. 1967. “The Metabolic Products of Microorganisms. Boromycin.” *Helvetica Chimica Acta* 50 (6): 1533–1539. doi:10.1002/hlca.19670500612.

Ikeda, Y, H Idemoto, F Hirayama, K Yamamoto, K Iwao, T Asao, and T Munakata. 1983. "Safracins, New Antitumor Antibiotics. I. Producing Organism, Fermentation and Isolation." *The Journal of Antibiotics* 36 (10) (October): 1279–1283.

Imhoff, Johannes F. 1986. "Osmoregulation and Compatible Solutes in Eubacteria." *FEMS Microbiology Letters* 39 (1–2) (July): 57–66. doi:10.1016/0378-1097(86)90061-3.

Irschik, H, D Schummer, K Gerth, G Höfle, and H Reichenbach. 1995. "The Tartrolons, New Boron-containing Antibiotics from a Myxobacterium, *Sorangium cellulosum*." *The Journal of Antibiotics* 48 (1) (January): 26–30.

Irschik, H, W Trowitzsch-Kienast, K Gerth, G Höfle, and H Reichenbach. 1988. "Saframycin Mx1, a New Natural Saframycin Isolated from a Myxobacterium." *The Journal of Antibiotics* 41 (8) (August): 993–998.

Ishida, Keishi, Thorger Lincke, Swantje Behnken, and Christian Hertweck. 2010. "Induced Biosynthesis of Cryptic Polyketide Metabolites in a *Burkholderia thailandensis* Quorum Sensing Mutant." *Journal of the American Chemical Society* 132 (40): 13966–13968. doi:10.1021/ja105003g.

Jasinski, M., E. Ducos, E. Martinoia, and M. Boutry. 2003. "The ATP-binding Cassette Transporters: Structure, Function, and Gene Family Comparison Between Rice and Arabidopsis." *Plant Physiology* 131 (3) (March): 1169–1177.

Jebbar, M., R. Talibart, K. Gloux, T. Bernard, and C. Blanco. 1992. "Osmoprotection of *Escherichia coli* by Ectoine: Uptake and Accumulation Characteristics." *Journal of Bacteriology* 174 (15) (August 1): 5027–5035.

Kallio, Pauli, Azmiri Sultana, Jarmo Niemi, Pekka Mäntsälä, and Gunter Schneider. 2006. "Crystal Structure of the Polyketide Cyclase AknH with Bound Substrate and Product Analogue: Implications for Catalytic Mechanism and Product Stereoselectivity." *Journal of Molecular Biology* 357 (1) (March 17): 210–220. doi:10.1016/j.jmb.2005.12.064.

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." *The Journal of Antibiotics* 35 (2) (February): 142–150.

Kerr, Russell G., Joseph Lawry, and Kim A. Gush. 1996. "In Vitro Biosynthetic Studies of the Bryostatins, Anti-cancer Agents from the Marine Bryozoan *Bugula neritina*."

*Tetrahedron Letters* 37 (46) (October 11): 8305–8308. doi:10.1016/0040-4039(96)01943-0.

Kohno, J, T Kawahata, T Otake, M Morimoto, H Mori, N Ueba, M Nishio, A Kinumaki, S Komatsubara, and K Kawashima. 1996. "Boromycin, an anti-HIV Antibiotic." *Bioscience, Biotechnology, and Biochemistry* 60 (6) (June): 1036–1037.

Koike, I, M Yamamuro, and PC Pollard. 1993. "Carbon and Nitrogen Budgets of Two Ascidians and Their Symbiont, *Prochloron*, in a Tropical Seagrass Meadow." *Marine and Freshwater Research* 44 (1) (January 1): 173–182.

Koike, Isao, and Takao Suzuki. 1996. "Nutritional Diversity of Symbiotic Ascidians in a Fijian Seagrass Meadow." *Ecological Research* 11 (3) (December): 381–386. doi:10.1007/BF02347795.

Kroiss, Johannes, Martin Kaltenpoth, Bernd Schneider, Maria-Gabriele Schwinger, Christian Hertweck, Ravi Kumar Maddula, Erhard Strohm, and Ales Svatos. 2010. "Symbiotic Streptomycetes Provide Antibiotic Combination Prophylaxis for Wasp Offspring." *Nature Chemical Biology* 6 (4) (April): 261–263. doi:10.1038/nchembio.331.

Kwan, Jason. 2012. "A New Family of Ascidian Symbionts Make Patellazoles and Other Complex Polyketides." Presented at the Marine Natural Products Gordon Research Conferences , Ventura, California, Feb 26-March 2.

Laureti, L., L. Song, S. Huang, C. Corre, P. Leblond, G. L. Challis, and B. Aigle. 2011. "Identification of a Bioactive 51-membered Macrolide Complex by Activation of a Silent Polyketide Synthase in *Streptomyces ambofaciens*." *Proceedings of the National Academy of Sciences* 108 (15) (March 28): 6258–6263. doi:10.1073/pnas.1019077108.

Lechene, Claude P, Yvette Luyten, Gregory McMahon, and Daniel L Distel. 2007. "Quantitative Imaging of Nitrogen Fixation by Individual Bacteria Within Animal Cells." *Science (New York, N.Y.)* 317 (5844) (September 14): 1563–1566. doi:10.1126/science.1145557.

Lee, J. J., P. M. Dewick, C. P. Gorst-Allman, F. Spreafico, C. Kowal, C. J. Chang, A. G. McInnes, J. A. Walter, P. J. Keller, and H. G. Floss. 1987. "Further Studies on the Biosynthesis of the Boron-containing Antibiotic Aplasmomycin." *Journal of the American Chemical Society* 109 (18): 5426–5432. doi:10.1021/ja00252a019.

Lesnik, Elena A., Rangarajan Sampath, Harold B. Levene, Timothy J. Henderson, John A. McNeil, and David J. Ecker. 2001. “Prediction of Rho-independent Transcriptional Terminators in *Escherichia coli*.” *Nucleic Acids Research* 29 (17) (September 1): 3583.

Lewer, Paul, Eleanor L Chapin, Paul R Graupner, Jeffrey R Gilbert, and Cathy Peacock. 2003. “Tartrolone C: a Novel Insecticidal Macrodiolide Produced by *Streptomyces* Sp. CP1130.” *Journal of Natural Products* 66 (1) (January): 143–145. doi:10.1021/np020451s.

Li, L., W. Deng, J. Song, W. Ding, Q. F. Zhao, C. Peng, W. W. Song, G. L. Tang, and W. Liu. 2008. “Characterization of the Saframycin A Gene Cluster from *Streptomyces lavendulae* NRRL 11002 Revealing a Nonribosomal Peptide Synthetase System for Assembling the Unusual Tetrapeptidyl Skeleton in an Iterative Manner.” *Journal of Bacteriology* 190 (1): 251–263.

Light, Alan R, Ronald W Hughen, Jie Zhang, Jon Rainier, Zhuqing Liu, and Jeewoo Lee. 2008. “Dorsal Root Ganglion Neurons Innervating Skeletal Muscle Respond to Physiological Combinations of Protons, ATP, and Lactate Mediated by ASIC, P2X, and TRPV1.” *Journal of Neurophysiology* 100 (3) (September 1): 1184–1201. doi:10.1152/jn.01344.2007.

Liles, M. R., L. L. Williamson, J. Rodbumrer, V. Torsvik, R. M. Goodman, and J. Handelsman. 2008. “Recovery, Purification, and Cloning of High-molecular-weight DNA from Soil Microorganisms.” *Applied and Environmental Microbiology* 74 (10) (April): 3302–3305.

Lin, Zhenjian, Christopher A. Reilly, Rowena Antemano, Ronald W. Hughen, Lenny Marett, Gisela P. Concepcion, Margo G. Haygood, Baldomero M. Olivera, Alan Light, and Eric W. Schmidt. 2011. “Nobilamides A–H, Long-acting Transient Receptor Potential Vanilloid-1 (TRPV1) Antagonists from Mollusk-associated Bacteria.” *Journal of Medicinal Chemistry* 54 (11): 3746–3755. doi:10.1021/jm101621u.

Linton, K. J., and C. F. Higgins. 1998. “The *Escherichia coli* ATP-binding Cassette (ABC) Proteins.” *Molecular Microbiology* 28 (1) (April): 5–13.

Long, Paul F, Walter C Dunlap, Christopher N Battershill, and Marcel Jaspars. 2005. “Shotgun Cloning and Heterologous Expression of the Patellamide Gene Cluster as a Strategy to Achieving Sustained Metabolite Production.” *Chembiochem: A*

*European Journal of Chemical Biology* 6 (10) (October): 1760–1765. doi:10.1002/cbic.200500210.

Lopanik, Nicole, Kirk R. Gustafson, and Niels Lindquist. 2004a. "Structure of Bryostatin 20: A Symbiont-produced Chemical Defense for Larvae of the Host Bryozoan, *Bugula neritina*." *Journal of Natural Products* 67 (8): 1412–1414. doi:10.1021/np040007k.

Lopanik, Nicole, Niels Lindquist, and Nancy Targett. 2004b. "Potent Cytotoxins Produced by a Microbial Symbiont Protect Host Larvae from Predation." *Oecologia* 139 (1) (February 15): 131–139. doi:10.1007/s00442-004-1487-5.

Marschner, Petra. 2011. *Marschner's Mineral Nutrition of Higher Plants*. Academic Press: London.

Martín, Juan F. 2004. "Phosphate Control of the Biosynthesis of Antibiotics and Other Secondary Metabolites is Mediated by the PhoR-PhoP System: an Unfinished Story." *Journal of Bacteriology* 186 (16): 5197–5201. doi:10.1128/JB.186.16.5197-5201.2004.

Masumi, Yamagishi, and Yamamoto Yukio. 1994. "Effects of Boron on Nodule Development and Symbiotic Nitrogen Fixation in Soybean Plants." *Soil Science and Plant Nutrition* 40 (2) (June): 265–274.

Mattheus, Wesley, Ling-Jie Gao, Piet Herdewijn, Bart Landuyt, Jan Verhaegen, Joleen Masschelein, Guido Volckaert, and Rob Lavigne. 2010. "Isolation and Purification of a New Kalimantacin/batumin-related Polyketide Antibiotic and Elucidation of Its Biosynthesis Gene Cluster." *Chemistry & Biology* 17 (2) (February 26): 149–159. doi:10.1016/j.chembiol.2010.01.014.

McGovern, Tamara M, and Michael E Hellberg. 2003. "Cryptic Species, Cryptic Endosymbionts, and Geographical Variation in Chemical Defences in the Bryozoan *Bugula neritina*." *Molecular Ecology* 12 (5) (May): 1207–1215.

Medema, M. H., K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski, M. A. Fischbach, T. Weber, E. Takano, and R. Breitling. 2011. "antiSMASH: Rapid Identification, Annotation and Analysis of Secondary Metabolite Biosynthesis Gene Clusters in Bacterial and Fungal Genome Sequences." *Nucleic Acids Research* 39 (Web Server) (May 14): W339–W346. doi:10.1093/nar/gkr466.

Mesak, Lili R, Shuhua Qi, Ivan Villanueva, Vivian Miao, and Julian Davies. 2010. "*Staphylococcus aureus* Promoter-lux Reporters for Drug Discovery." *The Journal of Antibiotics* 63 (8) (July): 492–498. doi:10.1038/ja.2010.74.

Meyer, F, D Paarmann, M D'Souza, R Olson, E M Glass, M Kubal, T Paczian, A Rodriguez, R Stevens, A Wilke, J Wilkening, and R A Edwards 2008. "The Metagenomics RAST Server - a Public Resource for the Automatic Phylogenetic and Functional Analysis of Metagenomes." *BMC Bioinformatics* 9: 386. doi:10.1186/1471-2105-9-386.

Miller, Stephen T., Karina B. Xavier, Shawn R. Campagna, Michiko E. Taga, Martin F. Semmelhack, Bonnie L. Bassler, and Frederick M. Hughson. 2004. "*Salmonella typhimurium* Recognizes a Chemically Distinct Form of the Bacterial Quorum-sensing Signal AI-2." *Molecular Cell* 15 (5) (September 10): 677–687. doi:10.1016/j.molcel.2004.07.020.

Milton, D L, R O'Toole, P Horstedt, and H Wolf-Watz. 1996. "Flagellin A Is Essential for the Virulence of *Vibrio anguillarum*." *Journal of Bacteriology* 178 (5) (March): 1310–1319.

Morris, P. F., and V. Phuntumart. 2009. "Inventory and Comparative Evolution of the ABC Superfamily in the Genomes of *Phytophthora ramorum* and *Phytophthora sojae*." *Journal of Molecular Evolution* 68 (5) (May): 563–575.

Mosmann, T. 1983. "Rapid Colorimetric Assay for Cellular Growth and Survival: Application to Proliferation and Cytotoxicity Assays." *Journal of Immunological Methods* 65 (1-2) (December 16): 55–63.

Moss, C., D. H. Green, B. Pérez, A. Velasco, R. Henríquez, and J. D. McKenzie. 2003. "Intracellular Bacteria Associated with the Ascidian *Ecteinascidia turbinata*: Phylogenetic and In Situ Hybridisation Analysis." *Marine Biology* 143 (1): 99–110.

Nakamura, H, Y Iitaka, T Kitahara, T Okazaki, and Y Okami. 1977. "Structure of Aplasmomycin." *The Journal of Antibiotics* 30 (9) (September): 714–719.

Nelson, J. T., J. Lee, J. W. Sims, and E. W. Schmidt. 2007. "Characterization of SafC, a Catechol 4-O-methyltransferase Involved in Saframycin Biosynthesis." *Applied and Environmental Microbiology* 73 (11) (May): 3575–3580.

Newman, Leslie J., Jon L. Norenburg, and Sherry Reed. 2000. "Taxonomic and Biological Observations on the Tiger Flatworm, *Maritigrella crozieri* (Hyman, 1939), New Combination (Platyhelminthes, Polycladida, Euryleptidae) from Florida Waters." *Journal of Natural History* 34 (6): 799–808. doi:10.1080/002229300299264.

Nguyen, TuAnh, Keishi Ishida, Holger Jenke-Kodama, Elke Dittmann, Cristian Gurgui, Thomas Hochmuth, Stefan Taudien, Matthias Platzer, Christian Hertweck, and Jörn Piel. 2008. “Exploiting the Mosaic Structure of Trans-acyltransferase Polyketide Synthases for Natural Product Discovery and Pathway Dissection.” *Nature Biotechnology* 26 (2) (January): 225–233. doi:10.1038/nbt1379.

Nielsen, F H. 2000. “The Emergence of Boron as Nutritionally Important Throughout the Life Cycle.” *Nutrition (Burbank, Los Angeles County, Calif.)* 16 (7-8) (August): 512–514.

O’Neill, M A, D Warrenfeltz, K Kates, P Pellerin, T Doco, A G Darvill, and P Albersheim. 1996. “Rhamnogalacturonan-II, a Pectic Polysaccharide in the Walls of Growing Plant Cell, Forms a Dimer that is Covalently Cross-linked by a Borate Ester. In Vitro Conditions for the Formation and Hydrolysis of the Dimer.” *The Journal of Biological Chemistry* 271 (37) (September 13): 22923–22930.

Okami, Y, T Okazaki, T Kitahara, and H Umezawa. 1976. “Studies on Marine Microorganisms. V. A New Antibiotic, Aplasmomycin, Produced by a Streptomycete Isolated from Shallow Sea Mud.” *The Journal of Antibiotics* 29 (10) (October): 1019–1025.

Olivera, B.M. 2000. “ω-omega;-Conotoxin MVIIA: From Marine Snail Venom to Analgesic Drug.” In *Drugs from the Sea*, ed. N. Fusetani, Karger Publisher, Basel, Switzerland 74–85.

Partida-Martinez, Laila P, Ingrid Groth, Imke Schmitt, Walter Richter, Martin Roth, and Christian Hertweck. 2007. “*Burkholderia rhizoxinica* Sp. Nov. and *Burkholderia endofungorum* Sp. Nov., Bacterial Endosymbionts of the Plant-pathogenic Fungus *Rhizopus microsporus*.” *International Journal of Systematic and Evolutionary Microbiology* 57 (Pt 11) (November): 2583–2590. doi:10.1099/ijs.0.64660-0.

Paungmoung, P., J. Punya, S. Pongpattanakitshote, W. Jeamton, T. Vichisoonthonkul, S. Bhumiratana, M. Tanticharoen, U. Linne, M. A. Marahiel, and S. Cheevadhanarak. 2007. “Detection of Nonribosomal Peptide Synthetase Genes in *Xylaria* Sp. BCC1067 and Cloning of XyNRPSA.” *FEMS Microbiology Letters* 274 (2) (September): 260–268.

Pawlik, Joseph R. 1993. “Marine Invertebrate Chemical Defenses.” *Chemical Reviews* 93 (5): 1911–1922. doi:10.1021/cr00021a012.

Peraud, Olivier, Jason S Biggs, Ronald W Hughen, Alan R Light, Gisela P Concepcion, Baldomero M Olivera, and Eric W Schmidt. 2009. “Microhabitats Within Venomous Cone Snails Contain Diverse Actinobacteria.” *Applied and Environmental Microbiology* 75 (21) (November): 6820–6826. doi:10.1128/AEM.01238-09.

Pérez, Marta, Cristina Crespo, Carmen Schleissner, Pilar Rodríguez, Paz Zúñiga, and Fernando Reyes. 2009. “Tartrolon D, a Cytotoxic Macrodiolide from the Marine-derived Actinomycete *Streptomyces* Sp. MDG-04-17-069.” *Journal of Natural Products* 72 (12) (December): 2192–2194. doi:10.1021/np9006603.

Perez-Brocal, V., R. Gil, S. Ramos, A. Lamelas, M. Postigo, J. M. Michelena, F. J. Silva, A. Moya, and A. Latorre. 2006. “A Small Microbial Genome: The End of a Long Symbiotic Relationship?” *Science (New York, N.Y.)* 314 (5797) (October 13): 312–313.

Perez-Matos, A. E., W. Rosado, and N. S. Govind. 2007. “Bacterial Diversity Associated with the Caribbean Tunicate *Ecteinascidia turbinata*.” *Antonie Van Leeuwenhoek* 92 (2) (July): 155–164.

Perlova, Olena, Klaus Gerth, Olaf Kaiser, Astrid Hans, and Rolf Müller. 2006. “Identification and Analysis of the Chivosazol Biosynthetic Gene Cluster from the Myxobacterial Model Strain *Sorangium cellulosum* So Ce56.” *Journal of Biotechnology* 121 (2) (January 24): 174–191. doi:10.1016/j.jbiotec.2005.10.011.

Piel, Jörn. 2002. “A Polyketide Synthase-peptide Synthetase Gene Cluster from an Uncultured Bacterial Symbiont of *Paederus* Beetles.” *Proceedings of the National Academy of Sciences* 99 (22) (October 29): 14002–14007. doi:10.1073/pnas.222481399.

Pommier, Y, G Kohlhagen, C Bailly, M Waring, A Mazumder, and K W Kohn. 1996. “DNA Sequence- and Structure-selective Alkylation of Guanine N2 in the DNA Minor Groove by Ecteinascidin 743, a Potent Antitumor Compound from the Caribbean Tunicate *Ecteinascidia turbinata*.” *Biochemistry* 35 (41) (October 15): 13303–13309. doi:10.1021/bi960306b.

Pospiech, A., B. Cluzel, J. Bietenhader, and T. Schupp. 1995. “A New *Myxococcus xanthus* Gene Cluster for the Biosynthesis of the Antibiotic Saframycin Mx1 Encoding a Peptide Synthetase.” *Microbiology (Reading, England)* 141 ( Pt 8) (Pt 8) (July): 1793–1803.

Quentin, Y., G. Fichant, and F. Denizot. 1999. “Inventory, Assembly and Analysis of *Bacillus subtilis* ABC Transport Systems.” *Journal of Molecular Biology* 287 (3) (April 2): 467–484.

Rath, Christopher M, Benjamin Janto, Josh Earl, Azad Ahmed, Fen Z Hu, Luisa Hiller, Meg Dahlgren, Rachael Kreft, Fengan Yu, Jeremy J. Wolff, Hye Kyong Kweon, Michael A. Christiansen, Kristina Håkansson, Robert M. Williams, Garth D. Ehrlich, and David H. Sherman 2011. “Meta-omic Characterization of the Marine Invertebrate Microbial Consortium that Produces the Chemotherapeutic Natural Product ET-743.” *ACS Chemical Biology* 6 (11) (November 18): 1244–1256. doi:10.1021/cb200244t.

Ren, Q., K. H. Kang, and I. T. Paulsen. 2004. “TransportDB: a Relational Database of Cellular Membrane Transport Systems.” *Nucleic Acids Research* 32 (Database issue) (January 1): D284–8.

Richardson, Adam D, William Aalbersberg, and Chris M Ireland. 2005. “The Patellazoles Inhibit Protein Synthesis at Nanomolar Concentrations in Human Colon Tumor Cells.” *Anti-Cancer Drugs* 16 (5) (June): 533–541.

Riley, Margaret A, and John E Wertz. 2002. “Bacteriocins: Evolution, Ecology, and Application.” *Annual Review of Microbiology* 56: 117–137. doi:10.1146/annurev.micro.56.012302.161024.

Rinehart, K. L. 2000. “Antitumor Compounds from Tunicates.” Medicinal Research Reviews 20 (1): 1–27.

Rinehart, K.L, and A. M. Lithgow-Bertelloni. Dehydrodidemnin B. US Patent 5834586, filed Jul 25, 1994, and issued Nov 10, 1998.

Rinehart, Kenneth L., Tom G. Holt, Nancy L. Fregeau, Justin G. Stroh, Paul A. Keifer, Furong Sun, Li H. Li, and David G. Martin. 1990. “Ecteinascidins 729, 743, 745, 759A, 759B, and 770: Potent Antitumor Agents from the Caribbean Tunicate *Ecteinascidia turbinata*.” *Journal of Organic Chemistry* 55 (15): 4512–4515.

Saier, M. H., C. V. Tran, and R. D. Barabote. 2006. “TCDB: The Transporter Classification Database for Membrane Transport Protein Analyses and Information.” *Nucleic Acids Research* 34 (Database issue) (January 1): D181–6.

Sakai, R, K L Rinehart, Y Guan, and A H Wang. 1992. “Additional Antitumor Ecteinascidins from a Caribbean Tunicate: Crystal Structures and Activities in Vivo.”

*Proceedings of the National Academy of Sciences of the United States of America* 89 (23) (December 1): 11456–11460.

Salomon, C. E. 2002. "Localization of Bioactive Secondary Metabolites in Sponges, Ascidians and Their Associated Symbionts". Ph.D Dissertation. University of California, San Diego.

Sambrook, J., and D. W. Russell. 2001. *Molecular Cloning: a Laboratory Manual*. CSHL press, Long Island, New York.

Sarker, Satyajit D, Lutfun Nahar, and Yashodharan Kumarasamy. 2007. "Microtitre Plate-based Antibacterial Assay Incorporating Resazurin as an Indicator of Cell Growth, and Its Application in the In Vitro Antibacterial Screening of Phytochemicals." *Methods (San Diego, Calif.)* 42 (4) (August): 321–324. doi:10.1016/j.ymeth.2007.01.006.

Saurin, W., M. Hofnung, and E. Dassa. 1999. "Getting In or Out: Early Segregation Between Importers and Exporters in the Evolution of ATP-binding Cassette (ABC) Transporters." *Journal of Molecular Evolution* 48 (1) (January): 22–41.

Schmerk, Crystal L., Mark A. Bernards, and Miguel A. Valvano. 2011. "Hopanoid Production Is Required for Low pH Tolerance, Antimicrobial Resistance, and Motility in *Burkholderia cenocepacia*." *Journal of Bacteriology*. doi:10.1128/JB.05979-11. http://jb.asm.org/content/early/2011/09/30/JB.05979-11.

Schmidt, Eric W. 2008. "Trading Molecules and Tracking Targets in Symbiotic Interactions." *Nature Chemical Biology* 4 (8) (August): 466–473. doi:10.1038/nchembio.101.

Schmidt, Eric W. 2011. "An Enzymatic Route to Sunscreens." *Chembiochem: A European Journal of Chemical Biology* 12 (3) (January 11): 363–365. doi:10.1002/cbic.201000709.

Schmidt, Eric W, Mohamed S Donia, John A McIntosh, W Florian Fricke, and Jacques Ravel. 2012. "Origin and Variation of Tunicate Secondary Metabolites." *Journal of Natural Products* (January 10). doi:10.1021/np200665k. http://www.ncbi.nlm.nih.gov/pubmed/22233390.

Schmidt, Eric W, James T Nelson, David A Rasko, Sebastian Sudek, Jonathan A Eisen, Margo G Haygood, and Jacques Ravel. 2005. "Patellamide A and C Biosynthesis by a Microcin-like Pathway in *Prochloron didemni*, the Cyanobacterial Symbiont of

*Lissoclinum patella*." *Proceedings of the National Academy of Sciences of the United States of America* 102 (20) (May 17): 7315–7320. doi:10.1073/pnas.0501424102.

Schmidt, Eric W., Sebastian Sudek, and Margo G. Haygood. 2004. "Genetic Evidence Supports Secondary Metabolic Diversity in *Prochloron* Spp., the Cyanobacterial Symbiont of a Tropical Ascidian." *Journal of Natural Products* 67 (8): 1341–1345. doi:10.1021/np049948n.

Schneider, Kathrin, Xiao-Hua Chen, Joachim Vater, Peter Franke, Graeme Nicholson, Rainer Borriss, and Roderich D. Süssmuth. 2007. "Macrolactin is the Polyketide Biosynthesis Product of the Pks2 Cluster of *Bacillus amyloliquefaciens* FZB42." *Journal of Natural Products* 70 (9): 1417–1423. doi:10.1021/np070070k.

Schoenian, Ilka, Michael Spiteller, Manoj Ghaste, Rainer Wirth, Hubert Herz, and Dieter Spiteller. 2011. "Chemical Basis of the Synergism and Antagonism in Microbial Communities in the Nests of Leaf-cutting Ants." *Proceedings of the National Academy of Sciences of the United States of America* 108 (5) (February 1): 1955–1960. doi:10.1073/pnas.1008441108.

Schröder, Heinz C, Hans J Breter, Ernesto Fattorusso, Hiroshi Ushijima, Matthias Wiens, Renate Steffen, Renato Batel, and Werner E G Müller. 2006. "Okadaic Acid, an Apoptogenic Toxin for Symbiotic/parasitic Annelids in the Demosponge *Suberites domuncula*." *Applied and Environmental Microbiology* 72 (7) (July): 4907–4916. doi:10.1128/AEM.00228-06.

Schummer, Dietmar, Herbert Irschik, Hans Reichenbach, and Gerhard Höfle. 1994. "Antibiotics from Gliding Bacteria, LVII. Tartrolons: New Boron-containing Macrodiolides from *Sorangium cellulosum*." *Liebigs Annalen Der Chemie* 1994 (3) (March 11): 283–289. doi:10.1002/jlac.199419940310.

Schwyn, B, and J B Neilands. 1987. "Universal Chemical Assay for the Detection and Determination of Siderophores." *Analytical Biochemistry* 160 (1) (January): 47–56.

Scott, Jarrod J, Dong-Chan Oh, M Cetin Yuceer, Kier D Klepzig, Jon Clardy, and Cameron R Currie. 2008. "Bacterial Protection of Beetle-fungus Mutualism." *Science (New York, N.Y.)* 322 (5898) (October 3): 63. doi:10.1126/science.1160423.

Seyedsayamdost, Mohammad R, Josephine R Chandler, Joshua A V Blodgett, Patricia S Lima, Breck A Duerkop, Ken-Ichi Oinuma, E Peter Greenberg, and Jon Clardy. 2010.

"Quorum-sensing-regulated Bactobolin Production by *Burkholderia thailandensis* E264." *Organic Letters* 12 (4) (February 19): 716–719. doi:10.1021/ol902751x.

Shenkar, Noa, and Billie J. Swalla. 2011. "Global Diversity of Ascidiacea." *PLoS ONE* 6 (6) (June 20): e20657. doi:10.1371/journal.pone.0020657.

Siefert, J L, M Larios-Sanz, L K Nakamura, R A Slepecky, J H Paul, E R Moore, G E Fox, and P Jurtshuk Jr. 2000. "Phylogeny of Marine *Bacillus* Isolates from the Gulf of Mexico." *Current Microbiology* 41 (2) (August): 84–88.

Simon, R., U. Priefer, and A. Puhler. 1983. "A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria." *Nature Biotechnology* 1 (9) (October): 784–791. doi:10.1038/nbt1183-784.

Slattery, M., G. A. Hines, J. Starmer, and V. J. Paul. 1999. "Chemical Signals in Gametogenesis, Spawning, and Larval Settlement and Defense of the Soft Coral *Sinularia polydactyla*." *Coral Reefs* 18 (1) (April 22): 75–84. doi:10.1007/s003380050158.

Slattery, M., J. Starmer, and V. Paul. 2001. "Temporal and Spatial Variation in Defensive Metabolites of the Tropical Pacific Soft Corals *Sinularia maxima* and *S. polydactyla*." *Marine Biology* 138 (6): 1183–1193.

Smyrniotopoulos, Vangelis, Dennis Abatis, Leto-A. Tziveleka, Christina Tsitsimpikou, Vassilios Roussis, Anargyros Loukis, and Constantinos Vagias. 2002. "Acetylene Sesquiterpenoid Esters from the Green Alga *Caulerpa prolifera*." *Journal of Natural Products* 66 (1): 21–24. doi:10.1021/np0202529.

Stachelhaus, T, H D Mootz, and M A Marahiel. 1999. "The Specificity-conferring Code of Adenylation Domains in Nonribosomal Peptide Synthetases." *Chemistry & Biology* 6 (8) (August): 493–505. doi:10.1016/S1074-5521(99)80082-9.

Sudek, Sebastian, Margo G Haygood, Diaa T A Youssef, and Eric W Schmidt. 2006. "Structure of Trichamide, a Cyclic Peptide from the Bloom-forming Cyanobacterium *Trichodesmium erythraeum*, Predicted from the Genome Sequence." *Applied and Environmental Microbiology* 72 (6) (June): 4382–4387. doi:10.1128/AEM.00380-06.

Sudek, Sebastian, Nicole B Lopanik, Laura E Waggoner, Mark Hildebrand, Christine Anderson, Haibin Liu, Amrish Patel, David H Sherman, and Margo G Haygood. 2007. "Identification of the Putative Bryostatin Polyketide Synthase Gene Cluster from '*Candidatus* Endobugula sertula', the Uncultivated Microbial Symbiont of the

Marine Bryozoan *Bugula neritina*." *Journal of Natural Products* 70 (1): 67–74. doi:10.1021/np060361d.

Suwanborirux, K., K. Charupant, S. Amnuoypol, S. Pummangura, A. Kubo, and N. Saito. 2002. "Ecteinascidins 770 and 786 from the Thai Tunicate *Ecteinascidia thurstoni*." *Journal of Natural Products* 65 (6) (May): 935–937.

Suwanborirux, Khanit, Surattana Amnuoypol, Anuchit Plubrukarn, Sunibhond Pummangura, Akinori Kubo, Chieko Tanaka, and Naoki Saito. 2003. "Chemistry of Renieramycins. Part 3.(1) Isolation and Structure of Stabilized Renieramycin Type Derivatives Possessing Antitumor Activity from Thai Sponge *Xestospongia* Species, Pretreated with Potassium Cyanide." *Journal of Natural Products* 66 (11) (November): 1441–1446. doi:10.1021/np030262p.

Tachibana, Kazuo, Paul J. Scheuer, Yasumasa Tsukitani, Hiroyuki Kikuchi, Donna Van Engen, Jon Clardy, Yalamanchili Gopichand, and Francis J. Schmitz. 1981. "Okadaic Acid, a Cytotoxic Polyether from Two Marine Sponges of the Genus *Halichondria*." *Journal of the American Chemical Society* 103 (9): 2469–2471. doi:10.1021/ja00399a082.

Takano, Junpei, Kyotaro Noguchi, Miho Yasumori, Masaharu Kobayashi, Zofia Gajdos, Kyoko Miwa, Hiroaki Hayashi, Tadakatsu Yoneyama, and Toru Fujiwara. 2002. "Arabidopsis Boron Transporter for Xylem Loading." *Nature* 420 (6913) (November 21): 337–340. doi:10.1038/nature01139.

Takano, Junpei, Mayuki Tanaka, Atsushi Toyoda, Kyoko Miwa, Koji Kasai, Kentaro Fuji, Hitoshi Onouchi, Satoshi Naito, and Toru Fujiwara. 2010. "Polar Localization and Degradation of *Arabidopsis* Boron Transporters Through Distinct Trafficking Pathways." *Proceedings of the National Academy of Sciences of the United States of America* 107 (11) (March 16): 5220–5225. doi:10.1073/pnas.0910744107.

Takebayashi, Y, P Pourquier, D B Zimonjic, K Nakayama, S Emmert, T Ueda, Y Urasaki, A Kanzaki, S Akiyama, N Popescu, K Kraemer, Y Pommier. 2001. "Antiproliferative Activity of Ecteinascidin 743 is Dependent Upon Transcription-coupled Nucleotide-excision Repair." *Nature Medicine* 7 (8) (August): 961–966. doi:10.1038/91008.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. "MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0." *Molecular Biology and Evolution* 24 (8) (August): 1596–1599.

Tang, L, S Shah, L Chung, J Carney, L Katz, C Khosla, and B Julien. 2000. “Cloning and Heterologous Expression of the Epothilone Gene Cluster.” *Science (New York, N.Y.)* 287 (5453) (January 28): 640–642.

Thomas, Jacob, and John E Cronan. 2005. “The Enigmatic Acyl Carrier Protein Phosphodiesterase of *Escherichia coli*: Genetic and Enzymological Characterization.” *The Journal of Biological Chemistry* 280 (41) (October 14): 34675–34683. doi:10.1074/jbc.M505736200.

Tomii, K., and M. Kanehisa. 1998. “A Comparative Analysis of ABC Transporters in Complete Microbial Genomes.” *Genome Research* 8 (10) (October): 1048–1059.

Toyoshima, M., N. V. Sasaki, M. Fujiwara, S. Ehira, M. Ohmori, and N. Sato. 2010. “Early Candidacy for Differentiation into Heterocysts in the Filamentous Cyanobacterium *Anabaena* Sp. PCC 7120.” *Archives of Microbiology* 192 (1) (January): 23–31.

Trindade-Silva, Amaro E., Erik Machado-Ferreira, Marcus V. X. Senra, Vinicius F. Vizzoni, Luciana A. Yparraguirre, Orilio Leoncini, and Carlos A. G. Soares. 2009. “Physiological Traits of the Symbiotic Bacterium *Teredinibacter turnerae* Isolated from the Mangrove Shipworm *Neoteredo reynei*.” *Genetics and Molecular Biology* 32 (3): 572–581. doi:10.1590/S1415-47572009005000061.

Tsuge, K, T Akiyama, and M Shoda. 2001. “Cloning, Sequencing, and Characterization of the Iturin A Operon.” *Journal of Bacteriology* 183 (21) (November): 6265–6273. doi:10.1128/JB.183.21.6265-6273.2001.

Turner, Ruth D. 1966. A Survey and Illustrated Catalogue of the Teredinidae. The Museum of Comparative Zoology Harvard University, Cambridge, Mass. 02138.

Udwary, Daniel W, Lisa Zeigler, Ratnakar N Asolkar, Vasanth Singan, Alla Lapidus, William Fenical, Paul R Jensen, and Bradley S Moore. 2007. “Genome Sequencing Reveals Complex Secondary Metabolome in the Marine Actinomycete *Salinispora tropica*.” *Proceedings of the National Academy of Sciences of the United States of America* 104 (25) (June 19): 10376–10381. doi:10.1073/pnas.0700962104.

Uppström, Leif R. 1974. “The Boron/chlorinity Ratio of Deep-sea Water from the Pacific Ocean.” *Deep Sea Research and Oceanographic Abstracts* 21 (2) (February): 161–162. doi:10.1016/0011-7471(74)90074-6.

Valladares, Ana, María Luz Montesinos, Antonia Herrero, and Enrique Flores. 2002. “An ABC-type, High-affinity Urea Permease Identified in Cyanobacteria.” *Molecular Microbiology* 43 (3): 703–715. doi:10.1046/j.1365-2958.2002.02778.x.

Velasco, A., P. Acebo, A. Gomez, C. Schleissner, P. Rodriguez, T. Aparicio, S. Conde, R. Muñoz, F. De La Calle, J. Garcia, J. Sánchez-Puelles. 2005. “Molecular Characterization of the Safracin Biosynthetic Pathway from *Pseudomonas fluorescens* A2-2: Designing New Cytotoxic Compounds.” *Molecular Microbiology* 56 (1) (March): 144–154.

Vincent, G., J. W. Lane, and R. M. Williams. 2007. “Regioselectivity of Pictet-Spengler Cyclization Reactions to Construct the Pentacyclic Frameworks of the Ecteinascidin-Saframycin Class of Tetrahydroisoquinoline Antitumor Antibiotics.” *Tetrahedron Letters* 48 (21) (April 21): 3719–3722.

Walsh, Christopher T., and Michael A. Fischbach. 2010. “Natural Products Version 2.0: Connecting Genes to Molecules.” *Journal of the American Chemical Society* 132 (8): 2469–2493. doi:10.1021/ja909118a.

Wang, Qiong, George M Garrity, James M Tiedje, and James R Cole. 2007. “Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy.” *Applied and Environmental Microbiology* 73 (16) (August): 5261–5267. doi:10.1128/AEM.00062-07.

Waterbury, John B., C. Bradford Calloway, and Ruth D. Turner. 1983. “A Cellulolytic Nitrogen-fixing Bacterium Cultured from the Gland of Deshayes in Shipworms (Bivalvia: Teredinidae).” *Science* 221 (4618): 1401–1403. doi:10.1126/science.221.4618.1401.

Wiegand, Irith, Kai Hilpert, and Robert E W Hancock. 2008. “Agar and Broth Dilution Methods to Determine the Minimal Inhibitory Concentration (MIC) of Antimicrobial Substances.” *Nature Protocols* 3 (2) (January): 163–175. doi:10.1038/nprot.2007.521.

Windust, A. J., J. L. C. Wright, and J. L. McLachlan. 1996. “The Effects of the Diarrhetic Shellfish Poisoning Toxins, Okadaic Acid and Dinophysistoxin-1, on the Growth of Microalgae.” *Marine Biology* 126 (1): 19–25. doi:10.1007/BF00571373.

Woodyer, Ryan D, Zengyi Shao, Paul M Thomas, Neil L Kelleher, Joshua A V Blodgett, William W Metcalf, Wilfred A van der Donk, and Huimin Zhao. 2006. “Heterologous Production of Fosfomycin and Identification of the Minimal Biosynthetic Gene Cluster.” *Chemistry & Biology* 13 (11) (November): 1171–1182. doi:10.1016/j.chembiol.2006.09.007.

Wright, Amy E., Dorilyn A. Forleo, Geewananda P. Gunawardana, Sarath P. Gunasekera, Frank E. Koehn, and Oliver J. McConnell. 1990. “Antitumor Tetrahydroisoquinoline Alkaloids from the Colonial Ascidian *Ecteinascidia turbinata*.” *The Journal of Organic Chemistry* 55 (15): 4508–4512. doi:10.1021/jo00302a006.

Yadav, Gitanjali, Rajesh S Gokhale, and Debasisa Mohanty. 2003. “SEARCHPKS: A Program for Detection and Analysis of Polyketide Synthase Domains.” *Nucleic Acids Research* 31 (13) (July 1): 3654–3658.

Yancey, Paul H., Mary E. Clark, Steven C. Hand, R. David Bowlus, and George N. Somero. 1982. “Living with Water Stress: Evolution of Osmolyte Systems.” *Science* 217 (4566). New Series: 1214–1222.

Yang, Joyce C, Ramana Madupu, A Scott Durkin, Nathan A Ekborg, Chandra S Pedamallu, Jessica B Hostetler, Diana Radune, Bradley S. Toms, Bernard Henrissat, Pedro M. Coutinho, Sandra Schwarz, Lauren Field, Amaro E. Trindade-Silva, Carlos A. G. Soares, Sherif Elshahawi, Amro Hanora, Eric W. Schmidt, Margo G. Haygood, Janos Posfai, Jack Benner, Catherine Madinger, John Nove, Brian Anton, Kshitiz Chaudhary, Jeremy Foster, Alex Holman, Sanjay Kumar, Philip A. Lessard, Yvette A. Luyten, Barton Slatko, Nicole Wood, Bo Wu, Max Teplitski, Joseph D. Mougous, Naomi Ward, Jonathan A. Eisen, Jonathan H. Badger, Daniel L. Distel. 2009. “The Complete Genome of *Teredinibacter turnerae* T7901: An Intracellular Endosymbiont of Marine Wood-boring Bivalves (shipworms).” *PloS One* 4 (7): 6085. doi:10.1371/journal.pone.0006085.

Young, C. M., and B. L. Bingham. 1987. “Chemical Defense and Aposematic Coloration in Larvae of the Ascidian *Ecteinascidia turbinata*. ” *Marine Biology* 96 (4): 539–544. doi:10.1007/BF00397972.

Zhao, Z., J. H. Thomas, N. Chen, J. A. Sheps, and D. L. Baillie. 2007. “Comparative Genomics and Adaptive Selection of the ATP-binding-cassette Gene Family in *Caenorhabditis* Species.” *Genetics* 175 (3) (March): 1407–1418.

Zhu, Yinghuai, Xiao Siwei, John A Maguire, and Narayan S Hosmane. 2010. “Application of Cycloaddition Reactions to the Syntheses of Novel Boron Compounds.” *Molecules (Basel, Switzerland)* 15 (12): 9437–9449. doi:10.3390/molecules15129437.

## Biographical Sketch

Sherif Elshahawi grew up in Cairo, Egypt. He received his Bachelor degree in 2000 in Pharmacy and Pharmaceutical Sciences from Cairo University. During his studies, he worked in a community pharmacy, in addition to a pharmaceutical company. Over the years, Sherif developed a keen interest in natural products that possess pharmacological activities. He realized the importance of being at the interface of chemistry and biology in this field. After graduation, he worked as a teaching and research assistant in the Faculty of Pharmacy for almost four years. He decided to pursue his Master's degree, so in 2004 he joined the graduate program of the Department of Pharmacognosy, The University of Mississippi, where he worked in Dr. Marc Slattery's lab. Sherif gained experience in the chemistry aspect of these compounds. After earning his degree in 2006, he decided to focus on the biological aspect of natural products. Sherif joined the biochemistry and molecular biology Ph.D program at Oregon Health & Science University in the same year under the guidance of Dr. Margo Haygood and he earned his degree in 2012.