2012

The Application of Functional Metagenomics to Natural Products Research

Jeffrey William Craig

THE APPLICATION OF FUNCTIONAL METAGENOMICS TO NATURAL PRODUCTS RESEARCH

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Jeffrey William Craig

June 2012

THE APPLICATION OF FUNCTIONAL METAGENOMICS TO NATURAL PRODUCTS RESEARCH

Jeffrey William Craig, Ph.D.

The Rockefeller University 2012

The study of bacterial natural products offers important insights into the physiologies and behaviors of individual species of bacteria as well as the contributions of these organisms to complex microbial communities. Such information frequently has important downstream applications in medicine, industry, agriculture, biotechnology, basic sciences and applied microbiology. The majority of bacteria found in nature cannot be cultivated in a laboratory setting, however, leaving a wealth of genetic and chemical diversity unexplored by traditional microbiological methods. Culture-independent approaches for studying diverse microbial populations affords indirect access to these resources. Within the wide array of culture-independent approaches currently in use (collectively termed "metagenomics"), the strategies of functional metagenomics provide access to important functions, phenotypes and activities encoded on fragments of environmental DNA (eDNA) as evidenced by the expression of these traits in model cultured bacterial hosts. Functional metagenomic techniques can be tailored towards finding metagenome-derived small molecule biosynthetic capabilities, thereby

allowing access to novel metabolites encoded within the genomes of uncultured bacteria. From a practical standpoint, however, eDNA-derived biosynthetic gene clusters may frequently go undetected in many functional metagenomic screens due to the inherent expressions limitations associated with individual heterologous hosts. One way to circumvent these limitations is to develop novel technologies that allow for cloned eDNA to be screened in a wide variety of bacterial hosts, each possessing its own unique expression capabilites. I begin this thesis by describing the development of a novel cloning system that permits the functional metagenomic screening of eDNA libraries hosted by multiple Gram-negative bacterial species, along with the application of this technology to natural product discovery through the targeting of clones that express small molecule associated phenotypes (Craig, Chang et al. 2009). The results of these screens support the hypothesis that phylogenetically diverse host species differ in their abilities to heterologously express foreign genes and gene clusters (Craig, Chang et al. 2010). Important genes and secondary metabolites associated with individual eDNA clones identified in these screens are also described in detail (Craig and Brady 2011). I conclude this thesis with a bioinformatic and functional analysis of eDNA clones that encode for the biosynthesis of *N*-acyl amino acids (Craig, Cherry et al. 2011). This analysis indicates that *N*-acyl amino acid synthase enzymes are found predominantly within bacterial species that harbor the putative PEP-CTERM/exosortase protein-sorting system.

This thesis is dedicated to

My Grandfather.

In loving memory of William Raymond Holub

“Learning to garden with my Pop-Pop” (1985)

ACKNOWLEDGMENTS

My family has told me on numerous occasions that they long suspected I would become a scientist. Their prediction was undoubtedly based on years of observing my love of nature and numbers - as well as my inquisitive, and often quirky, childhood behaviors. Ever since I can remember I've had a love for animals, and not just for my family's pet dogs, cats, fish and turtles. As a child I remember catching grasshoppers, lightning bugs, toads, newts, garter snakes and caterpillars and keeping them for hours in little make-shift aquariums for observation. Their movements and behaviors were fascinating to me, and I couldn't help but image what life in their world must be like. My fascination with nature wasn't limited to living creatures, either. For example, rock collections were amongst my first and most cherished possessions.

As I grew older, however, I began collecting new things. In second grade I opened my first pack of baseball cards. My fascination with baseball and with card collecting grew exponentially from that moment. My Grandfather was an avid baseball enthusiast who willing fed my appetite for baseball knowledge by sharing stories of the greatest players from his generation. From these history lessons it was obvious to me that one player stood head and shoulders above the rest. That person was former New York Yankee centerfielder Micky Mantle. Even though I never saw him play a

single game (he had been retired for nearly 25 years at the time), "the Mick" became my favorite ball player, and the Yankees became my favorite team. I doubt that either of those things will ever change. As an adolescent I was also fortunate enough to be best friends with another teenaged baseball junkie named Lucas Hammonds. Lucas and I were statistically inclined, by nature, and this trait only made us more beholden to our obsession with baseball. When we weren't too busy memorizing the backs of our baseball cards, we were often outside in my Grandfather's front lawn playing homerun derby. As adults we have come to look back on those days as being particularly integral to our intellectual developments. Lucas once likened it to the skill and coordination developed by two young wolf cubs at play. In addition to the thanks I owe to my parents for providing me with a healthy dose of natural "book smarts", I will always be grateful to my Grandfather, to Lucas and to Micky Mantle for fostering my love of math, numbers, statistics and strategy. I find it fitting that I should have the opportunity to defend this thesis on June 8th, 2011 - the 42nd anniversary of the first official Mickey Mantle Day.

Inspired by my high school chemistry teacher, Mr. Dennis Dilliplane, I entered college with the idea that I might one day become a chemist. In my head I was picturing myself as a synthetic chemist, but at the time I honestly had no idea what that meant. Once in college, my first true scientific research experience didn't come until the end of my freshman year at Johns

Hopkins University. After deciding to pursue Biophysics as a major, I searched endlessly for a lab that would be willing to take on a freshman student with no prior research experience. After I didn't hear back from the first few professors I emailed, I became somewhat depressed, figuring that I would never make the cut. That's when I decided to send out one last email, addressed to a Professor named Blake Hill in the Department of Biology. Within a day Blake had sent me an eager response. He was willing to offer me a position in his lab working under his first (and possibly all-time favorite) graduate student, Guru Thuduppathy. Guru, or G-man, was the best lab mentor that I could have ever wished for. He was exceedingly bright, amazingly patient and more than willing to answer any and all questions I threw his way. In the Hill lab I had found a home away from home - a safe place where I could be curious and productive at the same time. Blake, Guru, and fellow lab members Fred Tan and Salvadore Casares deserve an enormous amount of credit for steering me towards science as a career, as a hobby and as a way of contributing to society. Without their guidance I doubt that I would be where I am today.

I would also like to thank my parents for the sacrifices they've made to allow me to pursue an elite education at the college of my choosing. Knowing how incredibly important this gift can be, I plan to one day return the favor to my own children. I must also extend this gratitude to my Grandfather, William Raymond Holub. Were it not for his financial planning and

generosity, I might have easily passed up some of the most meaningful opportunities of my lifetime. I would also like to thank my Sister, Andrea, whose love and support have sustained me through trying times, and my best-man, Patrick Christopher Marti, who embodies everything that a true friend should be. And last, but certainly not least, I'd like to thank my wife, Bridget Mary Murphy, who has supported me firsthand throughout my Ph.D. training. Cohabitating with a quirky science nerd can't be a whole lot of fun, and my neuroticism, long work-hours and up-and-down personality probably didn't make things any easier. Still, Bridget stuck with me the whole way through, offering advice and perspective to help maintain my sanity.

In the four-plus years I've spent in The Laboratory of Genetically Encoded Small Molecules under the guidance of Dr. Sean F. Brady I have grown more intellectually and professionally than in any other time-period of similar duration. Because of Sean's influence, I've learned how to sustain my focus and how to think critically about my own research. These two "life lessons" will undoubtedly serve me well in the years to come. Since joining Rockefeller, I have also become a more confident scientific writer and editor, and I now appreciate how essential these skills really are. Working at a place like Rockefeller has in many ways been a dream come true. The seemingly limitless resources and the caliber of students and faculty have made each day both stimulating and challenging. My colleagues at Rockefeller and in the M.D.-Ph.D. program are amongst the most talented,

driven and innovative people that I have ever met. Surrounded by such talent, I am frequently overcome with an enthusiasm for scientific discovery and a will to use science for the greater good. I cannot express how grateful I am for having had the opportunity to learn from this community.

I would also like to thank: Dr. Olaf Anderson, Dr. Jochen Buck and Dr. Melissa Niremberg for offering scientific, professional and personal advice; my Rockefeller Faculty Advisory Committee members Dr. Vincent Fischetti and Dr. Alexander Tomasz for the efforts they've made to foster my scientific development; my external thesis advisor Dr. Sabine Ehrt for graciously serving on my committee and reviewing this thesis; past and current members of the Brady Lab for their incredible support over these past few years, particularly Jacob Banik, John Bauer, Fang Yuan-Chang, Jeffrey Kim, Ryan King, Hala Iqbal, Paula Calle and Jonette Suiter; and finally, I would like to thank my God-daughters, Kelly Theresa McCarthy and Mara Lynn Mennard, for reminding me that the most important things in life are the people we love.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACP	Acyl carrier protein
API	Atmospheric pressure ionization
ATCC	American Type Culture Collection
BAC	Bacterial artificial chromosome
BLAST	Basic local alignment search tool
Cam	Chloramphenicol
CIP	Calf intestinal alkaline phosphatase
CoA	Coenzyme A
cos	λ phage cohesive end segment
DNA	Deoxyribonucleic acid
DMSO	Dimethylsulfoxide
eDNA	Environmental DNA
ESI	Electrospray ionization
ExoAT	PEP-CTERM/exosortase system associated acyltransferase
Gent	Gentimicin
GFP	Green fluorescent protein
HMW	High molecular weight
HPLC-MS	High pressure liquid chromatography - mass spectrometry
HRMS	High resolution mass spectrometry
IncP	Incompatibility group P
Kan	Kanamycin
kb	Kilobase
LB	Luria Bertani medium
NAS	*N*-acyl amino acid synthase
NEB®	New England Biolabs® Inc.
NMR	Nuclear magnetic resonance spectroscopy
NRPS	Non ribosomal peptide synthetase
ORF	Open reading frame
Patl	*Pseudoalteromonas atlantica* T6c
PCR	Polymerase chain reaction
PKS	Polyketide synthase
PPT	4'-phosphopantetheinyltransferase
rcf	relative centrifugal force
rpm	Rotations per minute
rRNA	Ribosomal ribonucleic acid
RT	Room temperature
Spec	Spectinomycin
Tet	Tetracycline
Tn	Transposon
UV-Vis	Ultraviolet-visible spectrum absorption spectroscopy

CHAPTER 1

1 Introduction and Background

1.1 The Importance of Natural Products

For decades, natural products have been an important source of functional chemical diversity (Corey and Li 1999; Trost and Dong 2008; Nicolaou, Dalby et al. 2009). In fact, many of the most important therapeutically relevant compounds ever discovered are natural products, a significant fraction of which is derived from microorganisms (Newman, Cragg et al. 2003; Newman and Cragg 2004; Newman and Cragg 2007). Although microbial natural products display a wide range of bioactivities, they have offered their greatest utility as antibiotics, chemotherapeutics and immunosuppressive agents (Howitz, Bitterman et al. 2003; Miao, Coëffet-Legal et al. 2005; Harrison, Strong et al. 2009) (Figure 1). The success of natural products as lead compounds in drug development is attributed to refinements in the chemical and structural properties of naturally occurring metabolites that have resulted from constant selective pressures. The therapeutic activities of natural products can therefore be considered secondary to their intended roles under normal environmental circumstances, which are most likely related to nutrient scavenging, virulence, self-defense, communication and various other physiological processes (Omura, Ikeda et

al. 2001; Wolfgang, Kulasekara et al. 2003; Nougayrède, Homburg et al. 2006; Wyatt, Wang et al. 2010). Many of the chemical attributes seen in natural products are achieved through enzymatic transformations that exhibit site-specificity, chirality and reactivity that cannot be easily reproduced using traditional synthetic strategies, further emphasizing the potential value of natural products research (Savile, Janey et al. 2010).

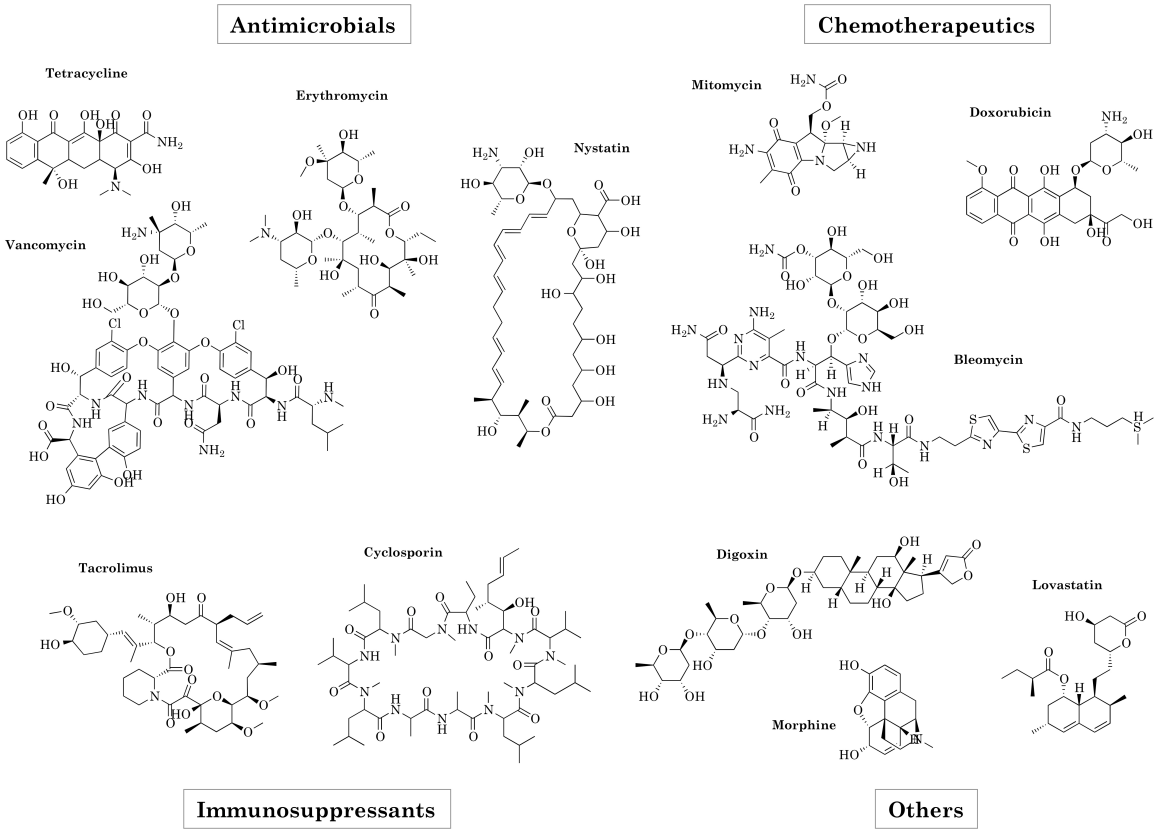


Figure 1: Medically important natural products.

The antimicrobials, chemotherapeutics and immunosuppressants shown above are all derived from soil bacteria. Digoxin is a plant natural product used to treat cardiac arrhythmias. Morphine is a plant natural product used for sedation and pain relief. Lovastatin is a fungal natural product used to control serum cholesterol levels.

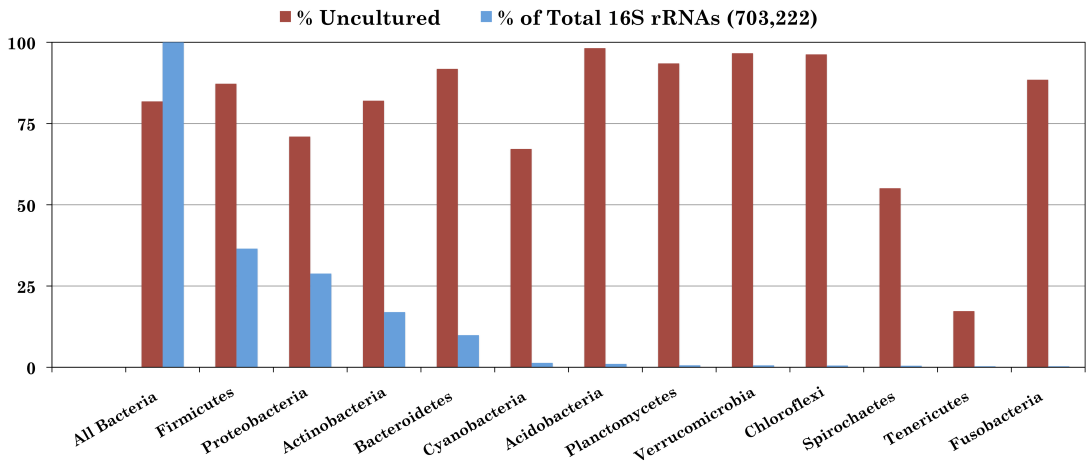
Amazingly, most of the structural diversity seen in natural products is generated through the use of relatively simple chemical building blocks. The high level of conservation amongst many of the enzymatic domains involved in natural product biosynthesis reflects this use of a common set of simple precursors. Such substrates are readily available, in most cases, as intermediates from the biosynthesis of proteins, fatty acids and other highly conserved pathways of central metabolism. The biosynthesis of many natural products can be conceptualized as occurring in two distinct phases, (i) the biosynthesis of core structural elements and (ii) the biosynthetic tailoring reactions that modify these core structures. Although several hundred thousand different natural products have been reported (Dictionary of Natural Product 19.2 Copyright © 2011 Taylor & Francis Group), those arising from four general classes of biosynthetic systems have been reported most frequently; These are the (1) polyketides, (2) non-ribosomal peptides, (3) isoprenoids and (4) shikimate derivatives (Lange, Rujan et al. 2000; Moore and Hertweck 2002; Schwarzer, Finking et al. 2003). The core structures produced by these four biosynthetic systems are frequently treated as "scaffolds" upon which "tailoring" modifications are performed. These tailoring reactions alter the final structures of natural products through various mechanisms, including oxidation/reduction, macrocyclization, halogenation, glycosylation, acylation, phosphorylation, sulfation, methylation and other chemical transformations (Savile, Janey et al. 2010).

1.2 The Chemistry of Uncultured Bacteria

The beginning of the 1940s marked the start of a productive era in the discovery of antibiotics from cultured soil bacteria (Schatz and Waksman 1944). The culture-based methods used to identify and study these first antimicrobials have not changed significantly since that time. Despite decades of historical productivity, pharmaceutical companies have recently deemphasized natural product discovery efforts due to the increasing rediscovery rates of common metabolites produced by easily cultured bacteria - a rate that now exceeds 99% (Zaehner and Fiedler 1999; Newman, Cragg et al. 2003; Newman and Cragg 2004; Newman and Cragg 2007). The difficulties associated with culturing environmental bacteria prevent traditional culture-dependent methods from being applied to a significant fraction of the bacterial species found in nature. The discrepancy between the number of bacteria present in environmental samples and the number that is easily cultured is a phenomenon known for over 50 years (referred to as the "great plate-count anomaly") (Jannasch and Jones 1959; Staley and Konopka 1985; Torsvik, Goksoyr et al. 1990). Recent molecular phylogenetic analyses have begun to quantitatively describe the extent of this anomaly by analyzing 16S rRNA genes amplified from environmental samples. Such surveys have indicated that a single gram of soil can contain up to 10,000 unique bacterial species and that less than 1% of these species have been grown in pure culture (Torsvik, Goksoyr et al. 1990; Torsvik, Salte et al.

1990; Tankere, Bourne et al. 2002; Torsvik and Ovreas 2002; Torsvik, Ovreas et al. 2002; Rappe and Giovannoni 2003; Webster, Yarram et al. 2007) (Figure 2). These same analyses also indicate that more than 80 major bacterial divisions exist, but that less than half are represented by cultured isolates (Schloss and Handelsman 2003; Keller and Zengler 2004; DeSantis, Hugenholtz et al. 2006; DeSantis, Hugenholtz et al. 2006; Cole, Chai et al. 2007; Webster, Yarram et al. 2007; Cole, Wang et al. 2009). Such insights suggest that uncultured bacteria are likely the largest remaining pool of genetic and chemical diversity on the planet. To circumvent the challenges of culturing environmental bacteria, novel culture-independent techniques have been developed for studying complex microbial communities. These techniques are collectively referred to as metagenomics (Handelsman, Rondon et al. 1998).

A)



B)

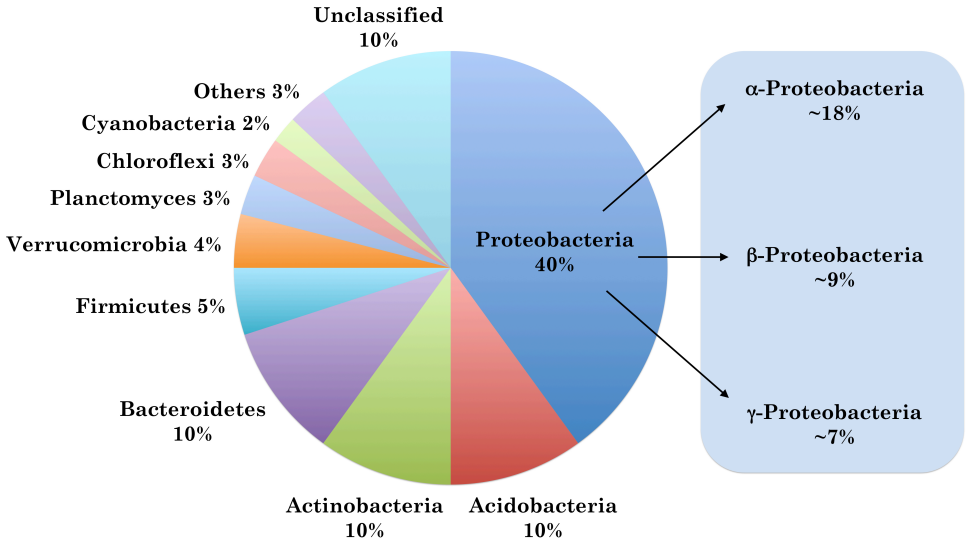


Figure 2: 16S rRNA sequence analysis of cultured and uncultured bacteria.

(A) High-quality 16S rRNA sequences from uncultured bacteria, collected by the Ribosomal Database Project (Release 10, update 26), are shown as a percentage of the total 16S rRNA sequences of major bacterial divisions. (B) The relative abundance of different phyla within 238 clone library data sets analyzed by Nemergut et el. (2011), corresponding to diverse terrestrial environments (Nemergut, Costello et al. 2011). Members of the superphylum *Proteobacteria* are prevalent throughout the biosphere, and the *α-Proteobacteria*, *β-Proteobacteria* and *γ-Proteobacteria* are the most abundant subphyla of *Proteobacteria* within bulk soil samples, according to Fierer et al. (2007) (Fierer, Bradford et al. 2007).

1.3 Metagenomics

Microbial 16S rRNA sequencing efforts indicate that only a small fraction of the world's environmental microbes have been cultured using standard microbiological methods (Jannasch and Jones 1959; Torsvik, Goksoyr et al. 1990; Kaeberlein, Lewis et al. 2002; Torsvik, Ovreas et al. 2002; Zengler, Toledo et al. 2002; Rappe and Giovannoni 2003; Gans, Wolinsky et al. 2005; Tringe, von Mering et al. 2005). From the bacterial kingdom, this minority is believed to represent less than 1% of all species in existence. The immense diversity of uncultured bacteria represents a promising source of novel natural products and biocatalysts that could potentially be exploited for use in agriculture, industry or as human therapeutics. Access to these resources is significantly limited, however, by the inability to culture these organisms in a laboratory setting. The field of metagenomics attempts to circumvent the challenges imposed by this culture-barrier by using culture-independent strategies based on DNA extracted directly from environmental samples (environmental DNA, eDNA) to explore naturally occurring bacterial populations (Handelsman, Rondon et al. 1998) (Figure 3). Frequently, collections of eDNA are archived through bulk cloning into various replicative plasmids maintained by model cultured bacterial hosts (Healy, Ray et al. 1995; Stein, Marsh et al. 1996). The resulting eDNA clone libraries are of potential interest to researchers looking to identify and recover metagenome-derived resources (novel genes, enzymes

or pathways). Metagenomic methods are universally applicable and have already been applied to microbial communities from soils/sediments (Rondon, August et al. 2000; Voget, Leggewie et al. 2003; Tringe, von Mering et al. 2005), rumen gut (Brulc, Antonopoulos et al. 2009), planktonic marine microbial assemblages (Beja, Suzuki et al. 2000; Breitbart, Salamon et al. 2002), deep sea microbiota (Sogin, Morrison et al. 2006), an acid mine site (Tyson, Chapman et al. 2004), arctic sediments (Jeon, Kim et al. 2009) and the Sargasso Sea (Venter, Remington et al. 2004).

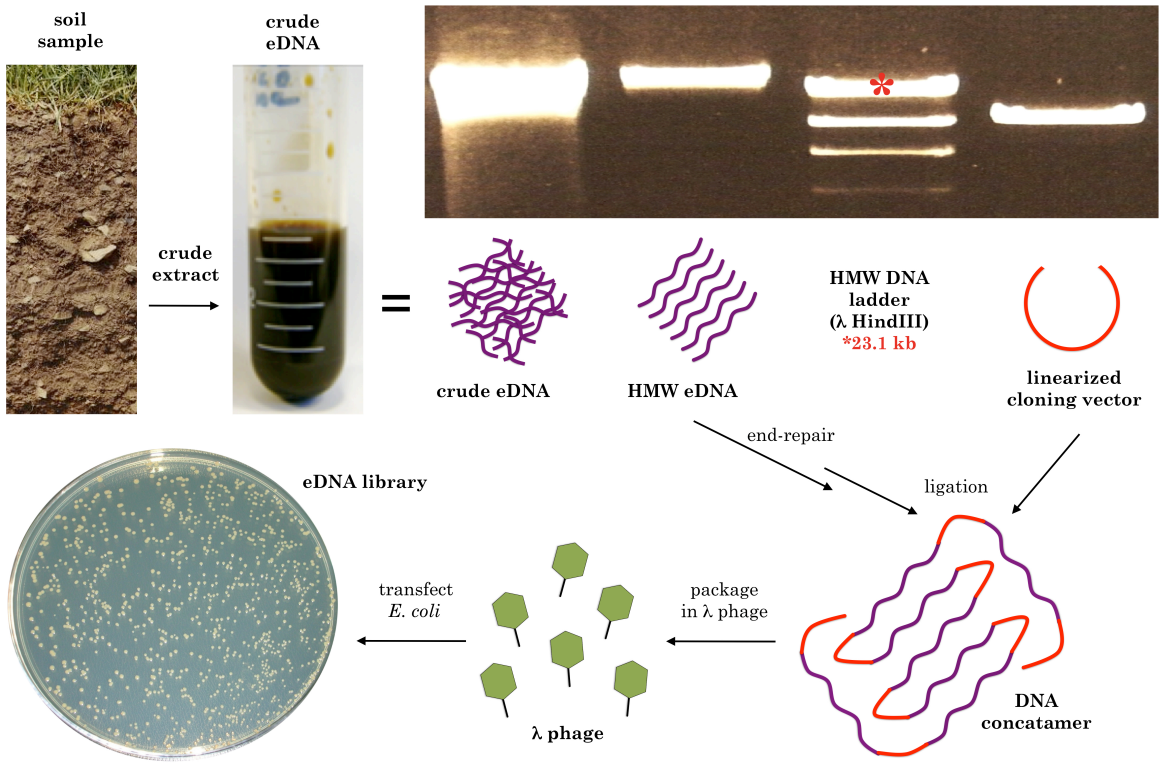


Figure 3: Metagenomic library construction.

For metagenomic library construction, environmental DNA from soil is crudely extracted, purified, size-selected, cloned, λ phage packaged and used to transfect *E. coli*.

Natural product researchers are most interested in the biosynthesis-related contents of eDNA libraries. Genes with related biological functions, including those required for the biosynthesis of secondary metabolites (i.e. natural product gene clusters, which often include genes for biosynthesis, regulation and resistance), are typically found in clusters on bacterial chromosomes. Libraries built to house large eDNA inserts may therefore contain clones harboring complete and functionally intact biosynthetic gene clusters. Once identified, this subset of clones may then be analyzed by genetic and chemical methods (e.g. heterologous expression experiments performed in model cultured bacterial hosts).

The identification of eDNA library clones harboring small molecule biosynthetic machineries is generally approached through two different, yet complementary strategies. These strategies have been generically labeled (1) phenotype or activity-based screening (a.k.a. functional metagenomics) and (2) DNA sequence or homology-based screening. In functional metagenomics, eDNA libraries are first examined using simple and robust criteria that target clones that express properties traditionally associated with small molecule production (i.e. changes in pigmentation, antibacterial activity and signaling phenomena). Clones identified during initial high-throughput screens are then isolated and examined in detail for the production of clone-specific small molecules. In DNA homology-based screening, degenerate primers based on conserved regions of natural product biosynthetic genes are

used to PCR amplify target sequences from library DNA. Recovered amplicons are used to design target-specific probes, which are then used to recover individual clones from pooled eDNA libraries. The biosynthetic enzymes encoded by these clones are often used for heterologous expression experiments.

Both functional metagenomics and DNA homology-based approaches offer certain advantages. Perhaps the most important advantage of functional metagenomics is that it allows for the identification of novel functional elements, the activities of which stem from genes whose functions could not be predicted by bioinformatic analyses and whose sequences have no homologs in the GenBank Database. In this sense, functional metagenomics is "sequence-blind" and independent of preconceived notions of what target gene(s) might look like. Unfortunately, functional metagenomic methods frequently suffer from problems related to their dependence on successful transcription and translation of foreign genetic material. Chapters 2 and 3 are focused specifically on functional metagenomic screening methods and the development of possible solutions to these problems.

1.3.1 eDNA Isolation

Large-insert clone libraries are ideal for many functional metagenomics applications, but require that eDNA extraction methods result in the isolation of sufficiently large DNA fragments. It is also imperative to isolate DNA as uniformly as possible from the diverse collections of microbes

present within environmental samples. If DNA isolation is in any way biased against or in favor of particular groups of microbes, all downstream analyses will reflect this bias as well (Liles, Manske et al. 2003; Feinstein, Sul et al. 2009).

There are two general approaches to the extraction of environmental DNA. The first approach is to simply lyse the microbes within bulk soil samples using various combinations of heat, detergents, enzymatic treatments, freeze/thaw cycles or other physical means (e.g. bead-beating) and to then collect the DNA released from these organisms (direct extraction). The second approach is to first recover the intact microbial cells from an environmental sample prior to the lysis step (indirect extraction). The direct extraction methods are advantageous in that they usually involve shorter processing times and provide larger yields of environmental DNA (Ogram, Sayler et al. 1987). Indirect extraction methods, however, are better suited to the isolation of DNA fragments >50 kb. One disadvantage of direct extaction methods is the inadvertent isolation of non-bacterial DNA and other contaminants (Ogram, Sayler et al. 1987; Tsai and Olson 1991; Tebbe and Vahjen 1993; Gabor, de Vries et al. 2003; Osborn and Smith 2005). However, this relatively minor complication is likely to be less problematic than the potential for indirect extraction methods to bias against microorganisms that are not readily dissociated from the environmental matrix.

Many different factors have been shown or hypothesized to affect the size and distribution of recovered eDNA, including the DNA extraction method itself, the collective status of microbial growth and the chemical composition of the sample (Bertrand, Poly et al. 2005). Protocols for the isolation of soil DNA must be gentle enough to prevent the excessive shearing of eDNA fragments, particularly if they are to be used for large-insert cosmid library construction. Many direct-extraction protocols, however, are sufficient for this purpose. Another nice feature of direct-extraction methods is that eDNA can frequently be sheared to fragments of the desired size through the random phycical forces of the extraction process. This effectively eliminates the bias introduced by restriction enzyme-based sizing strategies, however samples must usually undergo a second size-selection step prior to cloning (e.g. agarose gel electrophoresis).

Soils are heterogeneous mixtures that can vary substantially in their physical and chemical properties. The preliminary isolation of soil DNA (i.e. crude eDNA) is almost always accompanied by the co-isolation of organic compounds that inhibit PCR, cloning or other downstream applications. Although the exact composition of these organic inhibitors is sample-specific, they generally include humic acids, polyphenols, polysaccharides, and proteinaceous materials (Tebbe and Vahjen 1993; Zhou, Bruns et al. 1996; Frostegard, Courtois et al. 1999; Sylvia 2005). Secondary extraction and

purification methods are usually necessary to remove these contaminants prior to downstream events.

1.3.2 Standard Vectors for Library Construction

Selection of an appropriate cloning vector is important to ensure the overall maintenance and maximum expressivity of cloned eDNA pathways. There are multiple vector options for cloning high molecular weight (HMW) DNA from environmental samples, including cosmids, fosmids, and BACs. Cosmid vectors are hybrid plasmids that contain the cohesive-end (*cos*) sequences from the λ phage genome and were amongst the earliest vectors used for cloning (Collins and Hohn 1978). The packaging capacity of cosmids by λ phage proteins has a "sweet spot" between 45–48 kb, however the size of the eDNA fragments housed by cosmid vectors is intrinsically limited by the size of the vectors themselves (Bierman, Logan et al. 1992; Kim, Shizuya et al. 1992). The size selection performed by λ phage packaging ensures that vector self-ligation events, which result in empty cosmid clones, do not contribute significantly to the final library product. Fosmid vectors are similar to cosmids but are instead based on the *E. coli* F-factor replicon. The extremely low copy number (1-2 copies per cell) of fosmid vectors offers slightly greater stability for cloned eDNA fragments compared to cosmids, which often have low to medium copy numbers (5-20). Lower copy numbers are generally considered better for the long-term survival of foreign genetic material within a heterologous host. On the other hand, plasmid copy

numbers determine relative gene dosages. For functional metagenomics, higher gene dosages equate to higher "hit rates" due to their effect on expressivity and penetrance (i.e. the observability of eDNA-derived features). For DNA homology-based screening, it is often best to use vectors with lower gene dosages in order to limit the toxicity associated with certain foreign genes and proteins. Another class of vectors, the BACs, possess the ability to maintain inserts greater than 100 kb and, like fosmids, are based on the F-factor replicon (Shizuya, Birren et al. 1992). Both cosmid and fosmid vectors show reduced size compatibility ranges compared to BAC vectors, however cosmids and fosmids offer significantly greater cloning efficiencies.

1.4 Functional Metagenomics

Functional metagenomics is based on the screening of metagenomic libraries for the expression of specific observable phenotypes or activities conferred on heterologous hosts by cloned eDNA fragments (Henne, Daniel et al. 1999). The "hit rate" or frequency at which active clones are identified from functional metagenomic studies is often quite low and depends on a number of factors, including the vector-host system, the efficiency of heterologous expression, the size and abundance of the target gene(s), and the assay method. Consequently, high-throughput screening techniques are often implemented to increase the overall efficiency of the function-based screening process. Whenever possible, a direct selection strategy (i.e. selection versus screening) for obtaining positive clones can provide an

expedited path towards the isolation of eDNA clones displaying certain traits (e.g. acquired resistance to antibiotics, heavy metals or toxic xenobiotics) (Riesenfeld, Goodman et al. 2004; Mirete, de Figueras et al. 2007; Parsley, Consuegra et al. 2010) (Figure 4). Another variation of the selective approach is to use host strains that require heterologous complementation for growth under conditions that select for specific gene(s) of interest (Simon and Daniel 2009). Numerous studies have reported the use of such approaches for identifying biological targets, including the detection of enzymes involved in poly-3-hydroxybutyrate metabolism (Wang, Meek et al. 2006), DNA polymerase I (Simon, Herath et al. 2009), operons for biotin biosynthesis (Entcheva, Liebl et al. 2001), lysine racemases (Chen, Lin et al. 2009), glycerol dehydratases (Knietsch, Bowien et al. 2003) and naphthalene dioxygenase (Ono, Miyazaki et al. 2007).

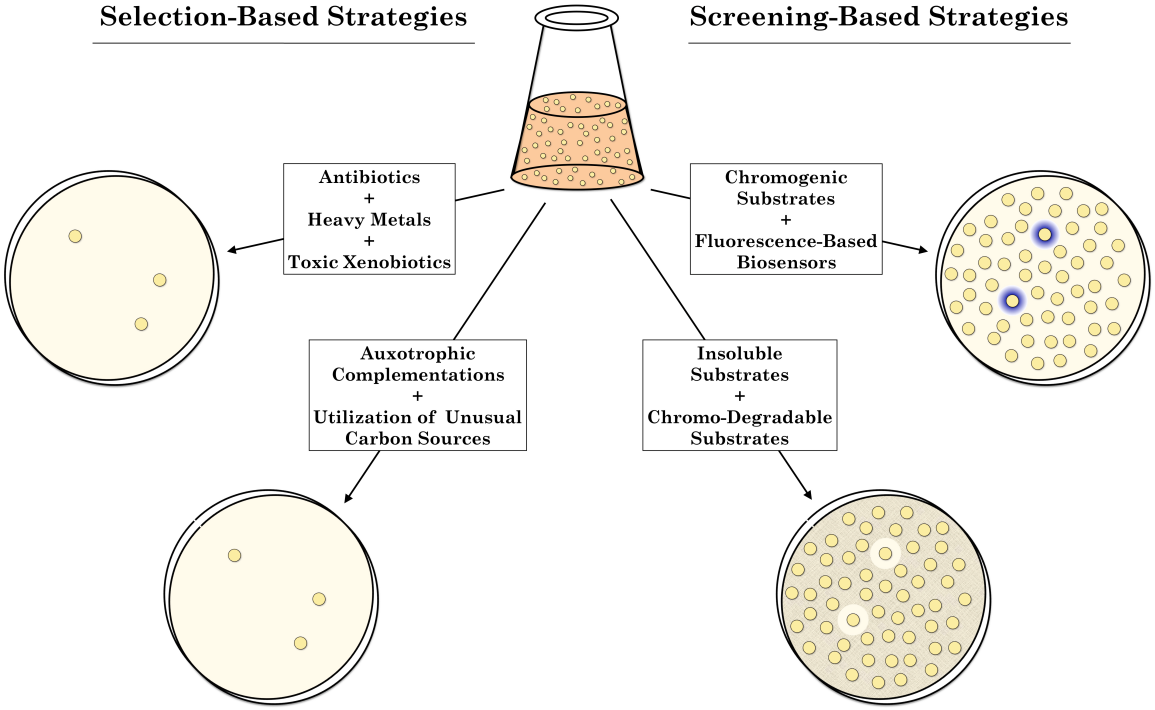


Figure 4: Common selection versus screening methods.

Selction-based strategies can be used to identify resistance elements, carbon-source utilization pathways or genes that complement auxotrophic host strains. Screening-based strategies can be used to identify various enzymes through the production of color or loss of color, through the degradation of insoluble substrates or through fluorescence-based biosensor systems.

When selection strategies are not available, screening methods are used instead. One of the first and most common strategies used to screen *E. coli*-based libraries involves the addition of various "indicators" to solid screening media. These indicators are often substrates for defined enzymatic classes that are able to convert the indicator to either clear or colored "halos" surrounding positive colonies (Figure 4). Incorporation of these substrates into the growth medium thereby permits the identification of eDNA clones harboring relevant enzymatic machineries. Examples of this type of strategy

include the identification of eDNA-derived esterases by the formation of clear halos around individual colonies (Elend, Schmeisser et al. 2006; Chu, He et al. 2008) and the identification of eDNA-derived extradiol dioxygenases through the production of a visible yellow pigment (Suenaga, Ohnuki et al. 2007). Functional screening can also be performed by visually inspecting arrayed metagenomic libraries for specific phenotypes. For example, metagenomic library clones displaying antimicrobial activity can be detected through growth inhibition assays utilizing chemically-sensitive assay strains encased in a layer of top-agar that has been overlaid on top of arrayed library clones (high-throughput) or the testing of supernatant extracts from individual eDNA clone cultures (low-throughput) (Rondon, August et al. 2000; Courtois, Cappellano et al. 2003; Brady 2007; Suenaga, Ohnuki et al. 2007).

1.4.1 Functional Metagenomics for Natural Product Discovery

The majority of microorganisms that remain uncultured collectively represent an important resource for the discovery of bioactive secondary metabolites. The biosynthetic pathways that encode for these metabolites can be captured in large-insert eDNA libraries and identified in functional metagenomic screens targeting small molecule induced phenotypes (Figure 5 and Figure 6). In the past, functional metagenomic screens of soil eDNA libraries for small molecule production have typically uncovered four common classes of biosynthetic genes and compounds. These include *N*-acyl amino

acid synthase (NAS)-expressing clones that produce long chain *N*-acyl amino acid antibiotics, aminolevulinic acid synthase (HemA)-expressing clones that produce dark-colored porphyrin-ring metabolites, 4-hydroxyphenylpyruvate dioxygenase-expressing clones that produce red pyomelanin pigments, and indole dioxygenase-expressing clones that produce mixtures of the blue dyes indigo and indirubin (Gillespie, Brady et al. 2002; Wilkinson, Jeanicke et al. 2002; Brady and Clardy 2005; Huang, Lai et al. 2009) (Figure 7). Other notable small molecule directed functional metagenomic studies include those describing the terragine family of metabolites (Wang, Graziani et al. 2000), isocyanide metabolites (Brady and Clardy 2005), the triaryl cation antibiotics turbomycin A and B (Gillespie, Brady et al. 2002) and an unidentified antifungal agent (Chung, Lim et al. 2008).

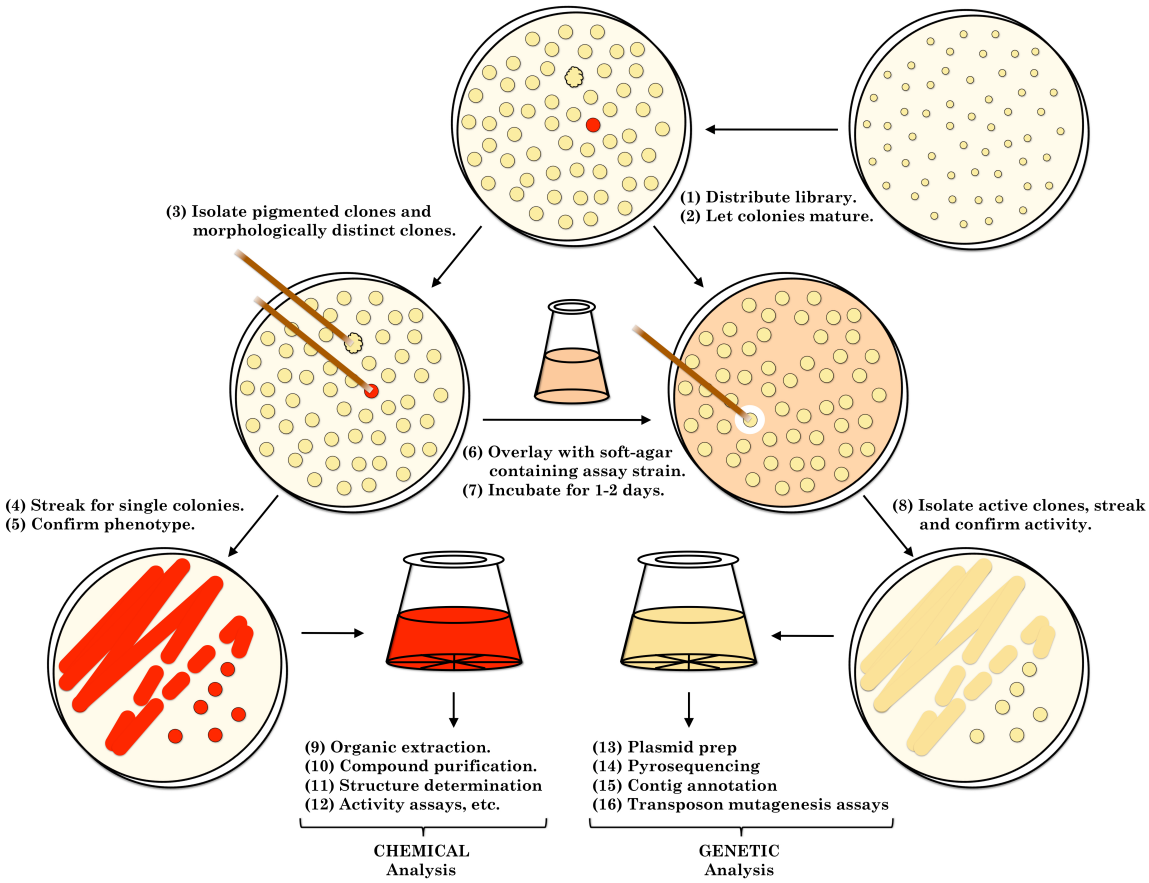


Figure 5: Screening for small molecule associated phenotypes.

eDNA libraries arrayed onto solid medium can be screened visually for phenotypic changes commonly associated with small molecule production, such as pigmentation, altered morphology or antibiosis.

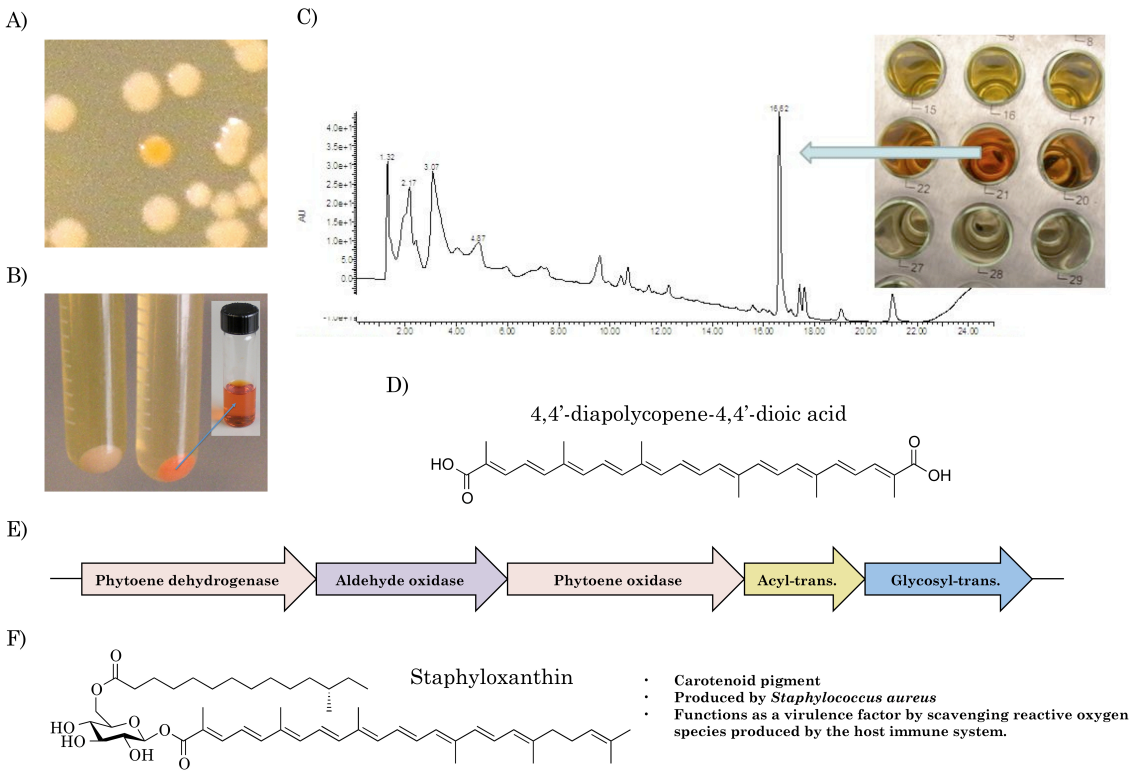


Figure 6: An example of small molecule functional metagenomics - clone RM101. (A) eDNA clone RM101, hosted by *Ralstonia metallidurans*, as it appeared on the original assay plate. (B) Liquid culture of RM101 indicates the presence of a cell-pellet associated pigment. (C) HPLC-MS purification of the major metabolite isolated from the acetone extract of RM101, which is most likely (D) 4,4'-diapolycopene-4,4'-dioic acid. (E) The RM101 eDNA insert contained a five-gene operon encoding the biosynthesis of a carotenoid pigment that may be similar in structure to the *S. aureus* metabolite (F) staphyloxanthin.

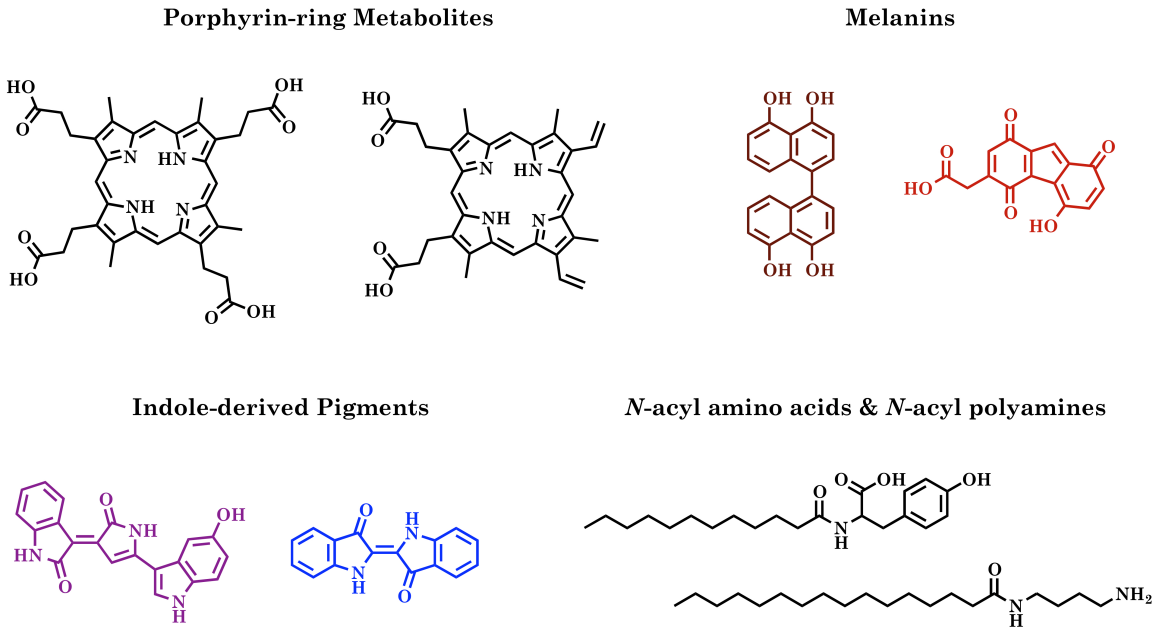


Figure 7: Four common classes of metabolites identified in *E. coli*-based screens. Porphyrin-ring metabolites appear as brown-pigmented clones. The examples shown are coproporphyrin III and protoporphyrin IX (referred to in Chapter 3). Melanins appear as diffusive red to brown pigments. The examples shown are DHN-melanin and pyomelanin (referred to in Chapter 3). Indole-derived pigments appear as blue or purple clones. The examples shown are violacein and indigo. *N*-acyl amino acids and *N*-acyl polyamines confer antibacterial activity to eDNA clones. The examples shown are *N*-dodecanoyl-tyrosine and *N*-hexadecanoyl-putrescine (referred to in Chapter 3 and Chapter 5).

1.4.2 Broad Host-Range Functional Metagenomics

E. coli strains that lack genes for homologous recombination (*recA*, *recBC*) and restriction (*mcrA*, *mcrBC*) are the most commonly used hosts for eDNA library construction. Commercial *E. coli* strains fitting this discription are available in both high-efficiency electrocompetent and chemically-competent preparations, and all varieties of cloning vector are available for *E. coli*-based libraries - plasmids (<10 kb DNA fragments), cosmids (25–35 kb),

fosmids (25–40 kb), and BACs (100–200 kb). Among these options, plasmids have the highest copy numbers and often the strongest vector-borne promoters, although these apparent merits improve hit rates only slightly (Simon, Herath et al. 2009). Because longer eDNA inserts will contain greater numbers of eDNA-derived genes, the relative hit rate of functional metagenomic screens increases proportionally with increasing eDNA insert size. Therefore, large-insert eDNA libraries are preferable, and can be built using either cosmid, fosmid or BAC vectors.

Bacterial soil metagenomes are complex mixtures of genomes from taxonomically, metabolically, and physiologically diverse species. Consequently, the genetic machineries of *E. coli* and other individual hosts are insufficient for the expression of the entire set of genes present within the global bacterial metagenome. All functional metagenomic efforts are therefore limited by host-dependent obstacles (deficiencies in the transcription, translation and processing of foreign genes and proteins), which are difficult to overcome (Figure 8). These obstacles are likely to culminate in intrinsic differences in the abilities of each host to express foreign genes and pathways (e.g., co-factors, post-translation modification enzymes, inducers, chaperones etc.). Due to these issues, functional metagenomics often fails to functionally access many encoded resources. The true extent of incidental transcription arising from natural regulatory sequences present within cloned eDNA is unknown, and the general feeling

within the functional metagenomics community is that the success or failure of most investigations is largely a matter of luck.

- **PROJECT-SPECIFIC VARIABLES**

  Size of the target gene(s)
  Abundance in the source metagenome

- **EXPERIMENTAL DESIGN VARIABLES**

  The assay method (sensitivity and selectivity)
  Vector choice (insert size, copy number, cloning efficiency)

- **HOST-SPECIFIC VARIABLES**

  Differential Transcription
  - G+C content
  - Codon bias
  - Promoter recognition
    - Sigma factors
    - Transcriptional regulators

  Differential Translation
  - Ribosomal-binding-sites
  - Translational start codon bias

  Post-translational processing
  - Protein folding
    - Inadequate chaperone function
  - Protein targeting and localization
    - Signal peptide recognition and transport
    - Postranslational modifications
      - Lipidation and Prenylation
      - Glycosylation
      - Proteolytic processing
      - Transpeptidation
  - Protein stability
    - Protease susceptibility
    - N-terminal processing

  Biochemical and enzymatic variables
  - Substrate availabilities (e.g. methylmalonyl-CoA)
  - Accessory enzymes (e.g. 4'-phosphopantetheinyltransferase)

  Other potential issues
  - Expression of toxic genes, proteins and metabolites
  - Unforeseen nuclease activity or recombination

Figure 8: Variables that affect the outcome of functional metagenomic screens. Host-specific variables, project-specific variables and experimental design variables can influence the outcomes of small molecule focused functional metagenomic studies.

This is not to say that attempts have not been made to quantify the relative "accessibility" of metagenomic samples. For example, Gabor *et al.*

used an *in silico* approach to describe the probability of identifying foreign genes from random expression cloning in *E. coli* (Gabor, Alkema et al. 2004). In this study, 32 genomes from diverse prokaryotes were analyzed for the presence of well-characterized expression signals (promoters, terminators and ribosomal binding sites) known to function in *E. coli*. This study found significant differences in the predicted levels of expression between various taxonomic groups and suggested that a maximum of only 40% of metagenome-derived genes could be expressed and detected using *E. coli*-based systems. Other studies have supported the widely held notion that eDNA expression is likely to be heavily biased in *E. coli* by pointing out intrinsic *E. coli* preferences, such as those for translation initiation codon usage (AUG), σ^{70} (RpoD)-binding site usage, and codon usage (Villegas and Kropinski 2008; Warren, Freeman et al. 2008; Kudla, Murray et al. 2009). Several functional metagenomic efforts have also indicated that clones positive for a specific activity detected using one host may not be detected in a different host and vice-versa (Li, Wexler et al. 2005; Wang, Meek et al. 2006; Craig and Brady 2011). Finally, a study by Martinez *et al.* showed that *E. coli, Pseudomonas putida* and *Streptomyces lividans* differed in their abilities to express heterologous gene clusters (Martinez, Kolvek et al. 2004).

Through the simultaneous use of multiple heterologous expression hosts, the inherently low probability of accessing eDNA-derived resources using functional metagenomics may be improved substantially. One of the

most efficient ways for achieving the potential to use multiple different host species is the creation of new vectors for eDNA library construction that allow for the efficient transfer of entire large-insert eDNA libraries from the library construction host (e.g. *E. coli*) to a variety of phylogenetically diverse species that could then serve as screening hosts. In this area of research, focus has centered on smaller derivatives of naturally occurring broad host-range plasmids that are able to replicate in multiple, phylogenetically distant species. Plasmids are typically characterized with respect to host range, replication mechanisms, incompatibility groups, and conjugal abilities. Plasmids with similar or related replication systems are considered to be incompatible (and therefore members of the same "incompatibility group") if they cannot stably coexist with one another in prokaryotic hosts (Datta 1979; Novick 1987). Cotransformation experiments have facilitated the sorting of plasmids into these incompatibility groups, of which there are now over 30 (Bergquist 1987; Couturier, Bex et al. 1988). Four incompatibility groups are considered to contain plasmids exhibiting broad host-range; They are the IncP, IncQ, IncW and IncN groups (Sobecky, Mincer et al. 1998). Among these, IncP1 (RK2-type) group plasmids have drawn considerable interest by researchers interested in expanding the repertoire of bacterial species available for function-based screening (Ono, Miyazaki et al. 2007; Aakvik, Degnes et al. 2009). Large (>20 kb) and poorly characterized cosmid-derivatives of the RK2-replicon have already been used for metagenomic

library constructions (Wexler, Bond et al. 2005), and vectors based on the RK2-replicon are maintained in a self-replicative state in numerous Gram-negative bacterial species (Thomas and Helinski 1989). In theory, libraries constructed in broad host-range vectors could be conjugatively transferred from *E. coli* to other hosts with high efficiency and minimal loss of the representative balance of clones present within the original libraries. The copy number of IncP1-α group vectors is also low enough to permit the maintenance of large eDNA inserts.

CHAPTER 2

2 Initial screens using *Ralstonia metallidurans*

2.1 Chapter Summary

Metagenomic studies designed to access new small molecules from the heterologous expression of environmental DNA have largely focused on the use of two model hosts, *Escherichia coli* and *Streptomyces* spp. Accessing the biosynthetic potential of DNA extracted from environmental bacteria will require the coordinated development of a diverse collection of model bacterial hosts that can be used for screening environmental DNA libraries. In this Chapter (Chapter 2) the *β-Proteobacteria Ralstonia metallidurans* was explored as a heterologous host using a novel broad host-range cloning system. The isolation and characterization of both novel and known metabolites from pigmented and antibacterially active clones identified in *R. metallidurans* based environmental DNA libraries are described. The clones found in this preliminary investigation did not confer the production of clone-specific metabolites to *E. coli*, validating *R. metallidurans* as an orthogonal expression host that can be used to expand the number of metabolites found in future metagenomic discovery efforts.

2.2 Introduction

It is now well established that the vast majority of bacteria present in environmental samples is not easily grown in the laboratory (Torsvik, Goksoyr et al. 1990; Rappe and Giovannoni 2003). Culture-dependent techniques for the discovery of natural products therefore fail to access the full chemical and metabolic diversity present within environmental samples. One means of accessing the biosynthetic potential of uncultured bacteria is to circumvent the culture-barrier by extracting DNA directly from environmental samples (environmental DNA, eDNA) and introducing this DNA into bacteria that are amenable to laboratory cultivation (Handelsman, Rondon et al. 1998). All potential cultured bacterial hosts are restricted, however, by their intrinsic abilities to recognize and utilize the foreign DNA captured within eDNA libraries. To date, only eDNA libraries based in *Escherichia coli* and *Streptomyces* spp., the traditional model systems used to study natural product biosynthesis, have yielded novel metabolites (Wang, Graziani et al. 2000; Brady, Chao et al. 2002; Brady and Clardy 2005). To expand upon these traditional hosts, the *β-Proteobacteria Ralstonia metallidurans* was explored as a potential new host for eDNA libraries built using a novel broad host-range cloning vector (Table 1). The isolation and characterization of both novel and known metabolites produced by clones identified in these initial eDNA library screens are also described. The development of a phylogenetically diverse collection of model bacterial

systems that can serve as heterologous hosts for eDNA libraries should expand the repertoire of natural products that can be accessed from uncultured bacteria.

Table 1: Strains, eDNA clones and plasmids discussed in Chapter 2.

Strain, eDNA clone, plasmid or library	Description	Source or reference
Strains		
E. coli EC100	*γ-Proteobacteria* host	Epicentre® Biotechnologies
E. coli EC300	Copy control *E. coli*	Epicentre® Biotechnologies
E. coli S17.1	Strain used for conjugal mating	(Simon, Priefer et al. 1983), ATCC 47055™ (S17-1)
B. subtilis 168	Source of *sacB* gene and promoter	(Kunst, Ogasawara et al. 1997), BGSCID 1A1, ATCC 23857™
B. subtilis 1E9	Antibiosis screening strain; [pBC16] Tc (Tet^r) trpC2	DSM402, BGSCID 1E9
S. aureus FDA 209	Sensitivity testing strain	FDA 209, ATCC 6538™
E. coli DRC39	Sensitivity testing strain; ΔAcrAB	(Margalit, Romberg et al. 2004)
E. coli BAS849	Sensitivity testing strain; *imp*-4213	(Sampson, Misra et al. 1989)
R. metallidurans CH34	*β-Proteobacteria* host	(Mergeay, Nies et al. 1985), ATCC 43123™
eDNA clones		
RM3	Carotenoid producing eDNA clone, hosted in *R. metallidurans* (RM3)	(Craig, Chang et al. 2009)
RM57	Antibacterially active eDNA clone, hosted in *R. metallidurans*, produces type III polyketides	(Craig, Chang et al. 2009)
Plasmids		
pGEX-3X	*tac* promoter-based expression vector; Amp^r	GE Healthcare 27-4803-01
pGEX-rmp57	pTAC-MAT-Tag-2 with *nasA*, S2K and I3L	(Craig, Chang et al. 2009)
pWEB	ColE1-replicon based cosmid cloning vector	Epicentre® Biotechnologies
pLAFR3	~23 kb RK2-derivative broad host-range cosmid	(Staskawicz, Dahlbeck et al. 1987)
pTR101	8.1 kb RK2-derivative broad host-range plasmid	(Weinstein, Roberts et al. 1992)
pJWC1	IncP1-α group broad host-range cosmid vector	(Craig, Chang et al. 2009)
pJWC1-rmp57	pJWC1 with the *rmp57* gene and its promoter	(Craig, Chang et al. 2009)
eDNA libraries		
JPA	450,000 clones; Wellsboro, Pennsylvania; creek bed mud/sediment eDNA	(Craig, Chang et al. 2009)
SROR	130,000 clones; Smith Rock, Oregon; sand/clay-covered desert soil eDNA	(Craig, Chang et al. 2009)

2.3 Results

A recent meta-analysis of soil metagenomic libraries found that the vast majority of bacteria present in environmental samples reside in five major divisions: *α-, β-, γ-Proteobacteria, Acidobacteria* and *Actinobacteria* (Fierer, Bradford et al. 2007). Although molecules have been found from

screening eDNA libraries hosted in representative *γ-Proteobacteria* (*Escherichia coli*) and *Actinobacteria* (*Streptomyces* spp.), the remaining three major divisions of bacteria have largely been overlooked (Wang, Graziani et al. 2000; Brady, Chao et al. 2002; Courtois, Cappellano et al. 2003; Brady and Clardy 2005). These three divisions were examined for potential hosts that would be genetically tractable, easily screenable and metabolically versatile. The *β-Proteobacteria*, *R. metallidurans* met all of these criteria. *Ralstonia* species are easily transformable, rapidly form small colonies that can be easily examined for interesting phenotypes and harbor a large number of genes for the metabolism of aromatic compounds (Taghavi, van der Lelie et al. 1994; Salanoubat, Genin et al. 2002; Mergeay, Monchy et al. 2003; Pohlmann, Fricke et al. 2006). *R. metallidurans* also contains a variety of cytochrome P450 monooxygenases that might be useful for transforming heterologously produced metabolites into novel entities (De Mot and Parret 2002; Urlacher, Lutz-Wahl et al. 2004).

E. coli remains the most efficient bacterium for creating large metagenomic libraries. Therefore each of the eDNA libraries was initially constructed in *E. coli* and then transferred to *R. metallidurans* by electroporation (Taghavi, van der Lelie et al. 1994; Brady 2007). The resulting *R. metallidurans* based libraries were screened for the production of color and antibiosis. Similar assays have served as successful high throughput primary screens for identifying clones that produce small

molecules in *E. coli* based eDNA libraries (Brady 2007). Two representative clones, one colored (RM3) and one antibacterially active (RM57), that produced organic extractable small molecules were selected for further study. Ethyl acetate extracts of the culture broth from the colored clone (RM3) contained a yellow pigment, while extracts from the antibacterially active clone (RM57) contained seven clone-specific metabolites. Neither the cosmid from RM3 nor the cosmid from RM57 conferred the production of any detectable clone-specific small molecules to *E. coli*, the host used most frequently for eDNA studies.

In an attempt to identify the genes responsible for encoding the biosynthesis of the metabolites produced by these two clones, both cosmids were sequenced (GenBank Accession Numbers: FJ151553 (RM3), FJ151552 (RM57)). RM3 was found to contain a six-gene operon resembling a hybrid of the carotenoid gene clusters from *Bdellovibrio bacteriovorus* and *Xanthobacter autotrophicus* (Figure 9A). 1D and 2D NMR and HRMS analysis of the yellow compound purified from cultures of RM3 showed that it was spectroscopically identical to the isoprenoid β-carotene (**1**). Sequence gazing of the cosmid from RM57 did not identify any obvious candidate genes as the source of the clone-specific metabolites produced by RM57. To identify the gene(s) responsible for the biosynthesis of these metabolites, the cosmid from RM57 was transposon mutagenized (<KAN-2> EZ-Tn5, Epicentre®) and retransformed into *R. metallidurans*. Extracts from 97 transposon mutants

were examined by TLC. Ninety-three mutant clones produced wild type levels of the seven clone-specific metabolites isolated from cultures of RM57; One showed reduced production of all seven metabolites and the remaining three produced no detectable clone-specific metabolites (Figure 10). The three knockout clones all contained transposon insertions in a putative type III polyketide synthase gene (RM57 protein, Rmp57). The transposon insertion in the mutant clone with reduced compound production was located just upstream of this ORF in a noncoding region that is likely the promoter for *rmp57*. *Rmp57*, along with its promoter, was subcloned from the RM57 cosmid as a 1680 bp PCR fragment and retransformed into *R. metallidurans*. Cultures of *R. metallidurans* transformed with *rmp57* alone produced the same mixture of metabolites seen with the entire RM57 cosmid indicating that Rmp57 is necessary and sufficient for the production of all seven clone-specific metabolites.

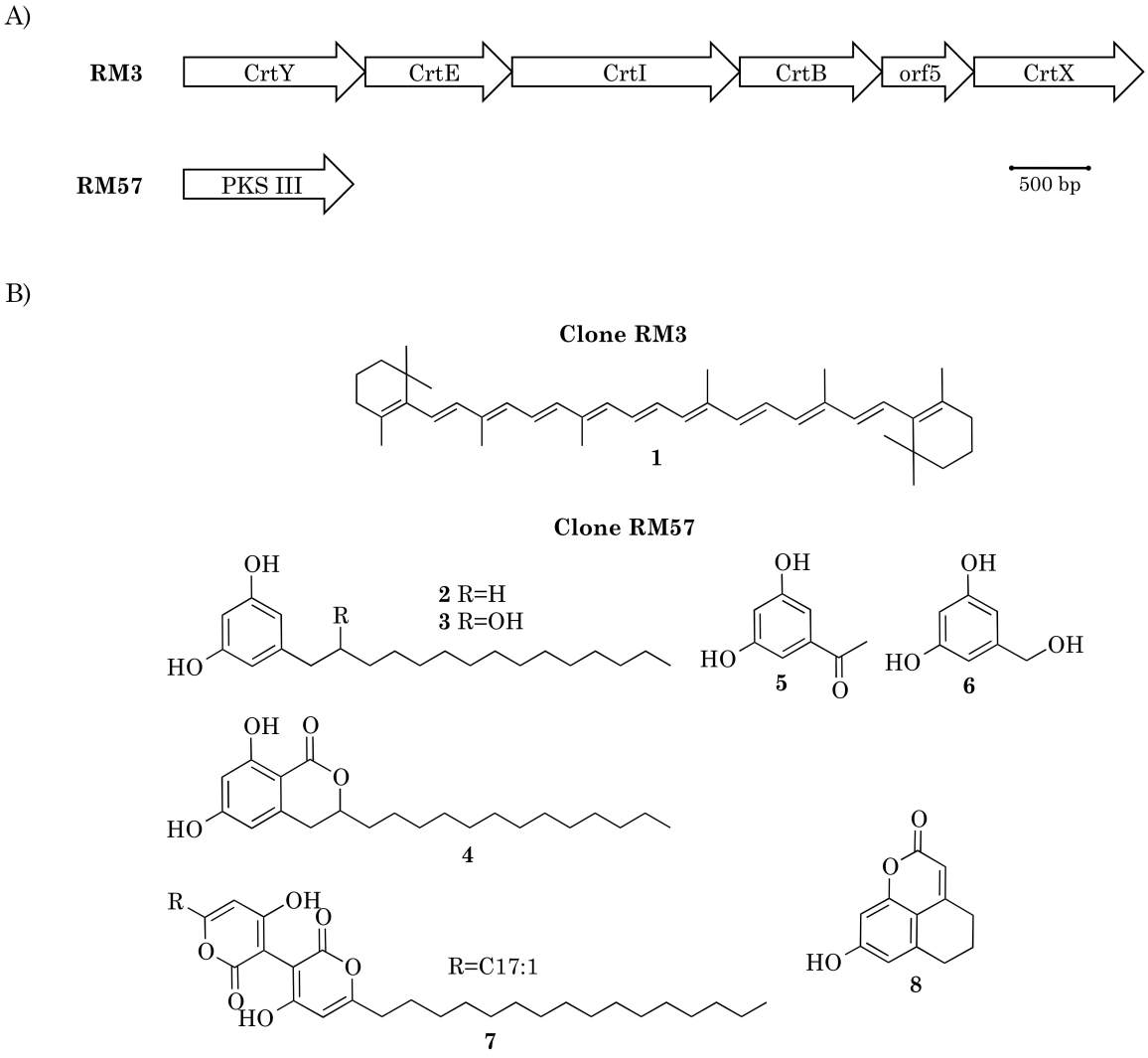


Figure 9: Genetic and chemical analysis of clones RM3 and RM57.

(A) The open reading frame(s) (ORF) responsible for the color and antibacterial phenotypes produced by RM3 and RM57, respectively. The ORF from RM57 is a predicted type III polyketide synthase (PKS) gene and the ORFs from RM3 are related to genes for carotenoid (Crt) biosynthetic enzymes. (B) Metabolites characterized from clones hosted in *R. metallidurans*. For compounds **2** and **3** only the major derivative present in the organic extract is shown. In addition to the C15 side chains shown, derivatives with C15:1 and C17:1 side chains were also detected in the organic extract.

Figure 10: Transposon mutagenesis analysis of cosmid RM57.

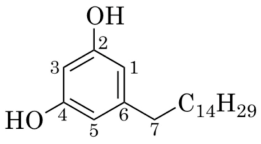
Transposon mutagenesis analysis of environmental clone RM57. Arrows depict ORFs, which are colored black for *rmp57* and grey for adjacent ORFs. The positions of transposon insertions within RM57 mutant clones deficient in metabolite production are shown as flags. Those insertions conferring a complete knockout of the production of all seven RM57-specific metabolites are colored black and the insertion conferring a decreased level of metabolite production is colored grey. The numbering scheme indicates basepairs from the beginning of the 454-assembled contig containing non-vector, environmental DNA.

Rmp57 is most closely related (~50-60% identity) to putative type III polyketide synthases found in the genomes of three different species of *Acidobacteria* (*Acidobacterium* sp. MP5ACTX9 - GenBank Accession No. ADW67770, "*Candidatus Koribacter versatilis* Ellin345" - GenBank Accession No. ABF42106, and *Acidobacterium capsulatum* ATCC 51196 - GenBank Accession No. ACO32650). With few exceptions, members of the phylum Acidobacteria remain recalcitrant to culturing (Barns, Takala et al. 1999; Joseph, Hugenholtz et al. 2003). The metabolites produced by RM57 were isolated from culture broth extracts using a combination of normal phase chromatography and reversed phase HPLC, and the structure of each metabolite was determined by 1D and 2D NMR and HRMS. The family of clone-specific metabolites produced by RM57 is shown in Figure 9B. The 1D and 2D NMR spectra for five of the seven metabolites (**2**-**6**) contained characteristic resorcinol-like aromatic signals that permitted structural

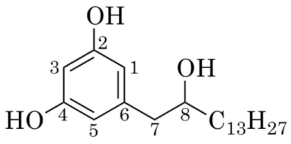
assignments as long- and short-chain substituted resorcinol derivatives (San Feliciano, Miguel del Corral et al. 1990; Dixit, Chen et al. 2003) (Table 2 and Table 3). The NMR spectra for compounds **7** and **8** did not show the same characteristic signals and could not be reconciled with any known metabolites.

Table 2: NMR data of compounds 2-6 (solvent).

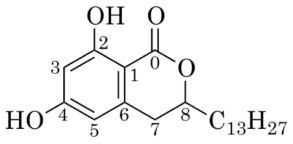
position	**2** (dichloromethane-d_2)[c]		**3** (methanol-d_4)[c]	**4** (methanol-d_4)[c]	**5** (methanol-d_4)	**6** (methanol-d_4)
	δ_C[a]	δ_H[b] (J in Hz)	δ_H[b] (J in Hz)	δ_H[b] (J in Hz)	δ_H[b] (J in Hz)	δ_H[b] (J in Hz)
1	108.4	6.26, d (2.2)	6.18, d (2.2)		6.90, d (2.2)	6.31, d (2.2)
2	157.3					
3	100.4	6.17, t (2.2)	6.13, t (2.2)	6.32, d (2.2)	6.50, t (2.2)	6.17, t (2.2)
4	157.3					
5	108.4	6.26, d (2.2)	6.18, d (2.2)	6.26, s	6.90, d (2.2)	6.31, d (2.2)
6	146.9					
7	36.3	2.51, t (7.9)	2.60, dd (13.2, 7.2) 2.55, dd (13.2, 6.6)	2.89, t (6.6)		4.46, s
8	32.5	1.60, p (8.1)	3.73, m	4.54, m	2.53, s	



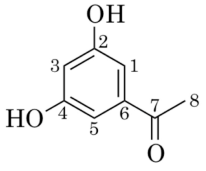
2



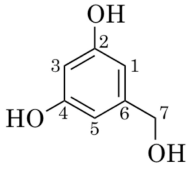
3



4



5



6

a. Measured at 100 MHz.
b. Measured at 600 MHz.
c. Chemical shift data is provided for head-group only.

Table 3: Chain length analysis of compounds 2-4.

Compound	$(MW\text{-}^1H)^-_{observed}$[a]	Length:Saturation of Alkyl Side Chain	Relative Intensity (% of total)[b]
2	317.4	15:1	15
	319.4	15:0	58
	345.4	17:1	27
3	335.4	13:0	74
	361.4	15:1	26
4	361.4	13:0	100

a. HPLC-MS analysis using ESI- mode on a Waters Micromass ZQ2000 ionization apparatus.
b. Relative intensities were calculated as a percentage of the total ion counts encompassing all species within an individual compound group.

The ^1H NMR spectrum of compound **7** contained a large methylene envelope similar to that seen in spectra from the long-chain resorcinol derivatives. However, unlike the other metabolites derived from RM57 which each contained at least six deshielded carbon chemical shifts, the ^{13}C NMR spectrum of compound **7** contained only five deshielded carbons (Table 4). The presence of only five deshielded carbons indicated that **7** could not contain the six-membered resorcinol head group seen in other RM57 metabolites. HMBC correlations from the lone olefinic proton singlet to the four remaining deshielded carbons along with HMBC correlations from the terminal methylene protons of the alkyl side chain to two of the deshielded carbons confirmed the presence of a tri-substituted six-membered pyrone head group in compound **7** (Figure 11). The molecular formula (HR-ESI-MS (*m*/*z*): $[M + H]^+$ calcd for $C_{44}H_{73}O_6$ 697.5407; found 697.5369) for compound **7** could only be reconciled with this NMR data if **7** were a dimer covalently linked through the C2 position of the pyrone head group. The most abundant pyrone dimer isolated from RM57 ethyl acetate extracts was a heterodimer that contained one saturated and one monounsaturated C17 side chain.

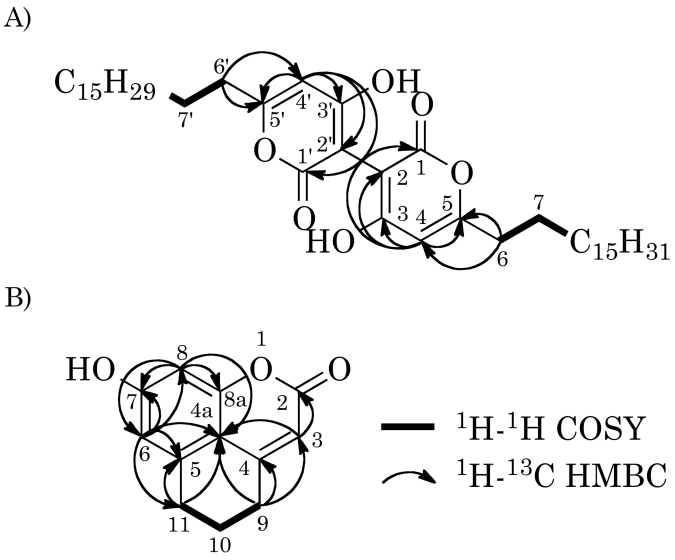
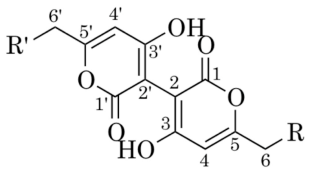


Figure 11: Structural characterization of Compounds 7 and 8.

^{1}H-^{1}H COSY and key ^{1}H-^{13}C HMBC correlations used to determine the structures of (A) compound 7 and (B) compound 8.

Table 4: NMR data of compound 7 in methanol-*d*4.

position	compound **7**			
	δ_C[a] mult.	δ_H[b] (J in Hz)	HMBC[c]	COSY
1	161.4			
2	87.7			
3	172.5			
4	100.3	5.94, s	1,2,3,5,6	
5	166.8			
6	31.3	2.49, t (7.6)	4,5,7	7
7	28.6	1.67, p (7.3)	6,8	6,8
8-21	21.9-28.9	1.29-1.37, m	6,7,22	7,22
22	13.0	0.92, t (7.0)	21	21



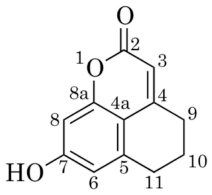
7

a. Measured at 150 MHz.
b. Measured at 600 MHz.
c. HMBC correlations are from proton(s) stated to the indicated carbon. All δ_C assignments were carried out on the basis of HSQC and HMBC NMR techniques.

The ^{1}H spectrum of compound **8** contained neither the large methylene envelope nor the characteristic pair of resorcinol derived doublets that were seen in other RM57 metabolites (Table 5). ^{13}C chemical shift data together with HMBC correlations associated with H6 and H8 (Figure 11) suggested, however, the presence of a tetra-substituted resorcinol-like system. This substituted resorcinol could be directly connected to a three-carbon methylene spin system (C9-C10-C11) by HMBC correlations from H11 to C4a, C5 and H6 to C11 (Figure 11). The other end of this spin system (C9) was linked by additional HMBC correlations to both the resorcinol-like ring and the C4-C3 olefin (Figure 11). The only way to place the final two atoms (CO) predicted by HRMS (HR-ESI-MS (*m*/*z*): $[M + H]^+$ calcd for $C_{12}H_{11}O_3$ 203.0708; found 203.0693) was to close the third ring and form the tricyclic isocoumarin (**8**).

Table 5: NMR data of compound 8 in dichloromethane-*d*2 and methanol-*d*4.

position	compound **8** (dichloromethane-d_2)		compound **8** (methanol-d_4)			
	δ_C[a] mult.	δ_H[b] (J in Hz)	δ_C[a]	δ_H[b] (J in Hz)	HMBC[c]	COSY
2	161.1		162.9			
3	107.5	5.97, s	105.9	5.98, s	2,4,4a	
4	153.2		156.0			
4a	110.3		109.4			
5	139.9		140.3			
6	111.4	6.61, s	111.3	6.62, t (1.1)	4a,5,7,8,11	
7	159.3		161.2			
8	100.5	6.64, d (2.2)	99.9	6.57, d (2.2)	4a,6,7,8a	
8a	155.1		155.3			
9	29.6	2.80, t (5.9)	29.4	2.84, td (6.1, 1.0)	3,4,4a,10,11	10
10	21.9	1.99, p (6.1)	22.0	1.98, p (6.2)	4,5,9,11	9,11
11	29.2	2.91, t (5.9)	28.7	2.92, t (6.1)	4a,5,9,10	10



8

a. Measured at 150 MHz.
b. Measured at 600 MHz.
c. HMBC correlations, are from proton(s) stated to the indicated carbon. All δ_C assignments were carried out on the basis of HSQC and HMBC NMR techniques.

Both compounds **7** and **8** are novel metabolites that have not been previously reported from studying cultured bacteria. The tricyclic isocoumarin based carbon skeleton seen in compound **8** has not been previously reported as a natural metabolite nor as a synthetic compound. Its closest naturally occurring relatives are plant metabolites containing additional methyl and hydroxyl substituents (Wang, Liu et al. 1988; Chen, Tang et al. 2006). While no naturally occurring long-chain pyrone dimers related to compound **7** have been reported, short-chain pyrone dimers have been produced in studies designed to synthesize unnatural variants of type III polyketide synthase derived natural products (Kwon, Lee et al. 2007).

Alkylresorcinol (**2**), the major known product isolated from the RM57 culture broth extract and the two novel metabolites (**7** and **8)** were assayed for antibacterial activity against *E. coli*, *Bacillus subtilis*, and *Staphylococcus aureus*. At the highest concentration tested (100 μg mL^{-1}) compound **2** showed no activity against any of these bacteria. The pyrone dimer (**7**) exhibited an MIC of 25 μg mL^{-1} against both *B. subtilis* and *S. aureus* and the tricyclic isocoumarin (**8**) exhibited an MIC of 100 μg mL^{-1} against *B. subtilis.* At 100 μg mL^{-1}, compound **8** showed no activity against *S. aureus* and neither compound **7** nor **8** exhibited activity against *E. coli*.

While a bacterial type III polyketide synthase that produces both resorcylic metabolites and six-membered pyrone ring containing metabolites was recently reported (Funabashi, Funa et al. 2008), the possibility that a single type III polyketide synthase could generate all of the molecules produced by RM57 seemed unlikely. To explore the origin of these metabolites, *rmp57* was cloned under an inducible promoter in *E. coli*. In this system, only pentadecylresorcinol (**2**) was found in the culture broth extract. To investigate the possibility that many of the compounds characterized from *R. metallidurans* RM57 cultures might arise from a mixed biosynthetic pathway involving both the eDNA derived type III polyketide synthase and endogenous *R. metallidurans* enzymes, compound **2** was fed (0.2 mg mL^{-1}) to cultures of *R. metallidurans* harboring the empty cosmid vector. After 5-6 days of incubation at 30°C both compounds **6** and **8**

accumulated in the culture broth of fed cultures. Neither compound appeared in control cultures that were not fed with compound **2**.

A proposed biosynthetic scheme that explains the biogenesis of the metabolites isolated from RM57 is shown in Figure 12 (Watanabe, Praseuth et al. 2007; Funabashi, Funa et al. 2008). The formation of pyrone and resorcinol head groups appears to largely depend on the length of the starter unit used by Rmp57. The incorporation of a C16 starter unit results in the enzyme-catalyzed condensation of three malonyl-CoA extender units, and upon cyclization and decarboxylation, a resorcinol based compound. The incorporation of a longer C18 starter unit largely results in the enzyme-catalyzed condensation of only two malonyl-CoA extender units, which upon cyclization yields a pyrone based compound. Endogenous *R. metallidurans* oxidases could then convert these long-chain resorcinol-based metabolites into compounds **6** and **8** and pyrone monomers into dimers. The biosynthesis of the various metabolites extracted from cultures of RM57 can thus be rationalized by the capacity of Rmp57 to produce both resorcinol and pyrone based compounds depending on the pool of available starter units and conversion of these compounds to novel metabolites by endogenous *R. metallidurans* oxidases.

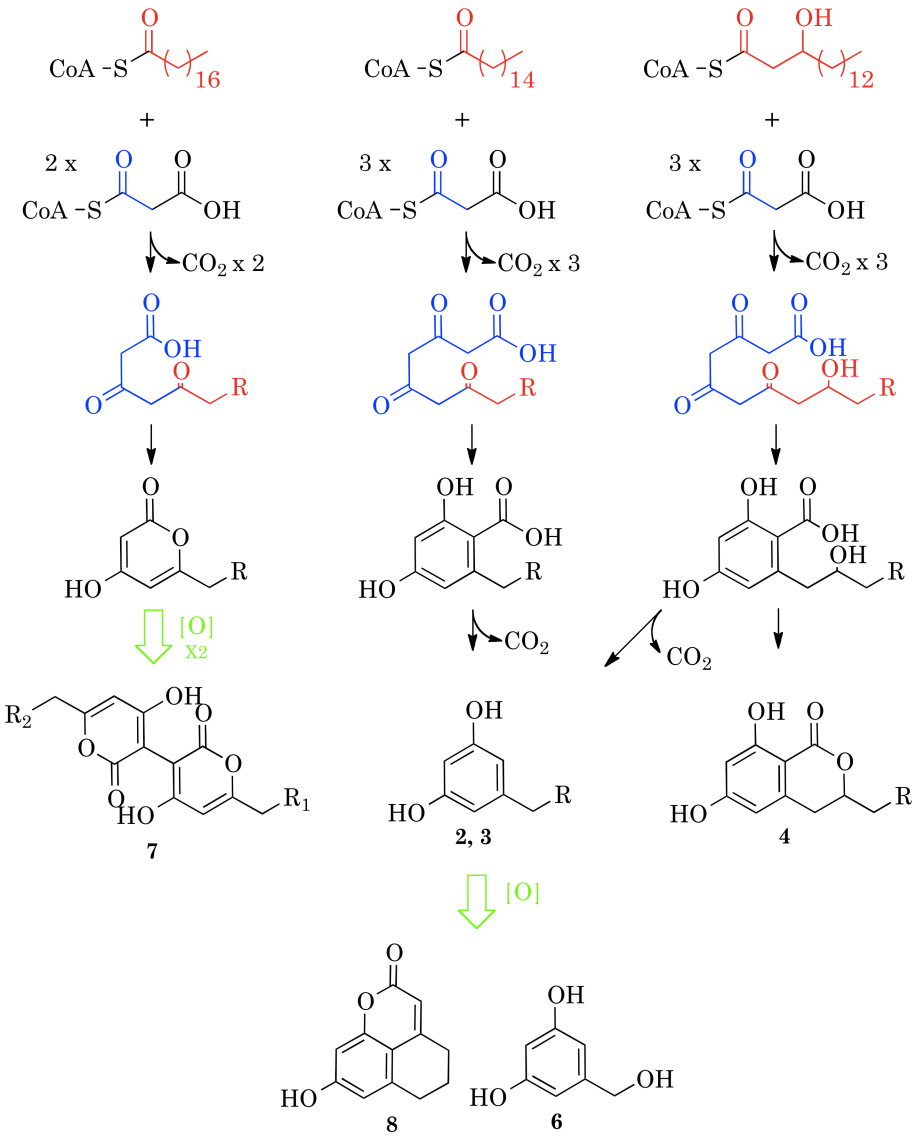


Figure 12: Biosynthesis of polyketide metabolites by Rmp57.

A biosynthetic scheme that includes both eDNA derived enzymatic steps (black arrows) and host enzymatic or spontaneous oxidation steps (green arrows) can be used to explain the biosynthesis of the clone-specific metabolites produced by RM57. Compound **5** was present in such low quantities in RM57 culture extracts (25 µg/L) that it was not possible to determine its biosynthetic origin.

2.4 Discussion and Future Directions

Compounds **7** and **8** are the first new metabolites to be identified from the heterologous expression of eDNA in a host other than *E. coli* or *Streptomyces* spp. The novel metabolites characterized in this Chapter (Section 2.3) appear to arise from the interaction of host enzymes with the metabolites produced from the expression of cloned eDNA. Both a host's native biosynthetic capacity as well as its propensity to recognize foreign promoters should be taken into consideration when choosing additional bacterial hosts for future eDNA studies. Accessing the biosynthetic potential encoded by DNA extracted from the mixtures of bacteria present in environmental samples will require the development of a diverse collection of model bacterial hosts that can be used to screen eDNA libraries. *R. metallidurans* represents a new model system that can now be used to expand the collection of small molecules found in future metagenomic studies.

2.5 Materials and Methods

2.5.1 Broad-host-range cosmid vector design

The 8.1 kb RK2-derivative plasmid, pTR101, was used as the foundation for the construction of a broad host-range cosmid vector (Weinstein, Roberts et al. 1992). A 1722 bp PCR fragment corresponding to a region from the *Bacillus subtilis* 168 genome containing the *sacB* gene and its

promoter was blunt-end cloned into the SmaI site of pWEB (Epicentre®; SmaI cuts at position 42; the *BSsacB* gene was cloned in the reverse orientation according to standard pWEB numbering). A 2741 bp PCR fragment of pWEB-bsSacB corresponding to 5'-CGA-1_{pWEB}-(BSsacB)-λcos-1115_{pWEB}-G-3' was then blunt-end cloned into the ScaI site of pTR101 (cloned in the reverse orientation according to standard pTR101 numbering; the ScaI site is in the β-lactamase gene, eliminating the ampicillin resistance marker of pTR101). pTR101-bsSacB-λcos was then digested with KpnI and ligated to a 3507 bp fragment of pLAFR3 containing the KorB/IncC/KorA stability region of RK2, obtained by PCR amplification (primer pair: KpnI-BamHI-Kor-F: 5'-GGTACCGGATCCTTTAGCTGAACAGTTCTCGACTTA-3', Kor-PmlI-BamHI-R: 5'-CACGTGGGATCCCACGTGTTTAGCGGCTAAAGGTGTTGACGT-3'). This final cosmid vector, pJWC1, was twice digested with ScaI, CIP-treated, and agarose gel purified in preparation for the blunt-end cloning of environmental DNA (Figure 13).

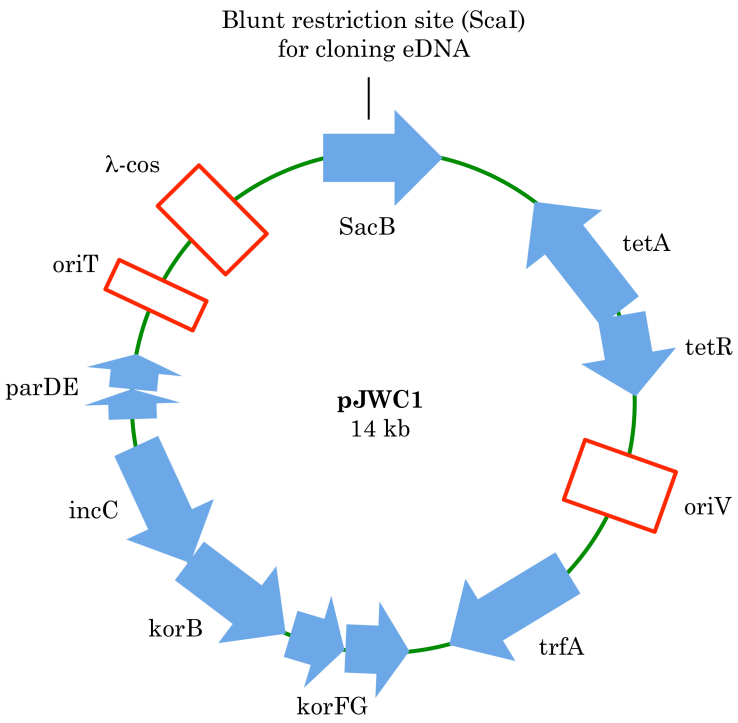


Figure 13: pJWC1 vector map.

Map of the broad host-range cosmid cloning vector pJWC1. *OriV* and *trfA* represent the minimal RK2 replicon required for replication in *E. coli (Lowbury, Lilly et al. 1969; Beringer 1974; Meyer, Figurski et al. 1975; Kolatka, Kubik et al. 2010).* *TetA* (exporter) and *tetR* (repressor), together, are the RK2-derived selectable marker for tetracycline resistance. *OriT* is the RK2-derived origin of conjugal transfer, when the *tra* and *mob* genes are provided in trans (Guiney and Yakobson 1983). *ParDE* is an RK2-derived stability region, termed the "plasmid addiction system", where ParE is a long-lived toxin and ParD is a short-lived antitoxin (Weinstein, Roberts et al. 1992). *IncC*, *korB* and *korFG*, together, represent an RK2-derived stability region implicated in the active partitioning of plasmids to bacterial progeny (Siddique and Figurski 2002). The λ-cos site confers packaging by λ-phage protein extracts, thereby allowing for cosmid library production (Collins and Hohn 1978). The *sacB* gene, from the genome of *B. subtilus* 168, encodes for levansucrase, which converts sucrose to glucose and levan (a homopolysaccharide composed of 2,6-β-D-fructosyl, which is toxic at high concentrations in the periplasm of Gram-negative bacteria) (Gay, Le Coq et al. 1985).

2.5.2 Environmental DNA isolation and library construction

Crude eDNA samples were prepared from soil using standard eDNA isolation methodology (Brady 2007). In brief: A one to one mixture of soil and lysis buffer (100 mM Tris-HCl, 100 mM Na EDTA, 1.5 M NaCl, 1% (w/v) CTAB, 2% (w/v) SDS, pH 8.0) was heated for 2 hours at 70°C, followed by removal of soil particulate matter by centrifugation (4,000 rcf, 30 minutes). Crude environmental DNA was precipitated from the supernatant with the addition of 0.7 volumes of isopropanol, collected by centrifugation (4,000 rcf, 30 minutes), washed with 70% ethanol and then resuspended in TE buffer (10 mM Tris-HCl, 1 mM Na EDTA, pH 8.0). The remaining soil particulate matter and humic substances were removed by large-scale gel purification on a 1% agarose gel (16 hours at 20 V). Purified HMW eDNA was then recovered by electroelution (2 hours at 100 V) and concentrated by isopropanol precipitation. Environmental DNA was blunt-end repaired (Epicentre®, End-It), ligated into a ScaI-digested and CIP-treated broad host-range cosmid vector, packaged into λ-phage (Epicentre®, MaxPlax Packaging Extracts) and transfected into *E. coli* EC100. After overnight selection on LB (20 μg mL^{-1} tetracycline, 50 g L^{-1} sucrose) agar plates the resulting cosmid library was resuspended from the selection plates and aliquots of the resuspended library were then used to inoculate liquid cultures. DNA miniprepped from these liquid cultures was used to electroporate *Ralstonia metallidurans*. 130,000 cosmid clones were constructed from Oregon soil

eDNA and 450,000 cosmid clones were constructed from Pennsylvania soil eDNA.

2.5.3 Transformation of *Ralstonia metallidurans* CH34 with eDNA cosmid libraries and phenotypic screening for antibacterial activity

Electrocompetent cells were prepared according to established protocols (Taghavi, van der Lelie et al. 1994). Electroporation reactions (1.0 mm cuvette, 1.8 kV pulse for 6 ms) for library production contained 2.5 mg of cosmid DNA and 80 μL aliquots of electrocompetent *Ralstonia metallidurans* CH34. After the electorporation pulse and the addition of 1 mL of LB supplemented with 10 mM $MgSO_4$, the cells were incubated with shaking at 37°C for 1.5-2.5 hours and then plated onto 150 mm diameter LB-tetracycline (20 μg mL^{-1}) selection plates at a titer of 1,500-2,500 colonies $plate^{-1}$. After 3-6 days at 30°C the plates were screened by visual inspection for colored colonies and a thin layer of LB-top-agar (15 mL $plate^{-1}$) containing *Bacillus subtilis* 1E9 (1:200 dilution of a culture at OD_{600nm} = 1.0) was poured onto each plate. After 1-2 days at 30°C the plates were screened for colonies that produced zones of growth inhibition. Active colonies were picked with a sterile toothpick and struck onto triple-selection media (tetracycline 20 μg mL^{-1}, kanamycin 3 μg mL^{-1}, chloramphenicol 1.5 μg mL^{-1}) to remove contaminating *B. subtilis*.

2.5.4 Organic extraction and compound isolation

For RM3 carotenoid extraction and isolation: The cell pellet from one liter liquid cultures of *R. metallidurans* CH34 (LB with tetracycline 20 μg mL^{-1} and kanamycin 3 μg mL^{-1}) grown at 30°C (200 rpm) for 5 days was extracted with acetone. Acetone extracts were then subjected to normal phase flash chromatography with 100% hexanes. Fractions containing pure yellow pigment by analytical TLC were pooled, yielding compound **1**. For RM57 metabolite extraction and isolation: One liter liquid cultures of *R. metallidurans* CH34 (LB with tetracycline 20 μg mL^{-1} and kanamycin 3 μg mL^{-1}) grown at 30°C (200 rpm) for 3-6 days were extracted with ethyl acetate and the dried extract was partioned by normal phase flash chromatography (step gradient: 100% hexanes, 95:5 hexanes:ethyl acetate, 90:10 hexanes:ethyl acetate (containing compound **4**), 75:25 hexanes:ethyl acetate (containing compound **2**), 50:50 hexanes:ethyl acetate (containing compounds **2** and **3**), 25:75 hexanes:ethyl acetate (containing compounds **3**, **5** and **8**), 100% ethyl acetate, 90:10 ethyl acetate:methanol (containing compounds **6** and **7**). Fractions containing compound **4** were subjected to a second round of normal phase flash chromatography (step gradient: 100% chloroform, 99:1 chloroform:methanol (containing compound **4**), 95:5 chloroform:methanol). Pooled fractions containing compound **4** were then subjected to preparative-TLC (95:5 acetone:dichloromethane) to give purified **4** (50 μg L^{-1}). Pooled fractions containing either compound **3**, **5** or **6** from the initial flash

chromatography step were separtely subjected to a second round of normal phase flash chromatography (step gradient: 100% chloroform, 98:2 chloroform:methanol, 95:5 chloroform:methanol (containing either compound **3** or compound **5**), 90:10 chloroform:methanol (containing compound **6**), 75:25 chloroform:methanol). Pooled fractions containing either compound **5** or **6** were subjected to preparative-TLC: compound **5**, 90:10 dichloromethane:methanol, yield 25 μg L^{-1} or compound **6**, 80:20 dichloromethane:methanol, yield 100 μg L^{-1}. Pooled fractions containing compound **3** were subjected to preparative-TLC (50:50 methanol:acetone) to yield purified **3** (150 μg L^{-1}). Pooled fractions containing either compound **2** or compound **8** from the initial round of flash chromatography were separately subjected to a second round of normal phase flash chromatography (four step gradient: 100% chloroform, 99:1 chloroform:methanol, 97.5:2.5 chloroform:methanol (containing compound **8**), 90:10 chloroform:methanol (containing compound **2**, 2 mg L^{-1}). Pooled fractions containing compound **8** were subjected to preparative-TLC (100% methanol) resulting in a final yield of 200 μg L^{-1}. Pooled fractions containing compound **7** from the initial round of flash chromatography were subjected to a second round of normal phase flash chromatography (four step gradient: 100% chloroform, 99:1 chloroform:methanol, 95:5 chloroform:methanol (containing compound **7**), 90:10 chloroform:methanol. Pooled fractions containing compound **7** were then subjected to preparative-HPLC (Waters XBridge C_{18} column [10 x 150

mm], 7 mL minute^{-1} flow rate); 90:10 H_2O:methanol w/ 0.1% formic acid to 100% methanol w/ 0.1% formic acid over 7 minutes, then holding at 100% methanol w/ 0.1% formic acid for 8 minutes, resulting in a final yield of 200 μg L^{-1}.

2.5.5 Compound 7, pyrone dimer

White powder; ^{1}H NMR (600 MHz, methanol-*d*4) 5.94 (C4/C4', 2H, s), 2.49 (C6/C6', 4H, t, 7.6), 1.67 (C7/C7', 4H, p, 7.3), 1.37-1.29 (C8-C21/C8'-C21', 56H, m), 0.92 (C22/C22', 6H, t, 7.0); ^{1}H-^{13}C HMBC derived ^{13}C δppm (150 MHz, methanol-*d*4) 161.4 (C1), 87.7 (C2), 172.5 (C3), 100.3 (C4), 166.4 (C5), 31.3 (C6), 28.9-21.9 (C7-C21), 13.0 (C22).

2.5.6 Compound 8, tricyclic isocoumarin

White powder; ^{1}H NMR (600 MHz, methanol-*d*4) 6.62 (C6, 1H, t, 1.1), 6.57 (C8, 1H, d, 2.2), 5.98 (C3, 1H, s), 2.92 (C11, 2H, t, 6.1), 2.84 (C9, 2H, td, 6.1, 1.0), 1.98 (C10, 2H, p, 6.2); ^{1}H-^{13}C HMBC derived ^{13}C δppm (150 MHz methanol-*d*4) 162.9 (C2), 105.9 (C3), 156.0 (C4), 109.4 (C4a), 140.3 (C5), 111.3 (C6), 161.2 (C7), 99.9 (C8), 155.3 (C8a), 29.4 (C9), 22.0 (C10), 28.7 (C11).

2.5.7 Cloning and expression of *rmp57* in *E. coli*

The 1077 bp *rmp57* gene was amplified (30 cycles of 97°C for 15 sec, 55°C for 20 sec and 72°C for 60 sec, NEB® Taq and Thermopol buffer) from environmental clone RM57 using the following primers: AE57PKS3EcoRIF: 5'-ATCCCCGGGAATTCAATGCGAATTGCATCTGTAGCAACAGCC-3',

AE57PKS3EcoRIR: 5'-GATAACGGGAATTCATCATGTACTTGAACCACGGATCAGTCC-3', which introduce EcoRI restriction sites upstream and downstream from *rmp57*. The PCR amplicon was digested with EcoRI and ligated into EcoRI-digested/CIP-treated pGEX-3X, then transformed into *E. coli* S17.1. Clones testing positive for the presence of the *rmp57* gene in the correct orientation were transferred to liquid cultures (LB supplemented with 50 μg mL^{-1} ampicillin) and incubated at 20°C, 30°C and 37°C with shaking, in both the presence and absence of 1.0 mM IPTG. After 3-5 days, cultures were extracted with an equal volume of ethyl acetate, dried under vacuum and analyzed by normal phase TLC and reversed phase HPLC.

2.5.8 Subcloning of *rmp57* in *R. metallidurans*

The *rmp57* gene and its upstream promoter region were amplified (30 cycles of 97°C for 20 sec, 50°C for 20 sec and 68°C for 120 sec, NEB® Taq and Thermopol buffer) from environmental clone RM57 using the following primers: 500UpsRmp57XbaIF: 5'-TACGTAAATCTAGAAAGACGCCGCGACGAGCCAATCAGCAGATG-3', Rmp57EcoRIR: 5'-GATAACGGGAATTCATCATGTACTTGAACCACGGATCAGTCC-3'. The PCR amplicon was blunt-end repaired (Epicentre®, End-It) and ligated into the ScaI-digested and CIP-treated broad host-range cosmid vector, then transformed into *E. coli* EC300. DNA miniprepped from successful ligation

transformants was then electroporated into *R. metallidurans*. Liquid cultures of *R. metallidurans* containing *rmp57* and its endogenous promoter were incubated at 30°C with shaking for 4 days, after which they were extracted with an equal volume of ethyl acetate, dried under vacuum and analyzed by normal phase TLC and reversed phase HPLC.

2.5.9 Antibacterial and antifungal activity assays

Activity assays were carried out against *Staphylococcus aureus* (ATCC 6538™; FDA 209), *B. subtilis* (BGSCID 1E9; DSM402 - [pBC16] Tc (Tet^r) trpC2), and *E. coli* DRC39 (ΔAcrAB) (Margalit, Romberg et al. 2004). The chemically sensitive *E. coli* strain BAS849 (*imp*-4213; LPS assembly OM complex LptDE - codominant LptD mutation) was not used in activity assays, but would have been an acceptable substitute for *E. coli* DRC39 (Sampson, Misra et al. 1989). Overnight cultures grown in LB were diluted 10^6 fold. 200 μL of cell suspension was aliquoted into each well of the first row of a 96-well microtiter plate. 100 μL of cell suspension was then aliquoted into all remaining wells. Compounds were resuspended in DMSO, added to the first well within a row (to a final concentration of 100 μg mL^{-1}) and then serially diluted in two-fold steps down each respective row. Plates were incubated overnight at 30°C with shaking, and the minimum inhibitory concentrations were recorded the following morning.

CHAPTER 3

3 Expanding small molecule functional metagenomics to diverse *Proteobacteria*

3.1 Chapter Summary

The small molecule biosynthetic diversity encoded within the genomes of uncultured bacteria is an attractive target for the discovery of natural products using functional metagenomics. Phenotypes commonly associated with the production of small molecules, such as antibiosis, altered pigmentation and altered colony morphology, are easily identified from screens of arrayed metagenomic library clones. Functional metagenomic screening methods are limited, however, by their intrinsic dependence on a heterologous expression host. Towards the goal of increasing the small molecule biosynthetic diversity found in functional metagenomic studies, broad host-range environmental DNA libraries were phenotyipically screened in six different *Proteobacteria*: *Agrobacterium tumefaciens*, *Burkholderia graminis*, *Caulobacter vibrioides*, *Escherichia coli*, *Pseudomonas putida* and *Ralstonia metallidurans*. Clone-specific small molecules found in the culture broth extracts of pigmented and antibacterially active clones are described, along with the genetic elements responsible for the biosynthesis of these metabolites. The host strains used in these screens provided access to unique

sets of clones showing minimal overlap, thus demonstrating the potential advantage conferred on functional metagenomics by the use of multiple, diverse host species.

3.2 Introduction

Uncultured bacteria are predicted to be a significant reservoir of novel small molecule biosynthetic machinery (Handelsman, Rondon et al. 1998; Rondon, August et al. 2000). One means to access the biosynthetic potential contained within the genomes of uncultured bacteria is through functional metagenomics (Handelsman, Rondon et al. 1998). This approach involves cloning DNA directly from naturally occurring microbial populations (environmental DNA, eDNA) and screening the resulting clone libraries for phenotypes traditionally associated with the production of secondary metabolites. A major limitation of functional metagenomics is its reliance on a foreign host to facilitate the expression of eDNA-derived genes and gene clusters (Gabor, Alkema et al. 2004). Codon bias, missing substrates and the inability to recognize foreign regulatory elements, including promoters and ribosomal binding sites, are just some of the obstacles that are likely to limit the success of expression-dependent studies with any single host organism. Circumventing these obstacles through an expansion of the collection of hosts available for functional metagenomic studies should increase the efficacy of this approach.

Soil ecosystems are rich in bacterial diversity, and the majority of soil-dwelling bacteria remain recalcitrant to standard microbial culture methods (Torsvik, Goksoyr et al. 1990; Rappe and Giovannoni 2003). Large-scale metagenomic sequencing studies indicate that soil microbiomes are often dominated by five bacterial phyla: the *α-*, *β-*, and *γ-Proteobacteria*, *Acidobacteria*, and *Actinobacteria* (Fierer, Bradford et al. 2007). Bacteria from these common phyla are therefore appealing hosts for use in functional metagenomic studies of soil-derived eDNA libraries.

In this Chapter (Chapter 3), six unique bacterial hosts (*Agrobacterium tumefaciens (α-Proteobacteria)*, *Burkholderia graminis (β-Proteobacteria)*, *Caulobacter vibrioides (α-Proteobacteria)*, *Escherichia coli (γ-Proteobacteria)*, *Pseudomonas putida (γ-Proteobacteria)* and *Ralstonia metallidurans (β-Proteobacteria)*) from the superphylum *Proteobacteria* were explored as hosts for functional metagenomic screening (Table 6). Soil eDNA libraries constructed in an IncP1-α group broad host-range cosmid vector were independently introduced into each of the six host *Proteobacteria*, and the resulting eDNA libraries were screened for colonies exhibiting antibacterial activity, altered pigmentation and altered colony morphology. Each of these phenotypes is easily detectable and frequently associated with small molecule production (Brady 2007). Individual clones possessing one or more of these natural product associated phenotypes were recovered from high-throughput primary screens and subsequently analyzed for the production of clone-

specific metabolites. Clone-specific small molecules responsible for the observed phenotypes and the genetic elements that encode their biosyntheses are described.

Table 6: Strains, eDNA clones and plasmids discussed in Chapter 3.

Strain, eDNA clone, plasmid or library	Description	Source or reference
Strains		
E. coli EC100	*γ-Proteobacteria* host	Epicentre® Biotechnologies
E. coli EC300	Copy control *E. coli*	Epicentre® Biotechnologies
E. coli S17.1	Strain used for conjugal mating	(Simon, Priefer et al. 1983), ATCC 47055™ (S17-1)
B. subtilis 1E9	Antibiosis screening strain; [pBC16] Tc (Tet^r) trpC2	DSM402, BGSCID 1E9
P. putida KT2440	*γ-Proteobacteria* host	(Bagdasarian, Lurz et al. 1981), ATCC 47054™
R. metallidurans CH34	*β-Proteobacteria* host	(Mergeay, Nies et al. 1985), ATCC 43123™
B. graminis C4D1M	*β-Proteobacteria* host	(Viallard, Poirier et al. 1998), ATCC 700544™
C. vibrioides CB15	*α-Proteobacteria* host	(Poindexter 1964), ATCC 19089™
A. tumefaciens LBA4404	*α-Proteobacteria* host	(Hoekema, Hirsch et al. 1983), Invitrogen™ 18313-015
eDNA clones		
EC5	Antibacterially active eDNA clone, hosted in *E. coli*, produces *N*-acyl amino acids	(Craig, Chang et al. 2010)
EC6	Antibacterially active (non-reproducible) eDNA clone, hosted in *E. coli*	(Craig, Chang et al. 2010)
RM19	Heme (protoporphyrin IX) producing eDNA clone, hosted in *R. metallidurans*	(Craig, Chang et al. 2010)
RM35	Antibacterially active eDNA clone, hosted in *R. metallidurans,* contains a cryptic NRPS gene cluster	(Craig, Chang et al. 2010)
RM44	Antibacterially active eDNA clone, hosted in *R. metallidurans*, produces *N*-acyl polyamines	(Craig, Chang et al. 2010)
RM57	Antibacterially active eDNA clone, hosted in *R. metallidurans*, produces type III polyketides	(Craig, Chang et al. 2009)
AT1/RM3	Carotenoid producing eDNA clone, hosted in either *A. tumefaciens* (AT1) or *R. metallidurans* (RM3)	(Craig, Chang et al. 2009)
AT3	eDNA clone with a wrinkled colony phenotype, hosted in *A. tumefaciens*	(Craig, Chang et al. 2010)
AT5	eDNA clone with a wrinkled colony phenotype, hosted in *A. tumefaciens*	(Craig, Chang et al. 2010)
Plasmids		
pTAC-MAT-Tag-2	*tac* promoter-based expression vector; Amp^r	Sigma Aldrich® E5405
pTAC-nasA	pTAC-MAT-Tag-2 with *nasA*, S2K and I3L	(Craig, Chang et al. 2010)
pJWC1	IncP1-α group broad host-range cosmid vector	(Craig, Chang et al. 2009)
eDNA libraries		
JPA	450,000 clones; Wellsboro, Pennsylvania; creek bed mud/sediment eDNA	(Craig, Chang et al. 2009)
SROR	130,000 clones; Smith Rock, Oregon; sand/clay-covered desert soil eDNA	(Craig, Chang et al. 2009)
RPA	170,000 clones; Reading, Pennsylvania; deciduous forest topsoil eDNA	(Craig, Chang et al. 2010)

3.3 Results

3.3.1 Broad host-range library construction and screening

The construction of an IncP1-α group broad host-range cosmid cloning vector, pJWC1, designed to facilitate functional metagenomic screening efforts was described in Chapter 2 (Section 2.5.1) (Craig, Chang et al. 2009). pJWC1 was used to create cosmid-based eDNA libraries from soil samples collected in Reading, Pennsylvania (RPA, deciduous forest topsoil, 170,000 clones), Wellsboro, Pennsylvania (JPA, creek bed mud/sediment, 450,000 clones), and Smith Rock, Oregon (SROR, sand/clay covered desert soil, 130,000 clones). These soils are representative of three different surface terrains, each of which was free from obvious pollution, contamination or other signs of human interference. Each library was initially constructed in *E. coli*, then transferred to *A. tumefaciens*, *B. graminis*, *C. vibrioides* and *R. metallidurans* by electroporation, and to *P. putida* by biparental conjugation. In each case, the retransformation of randomly chosen library clones back into *E. coli* confirmed that eDNA cosmids were stably maintained throughout the transfer and subsequent screening process despite the presence of functional recombination machineries in all hosts other than *E. coli*.

Environmental DNA libraries maintained in each host *Proteobacteria* were distributed onto 150 mm diameter plates at titers of 500-2500 colonies per plate. These arrayed eDNA libraries were allowed to mature for 4-6 days,

and then screened for colonies that exhibited antibacterial activity, altered pigmentation or altered colony morphology (Figure 14 and Table 7). Antibacterial screens were carried out using the two-layer overlay method with tetracycline resistant *B. subtilis* (BGSCID 1E9) as the assay strain. This approach provides a very simple method for identifying diffusible small molecule antibiotics. Alternative screening approaches would be required to identify compounds that remain sequestered within the host cell. Prior to the antibiosis overlay assay, each arrayed library was first examined for clones exhibiting altered pigmentation or altered colony morphology.

Table 7: eDNA clones identified in phenotypic screens of soil eDNA libraries.

Host Strain	Subphylum of *Proteobacteria*	Colonies displaying phenotype				
		Pigmentation			Antibiosis	Morphology
		Heme	Melanin	Other		
E. coli EC100	γ				2	
P. putida KT2440	γ	>10[a]				
R. metallidurans CH34	β	>10[a]	5	3	4	
B. graminis C4D1M	β	6			2	
A. tumefaciens LBA4404	α			1		2
C. vibrioides CB15	α					

a. Only a subset of brown colored clones was recovered from these libraries.

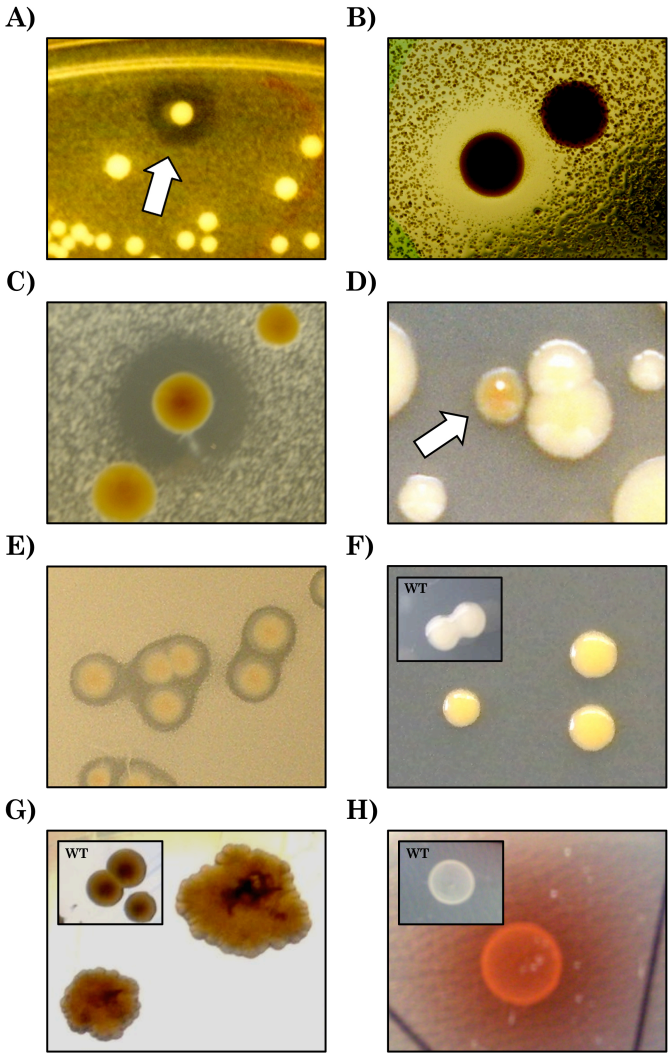


Figure 14: Small molecule associated phenotypes.

eDNA library clones displaying phenotypes of interest are shown. Representative colonies of interest as they appeared on the original assay plates: (A) antibacterially active colony found in an *E. coli* based library; (B) and (C) antibacterially active colony found in *R. metallidurans* based libraries; (D) pigmented colony found in a *P. putida* based library. Representative clones of interest after being recovered from primary assay plates: (E) antibacterially active colony recovered from an *E. coli* based library; (F) pigmented colony recovered from an *A. tumefaciens* based library; (G) wrinkled colony recovered from an *A. tumefaciens* based library; (H) pigmented colony recovered from a *R. metallidurans* based library.

3.3.2 *γ-Proteobacteria*

Standard cloning strains of the *γ-Proteobacteria E. coli* have been the hosts of choice for most functional metagenomic screening efforts. Therefore, *E. coli* was included in these screens to establish a base for comparing alternative bacterial hosts. No *E. coli*-hosted eDNA clones were found to display altered pigmentation or altered colony morphology during screens of any of the three libraries. Two antibacterially active eDNA clones, EC5 from JPA and EC6 from RPA, were, however, found during antibacterial-activity overlay assays. Upon retransformation, clone EC5 continued to confer antibacterial activity to its *E. coli* host, while EC6 did not reproduce the originally observed bioactivity (Figure 14E and 14A, respectively). Bioinformatic analysis of the fully sequenced eDNA insert from EC6 did not reveal any putative ORFs with homology to known biosynthetic enzymes, and therefore this clone was not investigated further (GenBank Accession GQ869384).

High-performance liquid chromatography mass spectrometry (HPLC-MS) analysis of ethyl acetate extracts derived from LB-broth grown cultures of the antibacterially active clone EC5 suggested the presence of a clone-specific collection of *N*-acyl aromatic amino acids. Clones that produce *N*-acyl amino acids are frequently encountered in antibacterial screens of *E. coli* hosted eDNA libraries (Brady, Chao et al. 2004). Individual *N*-acyl amino acids produced by EC5 were purified from ethyl acetate extracts by a

combination of normal phase flash chromatography and reversed phase HPLC. HPLC-MS analysis coupled with ^{1}H NMR studies of the two most abundant chemical species isolated from these ethyl acetate extracts indicated that the major clone-specific compounds produced by EC5 are spectroscopically and chromatographically identical to the 12-carbon *N*-acyl derivative of phenylalanine (observed m/z = 348.2, referred to as compound **11** beginning in Chapter 4) and the 14-carbon *N*-acyl derivative of tryptophan (observed m/z = 415.3, referred to as compound **24** beginning in Chapter 5) (Brady and Clardy 2005; Clardy and Brady 2007).

The fully sequenced EC5 eDNA insert (GenBank Accession GQ 869383) was found to contain a putative ORF showing low level (<20%) sequence identity to known *N*-acyl amino acid synthases (NAS). To test its role as a predicted NAS this ORF was PCR amplified and then cloned into the *E. coli* expression plasmid, pTAC-MAT-Tag-2 (Sigma-Aldrich® E5405). HPLC-MS analysis of the ethyl acetate extract derived from an *E. coli* culture transformed with this expression construct confirmed that it contained a mixture of long-chain *N*-acyl aromatic amino acids indistinguishable from that found in the extract derived from the original EC5 clone, confirming the NAS designation given to this gene.

The NAS (*nasA*) from EC5 is predicted to reside in a two gene operon (*nasAB*). The second ORF in this operon (*nasB*) contains a set of domains identical to those present in the E1α and E1β subunits of 2-oxo acid

dehydrogenase complexes (plus an E2-like lipoyl attachment site), as well as C-terminal β-keto-acyl synthase III (KAS III) domains, which are predicted to be involved in the synthesis of acyl carrier protein (ACP) linked fatty acids (Perham 1991; Finn, Tate et al. 2008). At the time of its discovery the only other putative enzyme found to contain the same predicted multi-domain architecture as nasB was encoded within the genome of *Agrobacterium radiobacter* K84 (GenBank Accession ACM27363) (Slater, Goldman et al. 2009). The putative *A. radiobacter* KAS III gene is also part of a predicted two gene operon sharing the same organization as the EC5 *nasAB* operon. In the operon from *A. radiobacter*, a putative fatty acid hydroxylase/sterol desaturase domain-containing enzyme (GenBank Accession ACM27362) sits in place of the NAS gene found in the eDNA derived operon.

The predicted domain functions of the eDNA-derived multi-domain KAS III (nasB) suggested that it might link the catabolism of 2-oxo acids to the production of ACP linked fatty acids, which are known substrates for NASs (Figure 15). The successful production of *N*-acyl amino acids in *E. coli*-based heterologous expression experiments with nasA alone indicates that, if nasB is a source of ACP linked fatty acids, it is not the exclusive source of ACP linked fatty acids used by nasA.

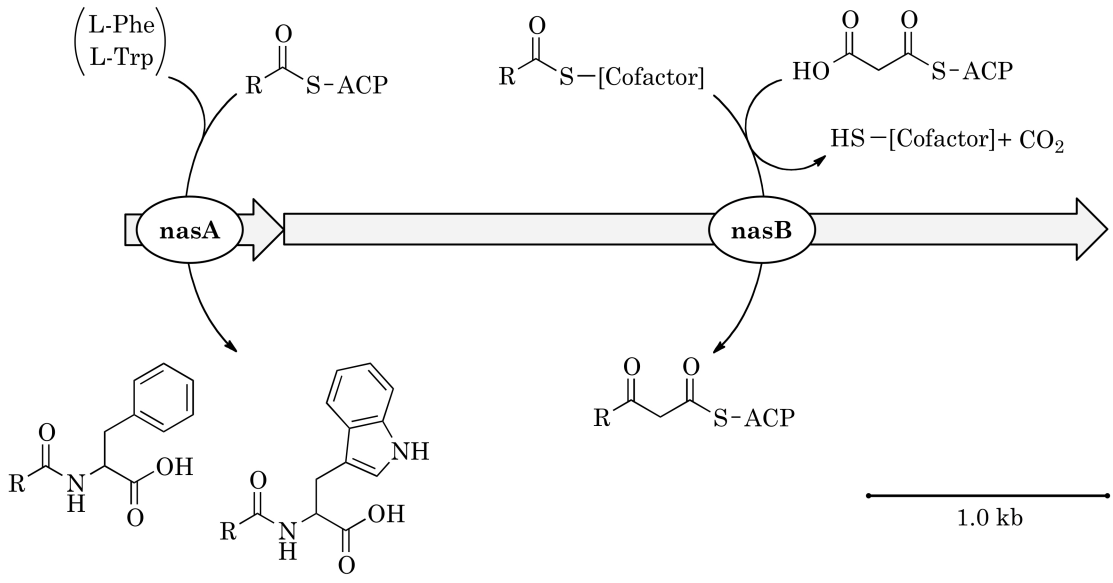


Figure 15: Structure and proposed funtion of the *nasAB* operon.

The EC5 putative *N*-acyl amino acid synthase operon (*nasAB*). The multi-domain ORF located immediately downstream of the EC5 NAS is predicted to produce ACP linked fatty acids that could serve as one source of acyl-ACPs used in the biosynthesis of *N*-acyl aromatic amino acids.

The soil-dwelling saprophyte *P. putida* was selected as a second representative host from the *γ-Proteobacteria* due to its frequent use as an expression host for natural product biosynthetic pathways (Zhang, Wang et al. 2008). The large colonies that are produced by *P. putida* are too easily dispersed by top-agar overlays to permit screening of *P. putida* based libraries for antibacterial activity using the two-layer overlay method. Surmounting this screening problem by generating mutant *P. putida* strains that are more compatible with high-throughput overlays could potentially transform *P. putida* into a productive host for antibiosis detection. While it is not possible to screen for antibacterially active *P. putida* based-clones, 11

pigmented clones, with colors ranging from deep-pink to brown, were identified in visual screens of the three eDNA libraries hosted in *P. putida* (Figure 14D). As seen with the brown-pigmented *R. metallidurans* clones described in Section 3.3.3 in more detail, HPLC-MS analysis of crude organic extracts derived from liquid cultures of these colored clones indicated that each one overproduced endogenous porphyrin-ring containing metabolites, and thus none were pursued further.

3.3.3 β-Proteobacteria

Chapter 2 (Section 2.3) described carotenoid and polyketide producing clones identified from phenotypic screens of the JPA and SROR libraries hosted by the *β-Proteobacteria R. metallidurans*, and showed that these same clones could not be identified using an *E. coli* host (GenBank accessions FJ151553 and FJ151552) (Craig, Chang et al. 2009). Screening of the RPA library in *R. metallidurans* led to the identification of four reddish-brown clones, six dark-brown clones and two clones displaying antibacterial activity (RM35 and RM44).

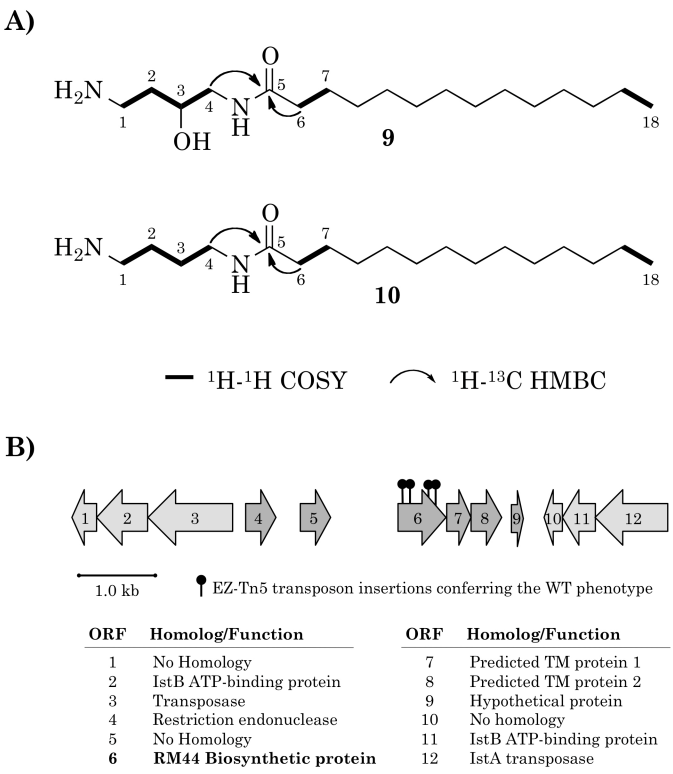
Reddish-brown melanin producing clones and dark brown heme producing clones have been previously reported from functional metagenomic screens of *E. coli*-hosted libraries (Brady and Clardy 2005; Lee, Yeo et al. 2008; Huang, Lai et al. 2009; Kim, Lim et al. 2009). The reddish-brown and dark brown clones identified in screens of *R. metallidurans* hosted libraries were therefore easily dereplicated as heme and melanin producers,

respectively. Mass and UV absorption spectra observed in the HPLC-MS analysis of acidified ethyl acetate extracts from the six dark-brown clones showed that each overproduced the heme-related pigment protoporphyrin IX (observed m/z = 563.3) (Herring 1972). One representative member of this family of clones, RM19, was transposon mutagenized and reintroduced into *R. metallidurans* for phenotypic screening. As reported from studies of brown clones found in *E. coli* hosted libraries, all of the transposon insertions that knocked out color production were found in a predicted glutamyl-tRNA reductase (*hemA* gene), which encodes for the enzyme that catalyzes the first committed step in heme biosynthesis. Acidified ethyl acetate extracts derived from cultures of the reddish-brown clones were all found to contain the same non-colored clone-specific metabolite (observed m/z = 169.1). This compound was purified from the ethyl acetate extract of a single member of this family of clones by normal phase flash chromatography. The ^{1}H NMR spectrum of the purified compound was found to be identical to that of homogentisic acid (commercial standard, Sigma-Aldrich® 53560). The accumulation of this common microbial metabolite, produced by the enzyme hydroxyphenylpyruvate dioxygenase (HPPD), is known to result in reddish-brown pyomelanin pigments (Figure 14H) (Plonka and Grabacka 2006).

Two clone-specific metabolites were isolated from ethyl acetate extracts derived from LB-broth cultures of the antibacterially active clone RM44 (Figure 14C). The chemical structure of each compound was

determined by 1D and 2D NMR experiments and high-resolution mass spectrometry (HRESIMS). HRESIMS indicated that the molecular formulas of compounds **9** and **10** were $C_{18}H_{38}N_2O_2$, and $C_{18}H_{38}N_2O$, respectively. Compound **9**, which was only predicted to differ from compound **10** by the presence of an additional oxygen atom, was present in the crude extract in 10-fold greater abundance than compound **10**.

^{1}H and ^{1}H-^{1}H COSY NMR spectra of compound **9** suggested the presence of a tri-substituted four-carbon spin system and a long-chain fatty acid substructure. Based on carbon chemical shift data, the tri-substituted four-carbon spin system was predicted to be functionalized at both C1 (H1 3.07 ppm) and C4 (H4A 3.29 ppm, H4B 3.18 ppm) with nitrogens and at C3 (H3 3.79 ppm) with an oxygen atom. Long range ^{1}H-^{13}C HMBC correlations from the C4 protons of the 4-amino-2-hydroxybutamine substructure to the carbonyl carbon of the fatty acid substructure established the structure of compound **9** as an *N*-acylated derivative of 4-amino-2-hydroxybutamine. Based on the molecular formula for **9**, the acyl side chain had to be a fully saturated 14-carbon fatty acid. The final structure of compound **9** was therefore *N*-(4-amino-2-hydroxybutyl)-tetradecanamide (Figure 16A, and Figure 32 in the Appendix).



ORF	Homolog/Function	ORF	Homolog/Function
1	No Homology	7	Predicted TM protein 1
2	IstB ATP-binding protein	8	Predicted TM protein 2
3	Transposase	9	Hypothetical protein
4	Restriction endonuclease	10	No homology
5	No Homology	11	IstB ATP-binding protein
6	**RM44 Biosynthetic protein**	12	IstA transposase

Figure 16: Genetic and chemical analysis of clone RM44.
(A) Key ^{1}H-^{1}H COSY and ^{1}H-^{13}C HMBC correlations used to establish the structures of compounds **9** and **10**. (B) Transposons that knockout the production of compounds **9** and **10** all fall in a single ORF (ORF6). This biosynthetic ORF is located within a genetic island flanked on both sides by IstA/IstB-like, IS21-family transposable sequences (ORFs 2-3, 11-12).

The 1D ^{1}H and ^{1}H-^{1}H COSY spectra of compound **10** indicated the presence of four methylene groups organized as a linear four-carbon spin system as well as a long chain fatty acid substructure. Chemical shift analysis suggested both ends of the four-carbon spin system were functionalized with nitrogen atoms (H1 2.94 ppm, H4 3.20 ppm). As seen in the ^{1}H-^{13}C HMBC spectrum of compound **9**, there was a correlation from protons on the terminal carbon (C4) of the four-carbon spin system to the

carbonyl carbon of the fatty acid substructure (Figure 16A, and Figure 33 in the Appendix). Based on the molecular formula for **10**, the acyl side chain was a fully saturated 14-carbon fatty acid. The final structure of compound **10** was therefore *N*-(4-aminobutyl)-tetradecanamide, or myristoylputrescine. Compound **9** was a new natural product that had not been previously isolated from cultured bacteria, and compound **10** was a trivially new metabolite, differing in acyl-chain length from the known bacterial natural product palmitoylputrescine (Brady and Clardy 2004).

Saturating transposon mutagenesis of the RM44 cosmid indicated that a single ORF was responsible for both the observed antibiosis and the production of compounds **9** and **10** (Figure 16B). This ORF was most similar to the biosynthetic enzyme responsible for the production of palmitoylputrescine (GenBank Accession AAV33349) and the N-terminus of a putative serine-pyruvate aminotransferase from the genome of *Bacillus* sp. NRRL B-14911 (GenBank Accession EAR68203). The entire RM44 eDNA insert was subsequently sequenced (GenBank Accession GQ869386), and the RM44 biosynthetic ORF was found to be located within an 8 kb region flanked on both sides by genes homologous to the tandem IstA/IstB transposase elements contained within IS21-family transposable sequences. In addition to the RM44 biosynthetic ORF, this region contained genes for a putative restriction endonuclease and two short hypothetical proteins, predicted to contain multiple transmembrane helices (Claros and von Heijne

1994). In combination, the latter two ORFs, which sit directly adjacent to the RM44 biosynthetic ORF, were predicted to contain a total of seven transmembrane helices. The predicted topology of these transmembrane helices was similar to the transmembrane regions of canonical receptors of the 7TM – diverse intracellular signaling family (*7TMR-DISM_7TM*, PF07695) (Anantharaman and Aravind 2003). Consistent with this observation, each of these putative ORFs shows greatest sequence similarity to the N-terminal transmembrane regions of putative multi-sensor signal transduction histidine kinases (GenBank accessions ACL15452 and ACA59209, respectively).

RM35, the second *R. metallidurans* based antibacterially active hit identified in overlay assays, lost activity upon restreaking and the cosmid isolated from this clone did not confer antibacterial activity to *R. metallidurans* upon retransformation (Figure 14B). This cosmid was fully sequenced (GenBank Accession GQ869384), and the eDNA insert was found to contain a 16 kb non-ribosomal peptide synthetase (NRPS) based gene cluster that resembles lipopeptide gene clusters known to produce antibiotics. A description of the RM35 NRPS based gene cluster and its predicted product can be found in Section 3.3.6.

In screens using libraries hosted in a second representative *β-Proteobacteria*, *B. graminis*, two antibacterially active clones surrounded by hazy zones of growth inhibition and six clones displaying light-brown to dark-

brown pigmentation were identified. Neither antibacterially active clone reproduced the antibacterial phenotype upon retransformation. As seen with brown clones found in *P. putida* and *R. metallidurans* hosted libraries, all of the brown *B. graminis* clones showed evidence of elevated heme metabolite production by normal phase TLC, and thus none of these clones were pursued further.

3.3.4 α-Proteobacteria

A non-pathogenic strain of the rhizosphere-dwelling bacterium *A. tumefaciens* was chosen as a representative of the *α-Proteobacteria* for its ease of transformation and compatibility with IncP group plasmids (Schmidhauser and Helinski 1985). Although no antibacterially active clones were detected in overlay assays, one yellow-pigmented clone and two clones with altered colony morphology were identified during screens using this host.

The yellow-pigmented clone, AT1, found while screening the JPA library (Figure 14F), was identical to the cosmid clone RM3 described in Section 2.3 as conferring the production of color to *R. metallidurans* (Craig, Chang et al. 2009). This clone contained a six-gene operon that was predicted to code for the production of the carotenoid β-carotene. This was the only case where the same clone was identified in screens using different bacterial hosts. The carotenoid biosynthetic gene cluster found in this clone sat directly adjacent to and collinear with the N-terminal fragment of the *sacB*

gene into which eDNA inserts are cloned in pJWC1. Only 59 bp of eDNA was captured upstream of the first ORF in the carotenoid biosynthetic operon, suggesting that the vector associated *sacB* promoter, and not an eDNA derived promoter, may be responsible for driving the expression of this gene cluster in both *R. metallidurans* and *A. tumefaciens*.

The two *A. tumefaciens* clones that displayed altered colony morphology, AT3 and AT5 (Figure 14G), were identified from the JPA and RPA libraries, respectively. These clones shared a similarly "wrinkled" appearance (i.e. crenated), and both were viscous upon physical manipulation. While recent studies suggest that small molecules may play a role in regulating colony phenotypes by exerting control over motility, adhesion and biofilm formation, no clone-specific metabolites were identified in the culture broths or cell pellets of either clone (Bassler and Losick 2006). The eDNA inserts from both AT3 and AT5 were fully sequenced (GenBank accessions GQ869381 and GQ869382), and no putative ORFs with homology to known biosynthetic enzymes were identified.

eDNA library screens using a second representative *α-Proteobacteria*, *C. vibrioides*, did not result in the identification of any clones displaying pre-defined phenotypes.

3.3.5 Retransformation of active cosmids into alternative hosts

The three eDNA libraries were screened at multiple coverage to help ensure that all library clones were examined in each host. To address the possibility that some cosmid DNAs were not actually screened in each host species during the initial high-throughput screens, each cosmid of interest was transformed separately into all six hosts. These transformants were then examined in phenotypic assays to determine the species specificities of the host-clone-phenotype relationships (Figure 17). In only one case did a clone confer an easily observable phenotype to a host species other than the one used for the initial clone identification. While these eDNA clones did not generally confer easily observable phenotypes to other hosts, there were several examples where new clone-host pairs were found to produce lower amounts of secondary metabolite(s) than that required for detection in the original phenotypic screens.

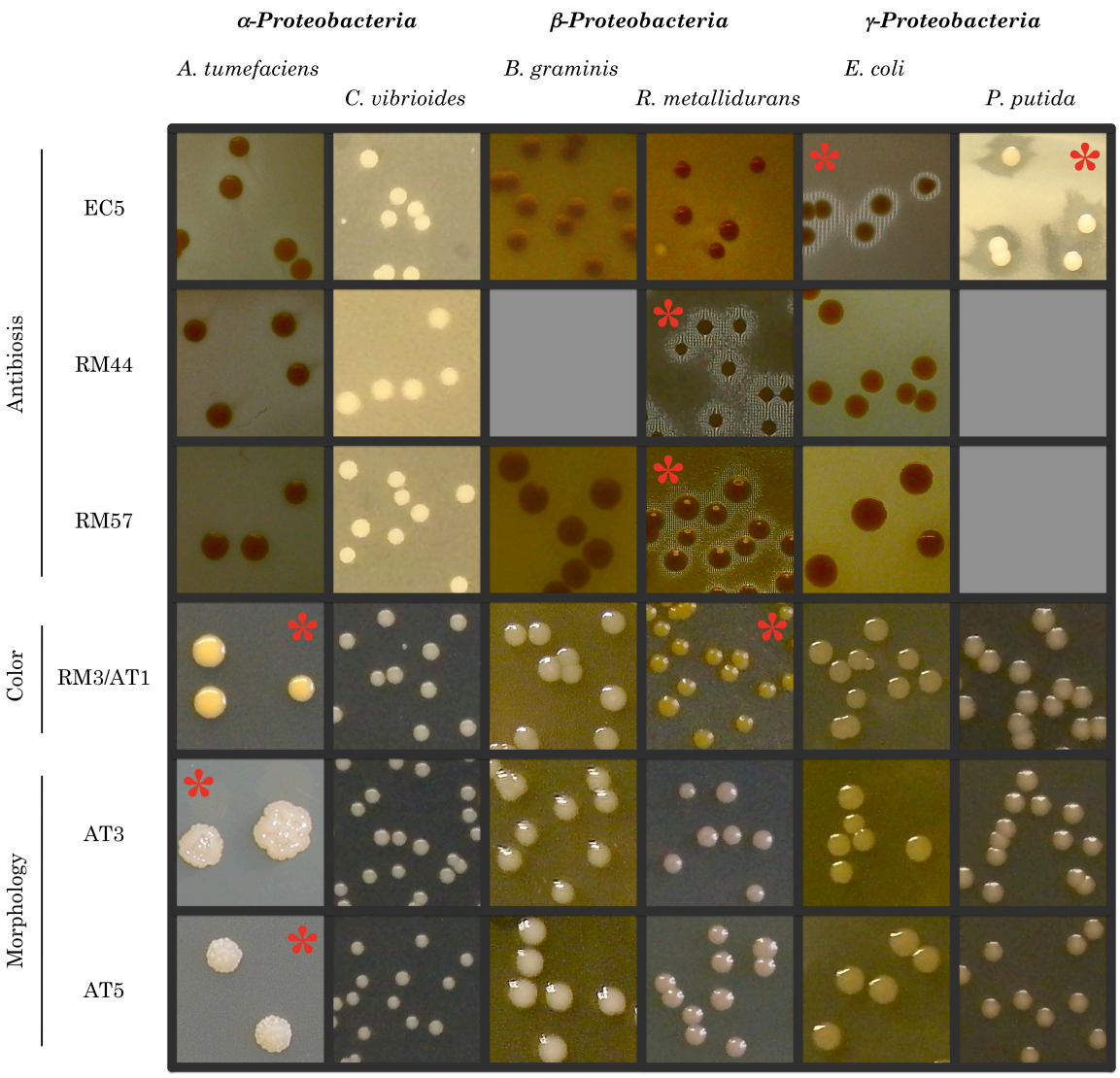


Figure 17: Phenotypic relationships of clone-host pairs.
Red asterisks indicate the expression of a clone-specific phenotype.

Upon re-transformation, clones RM3/AT1, RM44, RM57, AT3 and AT5 all conferred the originally observed phenotypes to the hosts they were initially recovered from, but none of these clones conferred phenotypes to any other host. Cosmid RM44 could not be successfully transformed into *B. graminis* despite numerous trials, and although clone RM44 did not confer an

antibacterial phenotype to *A. tumefaciens*, the novel metabolite *N*-myristoyl-histidinol (**28**) was recovered from the ethyl acetate extract of A. tumefaciens [RM44] cultures (see Figure 34 and Figure 35 in the Appendix).

Even though high-throughput screening of *P. putida* based libraries was not feasible using the two-layer overlay method, individual cosmid clones hosted in *P. putida* were amenable to this technique. All cosmids that conferred antibacterial activity to other hosts were therefore subjected to overlay assays in *P. putida.* From these overlay assays, cosmid EC5 was found to confer antibacterial activity to *P. putida* (Figure 18). HPLC analysis of the ethyl acetate extract of the culture broth of *P. putida* transformed with cosmid EC5 revealed the presence of mixed *N*-acyl aromatic amino acids in amounts comparable to that found using *E. coli* as the expression host (Figure 14A). A similar analysis was then performed on extracts of the culture broths of the remaining four hosts that did not produce an obvious antibacterial phenotype when transformed with the EC5 cosmid. This analysis revealed trace amounts of mixed *N*-acyl aromatic amino acids in extracts of both the *A. tumefaciens* and *B. graminis* culture broths. These trace levels were apparently too low to be detected by a two-layer overlay screen.

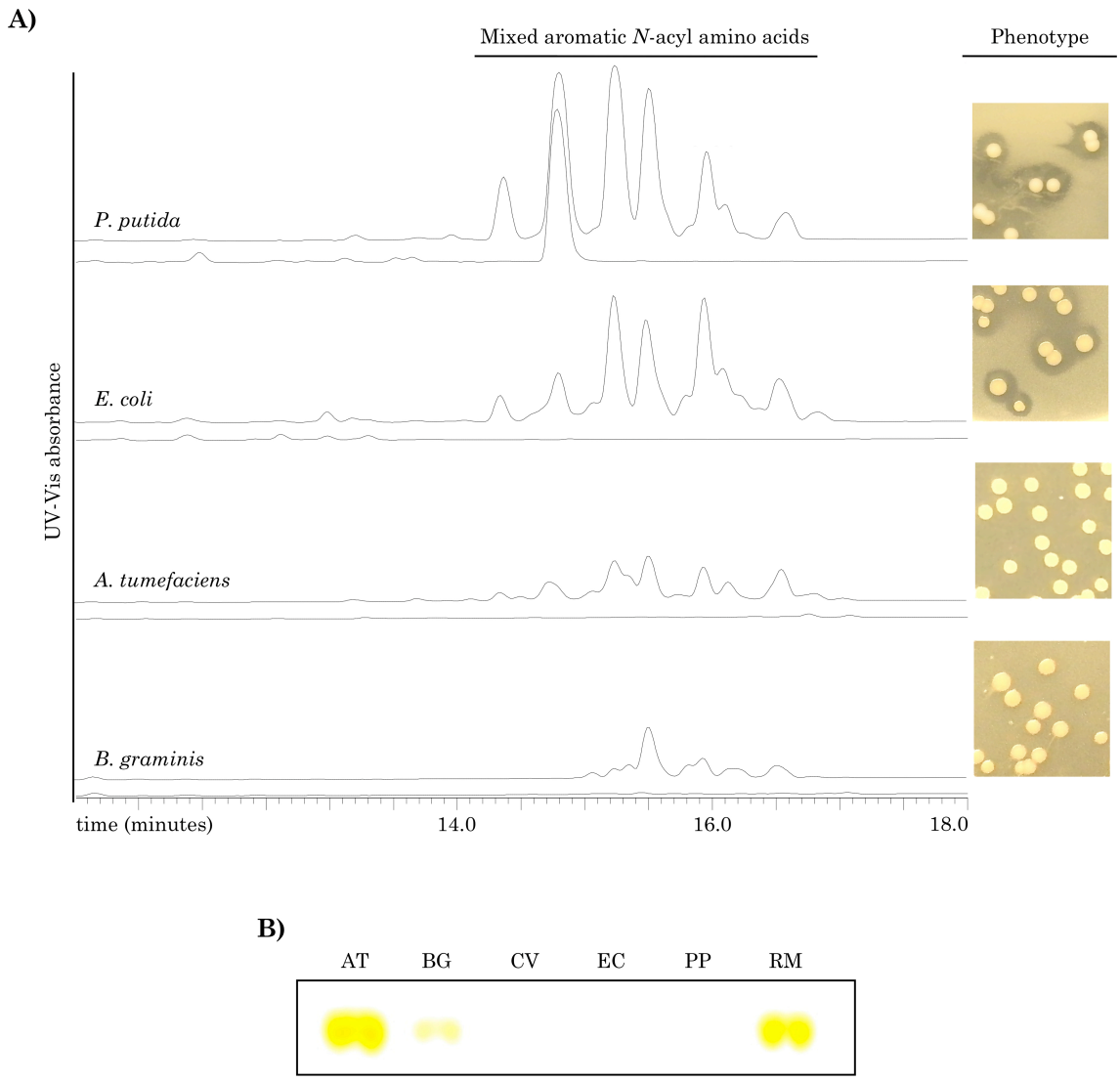


Figure 18: Host-chemotype analyses of clones EC5 and RM3/AT1.

(A) Analytical HPLC traces of the organic extracts derived from cultures of different *Proteobacteria* transformed with clone EC5. HPLC traces of extracts derived from vector control cultures are shown below each EC5 HPLC trace. The region containing *N*-acyl aromatic amino acids is highlighted. Representative images of the phenotype observed for each host in an antibacterial overlay assay is shown alongside the HPLC trace. (B) Analytical TLC analysis was performed on acetone extracts from the cell pellets of different *Proteobacteria* harboring cosmid RM3/AT1. The region of the TLC plate that is predicted to contain β-carotene is shown (AT - *A. tumefaciens*, BG - *B. graminis*, CV - *C. vibrioides*, EC - *E. coli*, PP - *P. putida*, RM - *R. metallidurans*).

Despite the potential for vector driven expression of the RM3/AT1 carotenoid biosynthetic gene cluster, no hosts other than *R. metallidurans* and *A. tumefaciens* were found to express the intense yellow-pigmented phenotype associated with this cosmid. Because of the potential role the vector-associated promoter may have played in the production of carotenoids by RM3/AT1, the production of β-carotene by each host was assessed by TLC analysis of acetone-extracted cell pellets. Significant amounts of β-carotene were detected only in the extracts from *R. metallidurans* and *A. tumefaciens* (Figure 18B). A small amount of β-carotene was also found in the extract derived from *B. graminis*, although this amount of pigment was not sufficient to appreciably alter the pigmentation of *B. graminis* colonies.

3.3.6 Environmental Clone RM35

Two large NRPS megasynthases constitute the majority of the RM35 gene cluster. The most closely related sequences to the RM35 NRPS-megasynthases are a portion of the putative linear pentadecapeptide gramicidin synthetase LgrB from *Brevibacillus brevis* NBRC 100599 (GenBank Accession BAH43870) and a portion of an amino acid adenylation domain-containing protein from *Nostoc punctiforme* PCC 73102 (GenBank Accession ACC85408), respectively. Upstream of these two large NRPS ORFs is the C-terminal fragment of a putative β-keto-acyl synthase, an isolated acyl-carrier protein and a methyltransferase. These putative genes, in combination with the domain architectures and the adenlyation domain

specificities (Ansari, Yadav et al. 2004) of the two NRPS megasynthases, suggest this gene cluster produces the acylated tripeptide, Acyl-Val-Cys-Phe (Figure 19). Lipodepsi-tripeptide antibiotics have been isolated from soil and marine fungi, as well as from a marine alga (Morino, Masuda et al. 1994; Hamann, Otto et al. 1996; Abbanat, Leighton et al. 1998; Schlingmann, Milne et al. 1998; Namatame, Tomoda et al. 1999).

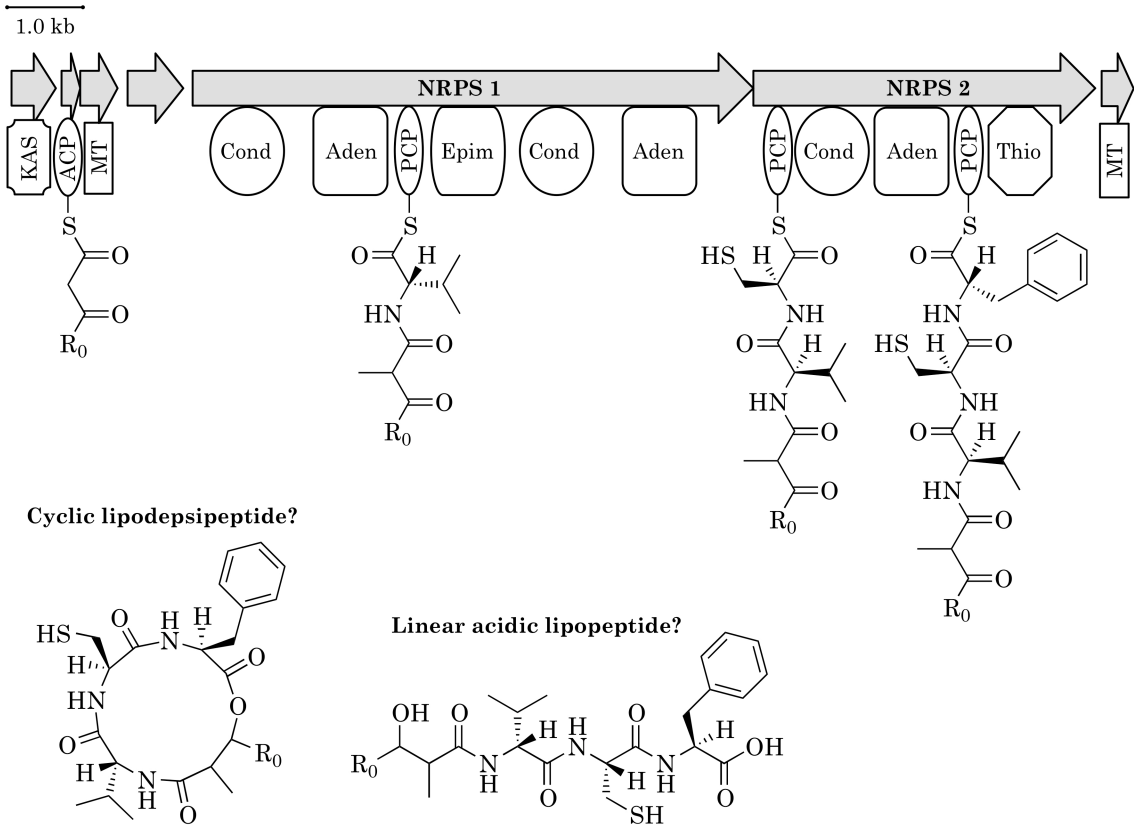


Figure 19: The RM35 NRPS-based gene cluster.

The eDNA-derived non-ribosomal peptide synthetase (NRPS)-based gene cluster from clone RM35 hosted by *R. metallidurans* is predicted to produce an acylated tripeptide. Abbreviations: Cond. – condensation domain, Aden. – adenylation domain, PCP – peptidyl-carrier-protein domain, Epim. – epimerization domain, Thio. – thioesterase domain, KAS – β-keto-(acyl-carrier-protein) synthase III domain, ACP – acyl-carrier-protein, MT – methyltransferase domain.

3.4 Discussion and Future Directions

The majority of bacteria from soil and aquatic habitats are not readily cultivated using standard microbiological methods. The development of methods to study these organisms by examining DNA cloned directly from environmental samples has been essential to the culture-independent analysis of microbial ecosystems. To access metagenome-derived enzymatic and biosynthetic abilities without prior knowledge of the genetic elements required for conferring these abilities, the emerging field of functional metagenomics relies on screening arrayed metagenomic libraries for clones displaying functional attributes and phenotypes. The success and/or failure of functional metagenomics is a function of the host strain, which is utilized to process the foreign genetic material within eDNA libraries. As the most commonly used host for metagenomic libraries, *E. coli* is estimated to readily express only 40% of genes derived from diverse microbial origins, with a strong bias against the expression of genes from certain groups of distantly related organisms (Gabor, Alkema et al. 2004). An expansion of the host repertoire used in functional metagenomics is therefore likely to increase both the number and diversity of eDNA clones identified from functional metagenomic studies.

The construction of eDNA libraries that are compatible with an expanded host repertoire could be achieved either through the parallel construction of libraries with unique origins of replication or through the

construction of eDNA libraries using broad host-range cloning vectors (Martinez, Kolvek et al. 2004; Li, Wexler et al. 2005; Wexler, Bond et al. 2005). The IncP1-*α* group of broad host-range plasmids boasts perhaps the most extensive list of compatible species and therefore represented an obvious choice as the foundation for a broad host-range eDNA cloning vector (Adamczyk and Jagura-Burdzy 2003; Aakvik, Degnes et al. 2009). Plasmids possessing the RK2-replicon, found within many IncP1-*α* group derivative cloning vectors, are thought to be stably maintained at low copy number in the majority of gram-negative bacterial species, and in some instances can be stably transferred to gram positive bacteria, yeast and mammalian cells (Friedman, Long et al. 1982; Heinemann and Sprague 1989; Waters 2001). From the work described in this Chapter (Chapter 3), it is obvious that the diverse host range of RK2-based cloning vectors may be taken advantage of to create eDNA libraries that are stably maintained in multiple different bacterial expression hosts.

It is clear that not all hosts perform equally well when subjected to simple phenotypic assays, and it remains a distinct possibility that some surrogate expression hosts will be better suited to using foreign genetic material than others. High-throughput screens of *E. coli*, *R. metallidurans* and *A. tumefaciens* based libraries each yielded clones that confer the production of small molecules to heterologous hosts, while the *C. vibrioides* based libraries did not yield any clones displaying phenotypes of interest.

Some phenotypes were also more readily detected in phenotypic screens than others (Table 7). In particular, heme and melanin producing clones, which have been reported numerous times from screening *E. coli* based libraries, were found in *P. putida*, *R. metallidurans* and *B. graminis* based libraries. Beyond screening for clones that produce antibiotics or colored metabolites, any easily reproducible phenotypic change has the potential to be exploited for the identification of new molecules or genes by functional metagenomics. The altered morphologies seen in environmental clones AT3 and AT5, hosted by *A. tumefaciens*, exemplify how the heterologous expression of eDNA can be used to confer readily observable phenotypes to a receptive expression host.

The functional metagenomic screens described in this Chapter (Chapter 3) also serve to highlight an important requirement for functional metagenomic screening. For a small molecule derived phenotype to be detected using functional metagenomics, the production of compound(s) associated with a given phenotype must be at or above a certain threshold level. The ability of several hosts to produce detectable amounts of clone-specific metabolites in sensitive analytical assays, yet not display relevant clone-specific phenotypes, appears to be the result of sub-threshold small molecule production by specific clone-host pairs. The propensity of some hosts to more commonly reach and surpass these threshold levels is an important mechanism by which individual hosts will likely confer an advantage to functional metagenomic screening efforts.

From the work described in Chapter 2, *R. metallidurans* was established as the first gram-negative host other than *E. coli* to lead to the successful identification of secondary metabolite producing clones from eDNA libraries (Craig, Chang et al. 2009). From the work described in this Chapter (Section 3.3.3), *R. metallidurans* again proved itself to be a valuable addition to the functional metagenomics screening toolkit, as screens of the RPA library identified several antibacterially active clones, including an antibacterially active clone that produced the novel natural product *N*-(4-amino-2-hydroxybutyl)-tetradecanamide. This clone, RM44, along with clone RM35, which harbors an orphan NRPS-type biosynthetic gene cluster, and the previously identified small molecule producing eDNA clones RM3 and RM57, demonstrate the potential *R. metallidurans* has for identifying biosynthetic genes and gene clusters from eDNA (Figures 16-19). The continued exploration of phylogenetically diverse bacteria as hosts for functional metagenomic screening is likely to identify additional microbes that will be rewarding hosts for exploring soil metagenomes.

During the course of these high-throughput eDNA library screens there was only one instance where the same small molecule producing clone was identified using two different host species. Distantly related host species therefore appear unlikely to identify the exact same subset of eDNA clones from phenotype-dependent screens of soil eDNA libraries, despite nearly identical assay conditions and selection criteria. This provides tangible

evidence that the use of broad host-range vectors can be effectively exploited for the creation of soil eDNA libraries suitable for functional metagenomic studies using multiple expression hosts. More importantly, these findings suggest that doing so is likely to expand the functional and genetic diversity of eDNA clones obtained from such screens as well as increase the number of phenotype-specific lead clones available for further study.

3.5 Materials and Methods

3.5.1 Environmental DNA isolation and library construction

Two soil samples collected in Pennsylvania (one from a deciduous forest and the other from the mud at the bottom of a small creek bed) and one collected in Oregon (from a cold desert covered with sand and clay) were used as sources for eDNA. Soils were collected manually using clean gardening spades and stored at 4°C for less than two weeks before processing. Large debris (rocks, sticks and roots) was removed prior to the addition of soil lysis buffer by passing each soil sample sequentially through 3.35 and 1.00 mm sieves. Following the general detergent-based lysis strategy (SDS and CTAB) of Zhou et al., a one to one mixture of soil and soil lysis buffer was heated briefly at 70°C, after which any remaining particulate matter was removed by centrifugation (Zhou, Bruns et al. 1996). Crude eDNA was precipitated from the resulting supernatant using isopropanol and collected by centrifugation. All crude eDNA samples were then washed with 70% ethanol

and resuspended in TE buffer (10 mM Tris–HCl, 1 mM Na EDTA, pH 8.0). The remaining soil particulate matter and humic substances were removed by large–scale gel purification on a 1% agarose gel (16 hours at 20 V). Purified HMW eDNA was recovered from the gel by electroelution (2 hours at 100 V) and concentrated by isopropanol precipitation. Environmental DNA was then blunt–end repaired (Epicentre®, End–It), ligated into ScaI digested and CIP–treated pJWC1 (Craig, Chang et al. 2009), packaged into λ-phage *in vitro* (Epicentre®, MaxPlax Packaging Extracts) and transfected into *E. coli* EC100 cells. Each 100,000–500,000 membered eDNA library was prepared as a collection of unique 20,000–50,000 membered sublibraries. The individual sublibraries were mini–prepped and the resulting DNA was used for either the direct transformation of heterologous expression hosts or for the transformation of the *E. coli* conjugal mating strain S17.1.

3.5.2 Preparation of electrocompetent cells

Electrocompetent cells were prepared using established protocols with only minor deviations. Briefly, all strains were grown at 30°C in their preferred media (*R. metallidurans* CH34, SOB medium (Taghavi, van der Lelie et al. 1994); *C. vibrioides* CB15, Caulobacter medium (CM, ATCC medium 36) (Gilchrist and Smit 1991); *B. graminis* C4D1M, Luria–Bertani medium (Dower, Miller et al. 1988); *A. tumefaciens* LBA4404, Yeast extract–Mannitol medium (YM) (Lin 1994)) to OD_{600nm} values between 0.5 and 1.0,

chilled on ice, harvested by centrifugation, washed three times with ice cold 10% glycerol, resuspended in 10% glycerol and flash frozen.

3.5.3 Transformation of eDNA cosmid libraries by electroporation and conjugation

Individual 80 μL aliquots of electrocompetent cells were thawed on ice and mixed with 10 μL of cosmid DNA (250 ng mL^{-1}), transferred to 1.0 mm electroporation cuvettes, pulsed at 1.8 kV (2.20 kV for *A. tumefaciens*) for 6 ms using a BioRad MicroPulser and then immediately mixed with 1 mL of recovery medium (SOC for *B. graminis* and *R. metallidurans*, YM for *A. tumefaciens*, and CM for *C. vibrioides*). The resulting cell suspensions were incubated with shaking at 30°C for 2–3 hours and subsequently plated at titers of 1,500–2,500 colonies per 150 mm diameter plate onto the following selection plates: *B. graminis* and *R. metallidurans*, LB–tetracycline (20 μg mL^{-1}); *A. tumefaciens*, LB–tetracycline (5 μg mL^{-1}); and *C. vibrioides*, CM–tetracycline (4 μg mL^{-1}). Biparental conjugation was used to transfer eDNA libraries from *E. coli* S17.1 to *P. putida* as previously described (Martinez, Kolvek et al. 2004). Tetracycline (50 μg mL^{-1}) and irgasan (20 μg mL^{-1}) were used to select for exconjugants and counter select against the *E. coli* donor, respectively. After 1-2 days exconjugants were scraped from the selection plates, resuspended in 10% glycerol and flash frozen in liquid nitrogen. Dilutions of frozen *P. putida* stocks containing the eDNA libraries were then

titered and plated at 500 colonies per plate onto 150 mm diameter plates of LB–tetracycline (50 μg mL^{-1}).

3.5.4 Phenotypic Screening of eDNA libraries

eDNA libraries transferred to individual host *Proteobacteria* were plated to achieve between 1.5 and 3–fold coverage of each library. After recovering pigmented colonies and colonies displaying altered morphologies, the screening plates were overlaid with a thin layer (15 mL $plate^{-1}$) of LB–top–agar (LB with 6 g L^{-1} agar) containing a 1:200 dilution of *B. subtilis* 1E9 (grown in LB–tetracycline to an OD_{600nm} = 1.0). Overlaid plates were allowed to solidify at room temperature before being transferred to 30°C. After 24–48 hours at 30°C the overlayed plates were screened visually for colonies that produced a zone of growth inhibition in the resulting *B. subtilis* lawn. Colonies displaying this phenotype were picked with sterile toothpicks, streaked onto selective media and retested to confirm the original phenotype. Cosmid DNA from each hit, regardless of host origin, was mini–prepped using standard Qiagen® protocols, then electroporated into CopyControl *E. coli* EC300 (Epicentre®) to facilitate genetic analysis. Each cosmid was sequenced by 454 pyrosequencing, and the sequences were analyzed for putative open reading frames (ORFs) using GLIMMER3.02 (Delcher, Harmon et al. 1999) and SoftBerry FGENESB: Bacterial Operon and Gene Prediction Program

(http://linux1.softberry.com/berry.phtml?topic=fgenesb&group=programs&su

bgroup=gfindb). Digital images were captured with a Nikon COOLPIX S550 digital camera and minimally processed (image cropping, brightness and contrast adjustment) with Microsoft Office Picture Manager.

3.5.5 Organic extraction, compound isolation, structural analysis

One liter liquid cultures (LB, tetracycline 20 μg mL^{-1}, kanamycin 12 μg mL^{-1}) of *R. metallidurans* CH34 clone RM44 were grown at 30°C with shaking (200 rpm) for 3-4 days. Liquid cultures were then extracted with ethyl acetate and the resulting extracts were dried under vacuum. Crude extracts were resuspended in a minimal volume of HPLC–grade methanol and then subjected to preparative–HPLC under the following conditions: (Waters XBridge C_{18} column [10 x 150 mm], 7 mL $minute^{-1}$ flow rate); 5 minutes at 50:50 H_2O:methanol with 0.1% formic acid, followed by a linear gradient of 50:50 H_2O:methanol with 0.1% formic acid to 100% methanol 0.1% formic acid over 15 minutes, followed by 100% methanol with 0.1% formic acid for 4 minutes. Pooled fractions containing compounds **9** and **10** (minutes 11.5–13.5) were resuspended in 50:50 methanol:acetonitrile and subjected to a second round of preparative–HPLC: 2 minutes at 25:75 acetonitrile:H_2O with 0.1% formic acid, followed by a linear gradient of 25:75 acetonitrile:H_2O with 0.1% formic acid to 100% acetonitrile over 20 minutes, followed by 100% acetonitrile for 2 minutes. Fractions containing pure compound **9** (minutes 8.5–10.0) were pooled and dried under vacuum resulting in a final yield of ~2.5 mg L^{-1}. Fractions containing pure compound

10 (minutes 8.5–10.0) were pooled and dried under vacuum resulting in a final yield of ~0.25 mg L^{-1}.

3.5.6 Cloning and expression of *nasA* in *E. coli*

The 657 bp *nasA* gene was PCR amplified (30 cycles of 97°C for 30 sec, 65°C for 30 sec, and 72°C for 15 sec) from cosmid EC5 using the following primers: EC5–NAS–HindIII–F: 5'–ACCATGAAGCTTCTGATTCGCCCGTGTATCGGGACAG–3', EC5–NAS–BglII–R: 5'–GGAGATCTTTCATCGGAGCGCCTCCGATTCCATGATC–3'. This primer set introduces the following amino acid changes into nasA: S2K and I3L. The PCR amplicon was double–digested with HindIII and BglII and then ligated to the correspondingly double–digested/CIP–treated pTAC–MAT–Tag–2 expression vector (Sigma-Aldrich® E5405). LB cultures (0.1 mM IPTG, 100 μg mL^{-1} ampicillin) of the *nasA* expression construct in *E. coli* EC100 were incubated at 30°C with shaking for three days and then extracted with an equal volume of ethyl acetate. The dried extracts were analyzed by normal phase TLC and reversed phase analytical HPLC for the presence of clone-specific metabolites. Analytical reversed phase HPLC conditions were as follows: (Waters XBridge C_{18} column [4.6 x 150 mm], 1.5 mL $minute^{-1}$ flow rate); 3 minutes at 80:20 H_2O:methanol with 0.1% formic acid, followed by a linear gradient from 80:20 H_2O:methanol with 0.1% formic acid to 100% methanol with 0.1% formic acid over 12 minutes, followed by 100% methanol with 0.1% formic acid for 5 minutes.

3.5.7 Nucleotide sequence Accession Numbers

The GenBank Accession Numbers corresponding to individual cosmid eDNA clones are: EC5, GQ869383; EC6, GQ869384; RM3/AT1, FJ151553 (Craig, Chang et al. 2009); RM57, FJ151552 (Craig, Chang et al. 2009); RM44, GQ869386; RM35, GQ869384; AT3, GQ869381; AT5, GQ869382.

3.5.8 Compound 9, 3-hydroxy myristoylputrescine (IUPAC: *N*–(4–amino–2–hydroxybutyl)tetradecanamide)

White powder; ^{1}H NMR (600 MHz, methanol-*d*4) 3.79 (C3, 1H, m), 3.29 (C4, 1H, dd, 13, 5.5), 3.18 (C4,1H, dd, 13, 7.4), 3.07 (C1, 2H, m), 2.23 (C6, 2H, t, 7.6) 1.82 (C2, 1H, m), 1.68–1.62 (C2, C7, 3H, m), 1.4–1.2 (m), 0.92 (C18, 3H, t, 7.0); ^{13}C NMR (150 MHz, methanol-*d*4) 176.9 (C5), 69.9 (C3), 46.3 (C4), 38.5 (C1), 37.0 (C6), 33.1, 32.7 (C2), 30.8 (m), 30.6, 30.5 (m), 30.4, 27.0 (C7), 23.7, 14.4 (C18); HRESIMS *m/z* 315.2998 [M]+ (calcd for $C_{18}H_{39}N_2O_2$, 315.3006).

3.5.9 Compound 10, myristoylputrescine

White powder; ^{1}H NMR (600 MHz, methanol-*d*4) 3.20 (C4, 2H, t, 7), 2.94 (C1, 2H, t, 7.5), 2.18 (C6, 2H, t, 8), 1.65 (C2, 2H, m), 1.64-1.56 (C3, C7, 4H, m), 1.2-1.4 (m), 0.90 (C18, 3H, t, 7); ^{13}C NMR (150 MHz, methanol-*d*4) 176.5 (C5), 40.3 (C1), 39.4 (C4), 37.2 (C6), 33.1, 30.8 (m), 30.6, 30.5 (m), 30.4, 27.5 (C3), 27.1 (C7), 26.0 (C2), 23.7, 14.4 (C18); HRESIMS *m/z* 299.3039 [M]+ (calcd for $C_{18}H_{39}N_2O$, 299.3057).

CHAPTER 4

4 Characterization of a metagenome-derived enzyme that produces branched-chain acyl-(acyl-carrier-protein)s from branched-chain 2-oxo acids

4.1 Chapter Summary

In bacteria, acyl-(acyl-carrier-protein)s (acyl-ACPs) serve as acyl-donating substrates for both primary and secondary metabolism. In addition to their principle role in lipid biogenesis, acyl-ACPs are used in the biosynthesis of a number of natural products, including *N*-acyl amino acids. The *N*-acyl amino acid synthase, *nasA*, is part of a two-gene biosynthetic operon, *nasAB*. NasB contained a previously uncharacterized domain-architecture predicted to encode for the production of acyl-ACPs. Heterologous expression of the *nasAB* operon in *Burkholderia graminis* C4D1M afforded the production of a novel *N*-acyl-phenylalanine containing a 13-carbon branched acyl-chain. Stable isotope feeding experiments in conjunction with homology arguments indicated that 2-oxoisocaproic acid, the product of L-leucine transamination, is used as a substrate by nasB to initiate the production of branched-chain acyl-ACPs. Analysis of nasB's multi-domain architecture suggests that this enzyme condenses the lipoate thioester of isovaleric acid with host-derived

malonyl-ACP to form 5-methyl-3-oxohexanoyl-ACP. Use of the endogenous acyl-carrier-protein enables the products of nasB to be recognized by the native type II fatty acid synthase system, thereby affording the production of branched long-chain acyl-ACPs that can subsequently be used in the biosynthesis of branched-chain *N*-acyl-phenylalanines by nasA. NasB and its homologs may be useful tools for generating branched-chain acyl-substrates within native metabolic backgrounds dominated by straight-chain fatty acids.

4.2 Introduction

Environmental DNA (eDNA) clones that confer the production of *N*-acyl amino acids to model cultured bacterial hosts are frequently identified in activity-based screens of soil DNA libraries. The *N*-acyl amino acids encoded by these clones are synthesized from acyl-(acyl-carrier-protein)s (acyl-ACPs, ACP) and amino acids by a diverse group of bacterial enzymes referred to as *N*-acyl amino acid synthases (NASs) (Brady, Chao et al. 2004; Van Wagoner and Clardy 2006). The acyl-ACP substrates of NAS enzymes are common intermediates in the *de novo* synthesis of fatty acids by the type II fatty acid synthase system (FAS II) (Lu, Zhang et al. 2004). In addition to *N*-acyl amino acids, FAS II-derived acyl-ACPs also serve as acyl-donors in the production of *N*-acyl homoserine lactone autoinducers, the cofactor lipoic acid and several lipopeptide antibiotics (Figure 20A, compounds **11**-**14**) (More, Finger et al. 1996; Schaefer, Val et al. 1996; Jordan and Cronan 1997; Val and Cronan 1998; Miller, Busby et al. 2000; Chooi and Tang 2010). During a

recent screen of soil metagenomic libraries for clones exhibiting antimicrobial activity, the NAS-containing clone (clone EC5), which confers the production of long-chain *N*-acyl-phenylalanines and *N*-acyl-tryptophans to multiple Gram-negative host species, was identified (Craig, Chang et al. 2010). The NAS-encoding gene found on clone EC5, *nasA*, is the first gene in a predicted two gene operon, *nasAB* (GenBank Accession No. GQ869383). The second gene in this operon, *nasB*, encodes for a putative 1118 residue protein that contains six well-characterized Pfam homology domains (Figure 20B) (Finn, Mistry et al. 2010). Based on generic functional predictions for these domains, this enzyme may be involved in the formation of ACP-linked fatty acids similar to those produced by bacterial FAS II systems. This Chapter (Chapter 4) describes results from a series of heterologous expression experiments in *Burkholderia graminis* C4D1M that show that nasB functions to provide branched-chain acyl-ACPs for nasA. The domain architecture of nasB also suggests that the production of branched-chain acyl-ACPs by nasB is independent of the soluble acyl-CoA intermediates required for canonical branched-chain fatty acid biosynthesis. Many bacteria, particularly Gram-negative species, do not natively produce branched-chain fatty acids (Lu, Zhang et al. 2004). NasB and its homologs are likely used by bacteria with straight-chain specific FAS II systems for generating branched-chain acyl-ACP substrates and may be useful as tools for synthesizing these substrates

in model heterologous expression systems in which they are not natively produced (e.g. *E. coli*).

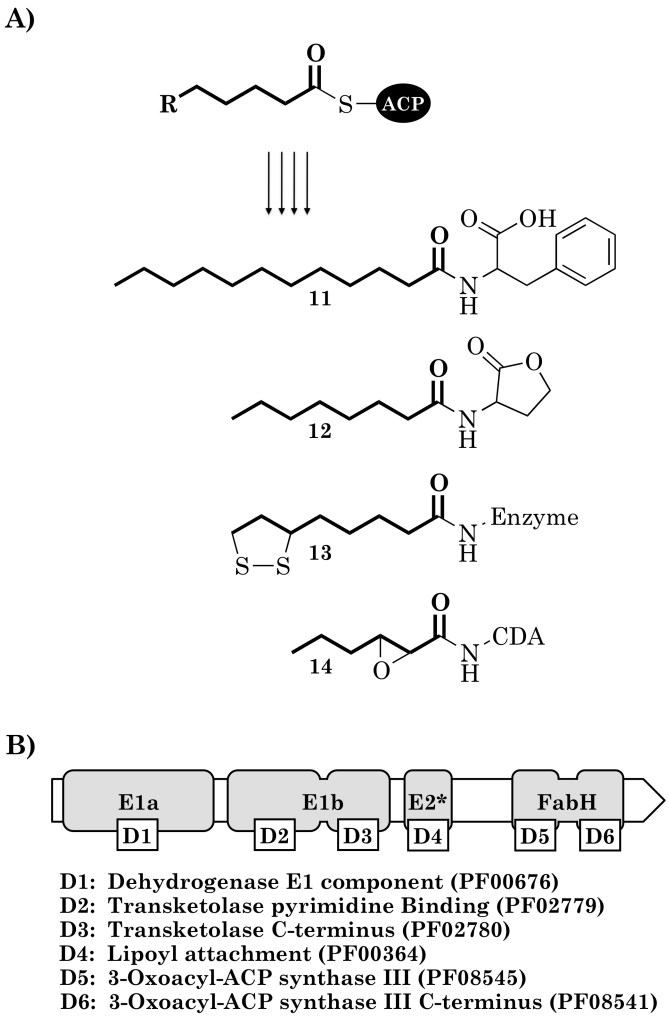


Figure 20: The domain architecture of nasB.

(A) Incorporation of FAS II intermediates (acyl-ACPs) into secondary metabolites and cofactors: *N*-dodecanoyl-phenylalanine (**11**), *N*-octanoyl homoserine lactone (**12**), enzyme-bound lipoic acid (**13**), and calcium-dependent antibiotic (CDA) (**14**). (B) The domain architecture of nasB includes all three Pfam homology domains found in the E1α/E1β subuntis of 2-oxo acid dehydrogenase complexes, as well as both of the Pfam homology domains found in the *E. coli* FabH protein (β-ketoacyl-ACP synthase III). Separating these two sets of domains is an E2-like (E2*) lipoyl-attachment site.

4.3 Results

4.3.1 Domain architecture of nasB

The N-terminal region of nasB contains the same set of protein domains that are found in the E1α and E1β subunits of 2-oxo acid dehydrogenase complexes (e.g. the pyruvate dehydrogenase complex, which converts pyruvate to acetyl-CoA) (Perham 1991). Following these domains is an E2-like (E2*) lipoyl-attachment site. Although lipoyl-attachment sites are found in the E2 subunits of 2-oxo acid dehydrogenase complexes, nasB lacks the catalytic lipoamide acyltransferase domain found at the C-terminus of canonical E2 subunit proteins (Pfam PF00198; exchanges the acyl-lipoate thioester with CoA, yielding a soluble acyl-CoA) (Mattevi, Obmolova et al. 1993). Instead, the C-terminal portion of nasB contains both of the protein domains that are found in the *E. coli* FabH protein, a β-ketoacyl-ACP synthase III (KASIII)-type enzyme that catalyzes the initial condensation reaction of FAS II (acetyl-CoA and malonyl-ACP, forming acetoacetyl-ACP) (Rock and Cronan 1985).

4.3.2 Heterologous expression of the *nasAB* operon in *B. graminis*

To test the original hypothesis that nasB provides substrates for the biosynthesis of *N*-acyl amino acids, the *nasA* gene was subcloned alone and in combination with *nasB* (i.e. the *nasAB* operon) into the broad host-range vector pJWC1, and these constructs were then electroporated into *B.*

graminis for heterologous expression studies (Craig, Chang et al. 2010) (Table 8). Ethyl acetate extracts from cultures of *B. graminis* were examined by HPLC-MS for differences in the *N*-acyl amino acids synthesized by each construct. Extracts of both cultures contained mixtures of *N*-acyl-phenylalanines, including significant amounts of the 12-carbon and 14-carbon saturated *N*-acyl derivatives of phenylalanine [m/z $(M+H)^+$ = 348 and 376, compounds **11** and **15**, respectively] (Figure 21A-B). Interestingly, the most abundant mass peak in the *nasAB* extract was from a compound with m/z $(M+H)^+$ = 362 (Figure 21B), which was notably absent from the *nasA*-only extract (Figure 21A). The presence of this additional mass peak suggested that *nasB* afforded the production of a new 13-carbon saturated *N*-acyl-phenylalanine derivative.

Table 8: Strains, eDNA clones and plasmids discussed in Chapter 4.

Strain, eDNA clone or plasmid	Description	Source or reference
Strain		
B. graminis C4D1M	*β-Proteobacteria* host	(Viallard, Poirier et al. 1998), ATCC 700544™
eDNA clone		
EC5	Antibacterially active eDNA clone, contains *nasAB* operon, confers *N*-acyl amino acid production	(Craig, Chang et al. 2010)
Plasmids		
pJWC1	IncP1-α group broad host-range cosmid vector	(Craig, Chang et al. 2009)
pJWC1-nasA	pJWC1 with the *nasA* gene and its promoter	(Craig and Brady 2011)
pJWC1-nasAB	pJWC1 with the *nasAB* operon and its promoter	(Craig and Brady 2011)

4.3.3 Structural Characterization of a branched-chain *N*-acyl amino acid

To determine the nature of the presumed 13-carbon acyl-chain, the metabolite corresponding to m/z = 362 (**16**) was isolated from the ethyl

acetate extract of *B. graminis* [*nasAB*] culture broth. Analysis of the 1D ^{1}H and ^{13}C NMR spectra of **16** suggested this compound was structurally similar to previously reported *N*-acyl-phenylalanines (see Figure 38 and Figure 39 in the Appendix) (Clardy and Brady 2007; Craig, Chang et al. 2010). The presence of phenylalanine was also apparent from the fragments observed in the LRESI MS spectrum (*m/z* 166 and 120). In the ^{1}H NMR spectrum of **16**, signals arising from the fatty acyl-chain were similar to those observed in the spectra of straight-chain *N*-acyl-phenylalanines isolated from *E. coli*, except that the three-proton triplet arising from the terminal fatty acid methyl-group was replaced by a six-proton doublet (0.88 ppm, 6.63 Hz) (Figure 21C-D). In the ^{1}H-^{1}H COSY spectrum of **16**, these methyl group protons were coupled to a single methine proton (1.52 ppm, m) that was in turn coupled to the methylene envelope of the fatty acyl-chain. Collectively, this data indicated that compound **16** was the structurally novel *N*-acyl amino acid *N*-(11-methyldodecanoyl)-phenylalanine, which contains an *iso*-branched 13-carbon saturated acyl-chain (*iso*-C13:0). The fatty acid profile of wild-type *B. graminis* does not contain branched-chain fatty acids of any type (Viallard, Poirier et al. 1998), indicating that nasB functions to provide a source of branched-chain acyl-ACP substrates for nasA.

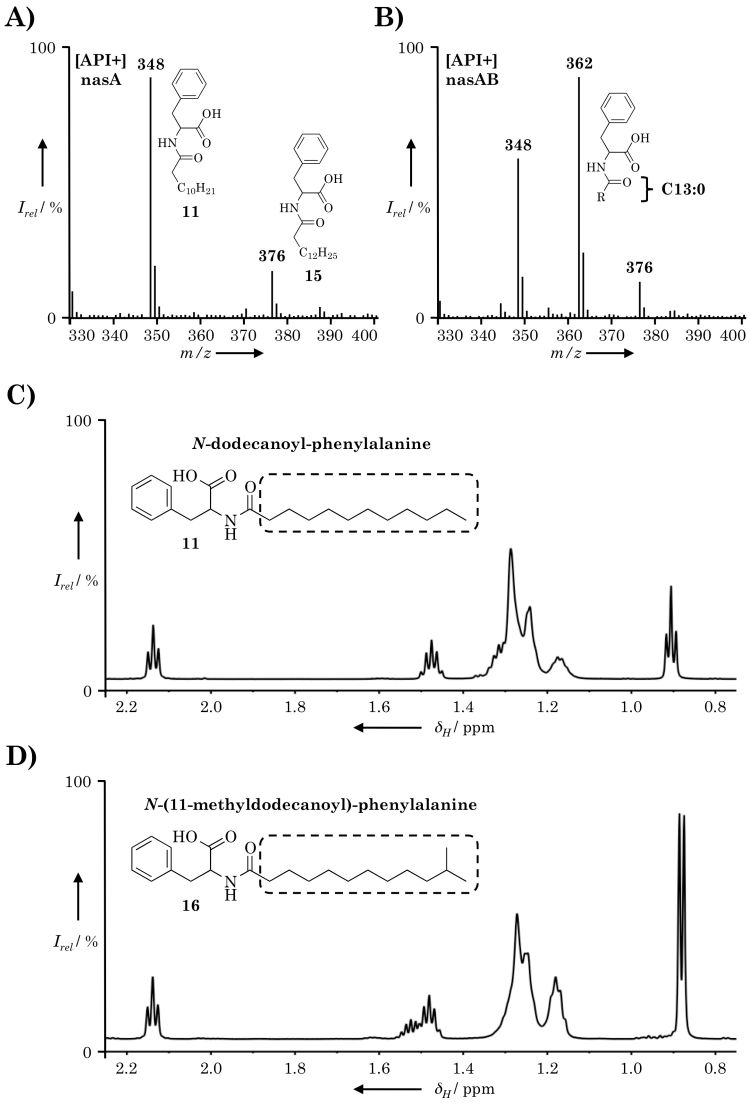


Figure 21: HPLC analysis of *B. graminis* [*nasAB*].

API-positive mode ionization data from the HPLC-MS analysis of extracts of *B. graminis* cultures transformed with (A) pJWC1-*nasA* and (B) pJWC1-*nasAB*. Comparison of the fatty acid region of the 1D ^{1}H NMR spectrum of (C) *N*-dodecanoyl-phenylalanine (**11**) with that of (D) *N*-(11-methyldodecanoyl)-phenylalanine (**16**).

4.3.4 Metabolic Labelling

The production of branched-chain fatty acids has been most extensively studied in the model Gram-positive bacterium *Bacillus subtilis*,

where the FAS II system is responsible for producing both *iso*- and *anteiso*-branched fatty acids (Kaneda 1991). The most important difference between the FAS II system of *B. subtilis* and that of *E. coli*, a model for straight-chain fatty-acid biosynthesis, is the substrate specificity of FabH, the β-ketoacyl-ACP synthase that performs the initial condensation reaction in FAS II-mediated fatty acid biosynthesis (Choi, Heath et al. 2000). In *B. subtilis*, but not in *E. coli*, FabH efficiently utilizes branched-chain acyl-CoA primers. These primers are produced from 2-oxo acids derived from the transamination of the proteinogenic branched-chain amino acids (Kaneda 1963; Willecke and Pardee 1971). The enzymatic machinery that carries out this function is a specialized 2-oxo acid decarboxylase referred to as the branched-chain α-keto acid dehydrogenase complex (BCKAD) (Oku and Kaneda 1988). BCKAD contains the standard repertoire of protein domains that are typically found in the E1α, E1β, E2 and E3 subunits of other 2-oxo acid dehydrogenase complexes.

The N-terminal domains of nasB are the same as those found in the E1α/E1β subunits of BCKAD, suggesting that nasB also uses 2-oxo acid substrates derived from the transamination of amino acids. The odd length and terminal branching pattern of the *iso*-C13:0 acyl-chain substructure of **16** suggested that L-leucine (**17**) was the likely source of the preferred 2-oxo acid substrate of nasB. To test this hypothesis and evaluate the 2-oxo acid decarboxylase activity of nasB, 1,2-$^{13}C_2$-labelled L-leucine was fed to a culture

of *B. graminis* [*nasAB*]. The most abundant mass peak in the organic extract of this culture was *m/z* $(M+H)^+$ = 363, one mass unit greater than that of unlabelled **16**. This result was consistent with the conversion of 1,2-$^{13}C_2$ L-leucine to 1,2-$^{13}C_2$ 2-oxoisocaproic acid by endogenous transaminase and the subsequent decarboxylation of this substrate by nasB to form a C1 isotopically-enriched isovaleryl-intermediate that feeds into the biosynthesis of **16** (Figure 22B; compare with Figure 22A). The mass peaks corresponding to linear-chain *N*-acyl-phenylalanines **11** and **15** (*m/z* $(M+H)^+$ = 348 and 376) were unaltered. To further confirm that the branched portion of the *iso*-C13:0 acyl-chain substructure of **16** originated from the branched side-chain of L-leucine, isopropyl-D_7-labelled L-leucine was also fed to a culture of *B. graminis* [*nasAB*]. As expected, this resulted in the appearance of a prominent mass peak with *m/z* $(M+H)^+$ = 369, corresponding to the incorporation of the leucine-derived isopropyl-D_7 label into **16** (Figure 22C). Again, the mass peaks corresponding to linear-chain *N*-acyl-phenylalanines were unaltered. In *B. graminis*, branched-chain amino acid aminotransferase (BCAT; GenBank Accession No. EDT09763) likely performs the initial transamination reaction that provides the 2-oxo acid **18** used by nasB.

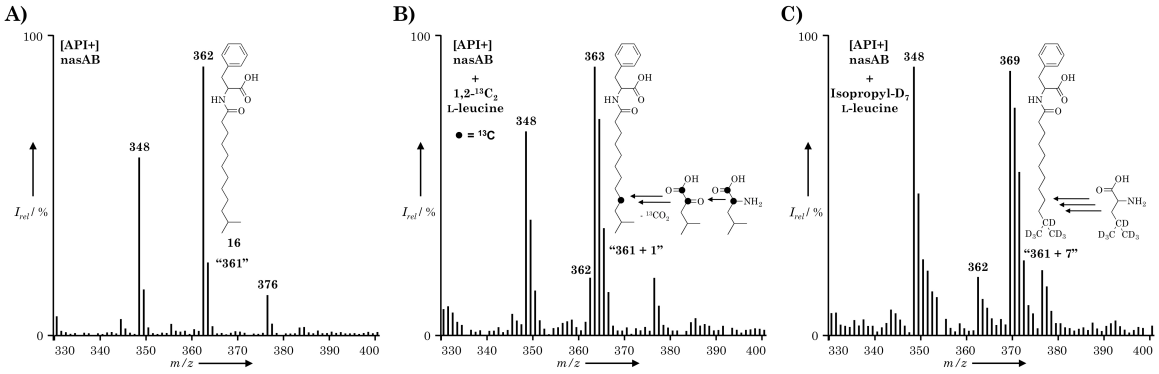


Figure 22: Stable isotope feeding experiments with *B. graminis* [*nasAB*].
API-positive mode ionization data from the HPLC-MS analysis of extracts of *B. graminis* [*nasAB*] cultures fed with either (A) nothing, (B) 1,2-$^{13}C_2$-labelled L-leucine or (C) isopropyl-D_7-labelled L-leucine.

4.3.5 Proposed mechanism of nasB function

These feeding experiments support a model in which the nasB enzyme preferentially converts the product of leucine transamination, 2-oxoisocaproic acid (**18**), to 5-methyl-3-oxohexanoyl-ACP (**20**) through an integrated series of biochemical reactions (Figure 23A). In this model, the E1α/E1β-like components of nasB first decarboxylate 2-oxoisocaproic acid (**18**), forming a covalent thiamine pyrophosphate (ThPP) adduct that is then transferred to the bound lipoic acid cofactor. As previously noted, nasB lacks the catalytic E2 lipoamide acyltransferase domain that would normally transfer this covalently-bound intermediate from the lipoic acid cofactor to CoA. The absence of this catalytic domain instead suggests that the FabH-like component of nasB condenses malonyl-ACP (**19**) directly with the enzyme-bound isovaleryl-lipoate thioester, forming 5-methyl-3-oxohexanoyl-ACP (**20**).

This FAS II intermediate can then undergo three rounds of canonical fatty acid chain elongation, yielding 11-methyldodecanoyl-ACP (**21**), which along with L-phenylalanine, is used by nasA for the production of *N*-(11-methyldodecanoyl)-phenylalanine (**16**) (Figure 23B).

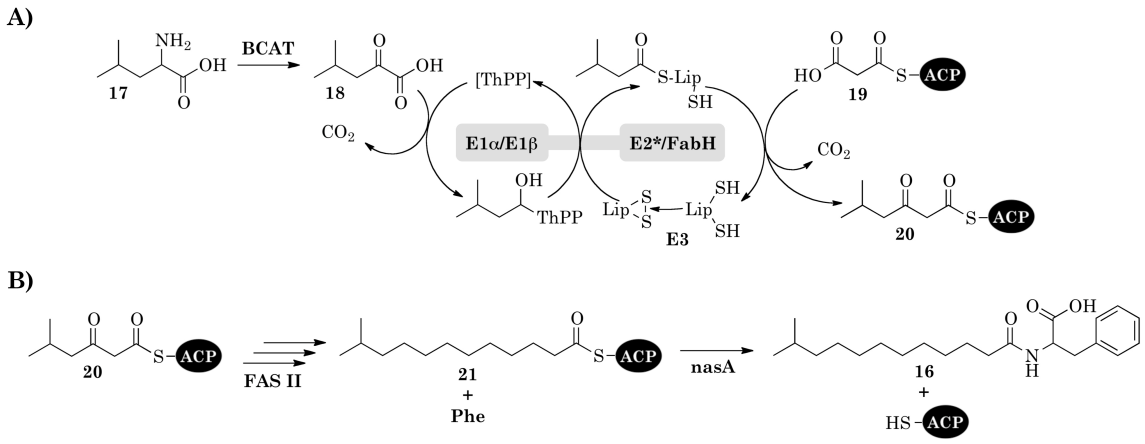


Figure 23: The biosynthesis of branched-chain *N*-acyl amino acids by nasA and nasB in *B. graminis*.

(A) The proposed condensation of 2-oxoisocaproic acid (**18**) and malonyl-ACP (**19**) by nasB to form 5-methyl-3-oxohexanoyl-ACP (**20**). (B) Scheme for the biosynthesis of *N*-(11-methyldodecanoyl)-phenylalanine (**16**), beginning with the FAS II-mediated chain elongation of **20**, yielding 11-methyldodecanoyl-ACP (**21**).

4.4 Discussion and Future Directions

The *nasAB* operon structure found on clone EC5 suggests that the branched-chain acyl-ACPs produced by nasB are specifically intended for use by nasA. Branched acyl-chains may therefore represent a characteristic feature used to distinguish the *N*-acyl amino acids produced by the uncultured EC5 bacterium from the more common straight-chain varieties

isolated from numerous other environmental DNA clones. While the precise role(s) of *N*-acyl amino acids remains to be determined, it has been proposed that they may function as signaling molecules in a manner similar to that of several other *N*-acylated small molecules, most notably the *N*-acyl homoserine lactones (e.g. compound **12**) (Brady, Chao et al. 2004). Variations in the acyl-chain length, degree of unsaturation and C-3 oxidation state of *N*-acyl homoserine lactones are the primary structural determinants used to target specific populations of receptor proteins (Winson, Camara et al. 1995; Dickschat 2010). Although less common by comparison, the incorporation of branched acyl-chains into *N*-acyl homoserine lactones has also been reported (Thiel, Kunze et al. 2009). The above examples highlight how acyl-chain diversification strategies can allow individual bacteria to increase the number of structurally and functionally unique metabolites they produce.

The genomes of three different species of bacteria were predicted to encode homologs of nasB containing the same putative domain architecture. These *nasB* homolgs do not form operon structures with *nasA* homologs, but are instead found within divergent genetic contexts (Figure 24). The *nasB* homolog from "*Cand. A. phosphatis*" is adjacent to a homolog of the *N*-acyltransferase lpxD, which transfers the acyl group from 3-hydroxytetradecanoyl-ACP to UDP-3-O-(3-hydroxytetradecanoyl)-D-glucosamine during lipopolysaccharide biosynthesis (GenBank Accession No. ACV34570; Pfam PF04613). The *nasB* homolog from *A. radiobacter* is

adjacent to a putative fatty acid hydroxylase superfamily enzyme predicted to be involved in the oxidative modification of lipids and sterols (GenBank Accession No. ACM27362; Pfam PF04116). Although the biosynthetic abilities of the known nasB homologs have not been investigated, the diversity of genes to which they are linked suggests that nasB homologs are involved in multiple different metabolic pathways.

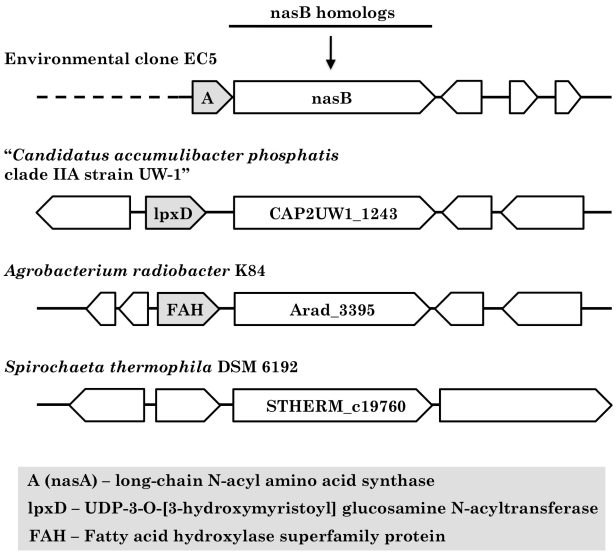


Figure 24: nasB homologs in sequenced bacterial genomes.
Putative nasB homolgs from sequenced bacterial genomes are found in a variety of genetic contexts. The labels for nasB homologs are taken from the corresponding KEGG database entries. The GenBank Accession Numbers for CAP2UW1_1243, Arad_3395 and STHERM_c19760 are ACV34571, ACM27363 and ADN02911, respectively.

The aggregation of distinct enzymatic domains enhances the catalysis of multi-step reactions (Reed 1974). This is particularly true for multi-enzyme complexes that utilize covalent transfer as a means of channeling substrates through successive active sites and for selecting against competing

metabolic pathways. In branched-chain fatty acid producing bacteria such as *B. subtilis*, BCKAD carries out the conversion of 2-oxo acids to the corresponding acyl-CoAs, which then become substrates for FabH. As a covalent alternative to the CoA-based strategy used by most bacteria, the FabH-like component of nasB may instead utilize the bound isovaleryl-lipoate thioester, eliminating the need for diffusible acyl-CoA intermediates along with the potential for these intermediates to be consumed by competing catabolic pathways. Although speculative, this proposed mechanism is supported by functional inferences for each of the Pfam homology domains found in nasB, as well as by the absence of the catalytic E2 lipoamide acyltransferase domain.

To function in a heterologous host like *B. graminis*, nasB must recognize host-derived malonyl-ACP. The use of host ACP allows the acyl-ACP products of nasB to be treated as native intermediates by the endogenous FAS II system (Choi, Heath et al. 2000), permitting subsequent utilization by acyl-ACP requiring enzymes that fail to discriminate between branched-chain and straight-chain substrates. The introduction of *nasB* into bacterial strains that normally produce only straight-chain fatty acids may therefore provide access to branched-chain variants of numerous membrane lipids and acylated small molecules. Previous attempts to engineer *E. coli* for the production of branched-chain fatty acids have relied on the replacement of *E. coli* FabH with alternative FabH homologs derived from bacterial

species that produce primarily branched-chain fatty acids. Such efforts have met with limited success (Smirnova and Reynolds 2001). The potential for nasB to be used as a tool for the engineering of branched-chain metabolites is intriguing and warrants additional study.

4.5 Materials and Methods

4.5.1 Bacterial culture conditions

Cultures of *Burkholderia graminis* C4D1M were grown aerobically in YENB medium [7.5 g L^{-1} yeast extract + 8.0 g L^{-1} nutrient broth] at 30°C on a rotary shaker (200 rpm). Broad host-range plasmids were selected for with tetracycline (30 μg mL^{-1}). Electrocompetent *B. graminis* cells were prepared according to Sharma and Schimke (1996) (Sharma and Schimke 1996).

4.5.2 Analytical and Preparative HPLC-MS

Organic extracts were subjected to analytical HPLC-MS using the following conditions: (Waters XBridge C_{18} 5 mm column [4.6 by 150 mm], 1.5 mL $minute^{-1}$ flow rate); 3 minutes at 50:50 H_2O:methanol with 0.1% formic acid, followed by a linear gradient from 50:50 H_2O:methanol with 0.1% formic acid to 100% methanol with 0.1% formic acid over 12 minutes, followed by 100% methanol with 0.1% formic acid for 5 minutes. Preparative HPLC-MS was performed using the same solvent conditions (Waters XBridge C_{18} 5 mm column [10 by 150 mm], 7 mL $minute^{-1}$ flow rate). Under these conditions,

compound **16** was eluted between minutes 14.4-14.6 at a final yield of approximately 1.5 mg L^{-1}.

4.5.3 Compound 16, *N*-(11-methyldodecanoyl)-phenylalanine

White crystalline powder; ^{1}H NMR (600 MHz, methanol-*d*4) 7.18-7.27 (C1 to C5, 5H, m), 4.67 (C8, 1H, dd, 4.77, 9.32), 3.22 (C7, 1H, dd, 4.72, 13.96), 2.93 (C7, 1H, dd, 9.61, 13.86), 2.14 (C11, 2H, t, 7.39), 1.52 (C20, 1H, m), 1.48 (C12, 2H, p, 7.29), 1.16-1.27 (C13 to C19, 14H, m), 0.88 (C21, C22, 6H, d, 6.63); ^{13}C NMR (150 MHz, methanol-*d*4) 176.3, 175.1, 138.8, 130.4, 129.5, 127.9, 55.1, 40.4, 38.6, 37.0, 31.2, 30.9, 30.7, 30.6, 30.3, 29.3, 28.7, 27.0, 23.2.

4.5.4 Chemical synthesis of *N*-acyl amino acids.

The total chemical synthesis of *N*-acyl amino acids was performed according to the scheme depicted in Figure 40 of the Appendix. Representative ^{1}H NMR spectra for chemically synthesized preparations of *N*-hexadecanoyl-tryptophan (**29**) and *N*-dodecanoyl-phenylalanine (**11**) are provided as Figure 36 and Figure 37 within the Appendix, respectively.

CHAPTER 5

5 *N*-acyl amino acid synthases are linked to the putative PEP-CTERM/exosortase protein-sorting system

5.1 Chapter Summary

Clones that encode the biosynthesis of *N*-acyl amino acids are frequently recovered from activity-based screens of soil metagenomic libraries. Members of a diverse set of enzymes referred to as *N*-acyl amino acid synthases are responsible for the production of all metagenome-derived *N*-acyl amino acids characterized to date. Based on the frequency at which *N*-acyl amino acid synthase genes have been identified from metagenomic samples, related genes are expected to be common throughout the global bacterial metagenome. Homologs of metagenome-derived *N*-acyl amino acid synthase genes are scarce, however, within the sequenced genomes of cultured bacterial species. Towards the goal of understanding the role(s) played by *N*-acyl amino acids in environmental bacteria, a search was performed for conserved genetic features that are positionally-linked to metagenome-derived *N*-acyl amino acid synthase genes. This analysis revealed that *N*-acyl amino acid synthase genes are frequently found

adjacent to genes predicted to encode PEP-CTERM motif-containing proteins and, in some cases, other conserved elements of the PEP-CTERM/exosortase system. The PEP-CTERM/exosortase system is the putative Gram-negative equivalent of the LPXTG/sortase protein-sorting system of Gram-positive bacteria. Several metagenome-derived members of an uncharacterized family of hypothetical acyltransferases previously associated with the PEP-CTERM/exosortase system were then found to possess *N*-acyl amino acid synthase activity, thereby establishing this family as the largest group of homologous *N*-acyl amino acid synthases reported to date. This discovery led to the identification of dozens of previously undetected *N*-acyl amino acid synthase-containing species within the *Proteobacteria* phylum.

5.2 Introduction

Diverse microbial communities are now routinely explored and characterized through metagenomic analysis. Generally, DNA extracted from complex environmental samples (environmental DNA, eDNA) is cloned into standard *E. coli*-based vectors and the resulting eDNA libraries are subjected to various functional and/or sequence-based analyses. Activity-dependent screens of eDNA libraries that target antibacterial-activity have been used to identify eDNA clones that contain genes encoding for small molecule biosynthetic machineries (Brady 2007). To date, the most common small molecules reported from antibiosis screens of *E. coli*-based soil eDNA libraries are long-chain *N*-acyl amino acids. Extracts of antibacterially-active

eDNA clones have yielded several distinct varieties of *N*-acyl amino acids, including *N*-acyl derivatives of tyrosine, phenylalanine, tryptophan and arginine (Figure 25, compounds **22**-**25** respectively) (Brady and Clardy 2000; Brady and Clardy 2005; Clardy and Brady 2007). eDNA clones that produce *N*-acyl amino acids have been identified in metagenomic studies spanning geographically and compositionally distinct soils, suggesting that *N*-acyl amino acid-producing bacteria may be common to the majority of soil ecosystems. Despite their apparent ubiquity, the biological purpose and ultimate fate of *N*-acyl amino acids is unclear.

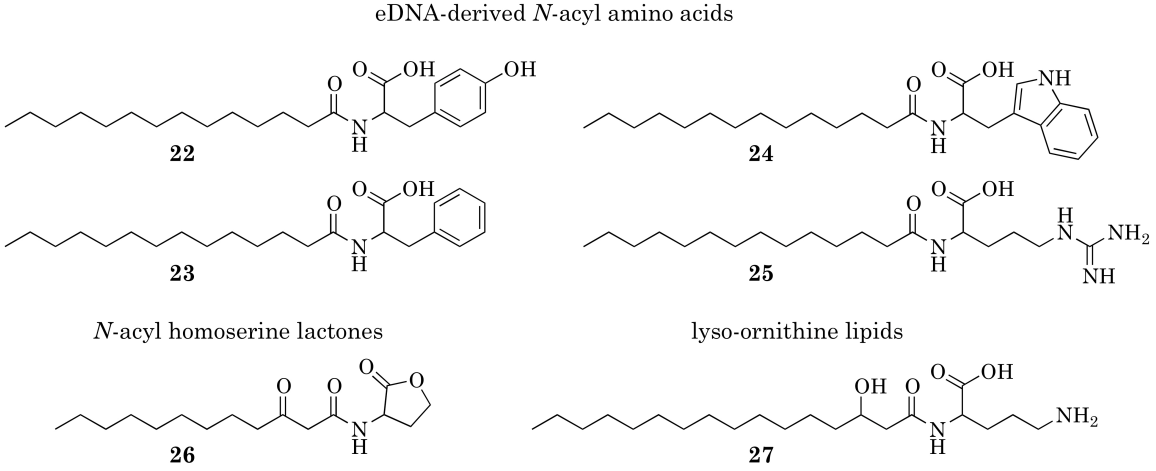


Figure 25: *N*-acyl amino acids and related compounds.
Long-chain *N*-acyl amino acids isolated from antibacterially-active eDNA clones: *N*-tetradecanoyl-tyrosine (**22**), *N*-tetradecanoyl-phenylalanine (**23**), *N*-tetradecanoyl-tryptophan (**24**) and *N*-tetradecanoyl-arginine (**25**). An *N*-acyl homoserine lactone produced by LasI from *P. aeruginosa* PAO1: *N*-(3-oxododecanoyl)-homoserine lactone (**26**). A lyso-ornithine lipid produced by OlsB from *S. meliloti* strain 1021: *N*-(3-hydroxyhexadecanoyl)-ornithine (**27**).

The biosynthesis of *N*-acyl amino acids is carried out by a diverse group of enzymes referred to as *N*-acyl amino acid synthases (NASs), which utilize acyl-(acyl-carrier proteins) (acyl-ACPs) and amino acids as substrates (Brady, Chao et al. 2004). Genes with homology to known eDNA-derived *N*-acyl amino acid synthase genes are scarce within the sequenced genomes of cultured bacteria, and in only one instance has a predicted NAS from a cultured organism been functionally verified (Brady, Chao et al. 2004). Consequently, appreciation of the microorganisms that produce *N*-acyl amino acids and their significance as members of microbial communities is limited. Most of what is currently known about bacterial NASs has been provided by structural and biochemical studies of the eDNA-derived *N*-acyl-tyrosine synthase (NasY) FeeM (Van Wagoner and Clardy 2006). In particular, it has been noted that FeeM shares several important characteristics with *N*-acyl homoserine lactone synthases, including structural resemblance to the GCN5-related *N*-acetyltransferases and the use of acyl-ACP rather than acyl-CoA as an acyl-donor (Brady, Chao et al. 2004; Van Wagoner and Clardy 2006).

The *N*-acyl homoserine lactone products of *N*-acyl homoserine lactone synthases are used as quorum sensing signaling molecules by Gram-negative bacteria (Smith, Wang et al. 2006). The chemical similarities between *N*-acyl-tyrosines and *N*-acyl homoserine lactones are readily apparent, and it has been proposed that *N*-acyl-tyrosines may play a similar role in bacterial

signaling (Figure 25, compounds **22** and **26**) (Brady, Chao et al. 2004). It has also been suggested that some eDNA-derived NASs may represent an expansion of the lipid biosynthetic machinery responsible for the synthesis of bacterial-type ornithine-containing lipids, which are structural lipids similar to phosphatidic acid (Geiger, Gonzalez-Silva et al. 2010). The first step of ornithine-lipid biosynthesis requires NAS-like activity and is performed by the enzyme OlsB, which catalyzes the formation of lyso-ornithine lipid from ornithine and 3-hydroxyacyl-ACP (Figure 25, compound **27**) (Gao, Weissenmayer et al. 2004).

The clustering of functionally related genes is common within bacterial chromosomes (Audit and Ouzounis 2003; Zheng, Anton et al. 2005), and a thorough accounting of the genes surrounding known *NAS* genes may provide valuable information that could be used to identify common features of *NAS*-containing bacteria and to begin to develop an understanding of the biological role(s) performed by *N*-acyl amino acids. Therefore, to study *N*-acyl amino acids and the organisms capable of producing them, the genetic contexts of eDNA-derived *NAS* genes reported from previous metagenomic studies were analyzed. In total, 14 eDNA cosmids containing characterized *NAS* genes were fully sequenced and annotated to identify genetic elements conserved across this family of clones (Table 9). From this analysis, it became apparent that there exists a genetic link between *NAS* genes and genes corresponding to conserved elements of the PEP-CTERM/exosortase system, a putative

protein-sorting system associated with exopolysaccharide (EPS) production in Gram-negative bacteria. This investigation has also indicated that members of a previously uncharacterized group of hypothetical proteins, the putative PEP-CTERM/exosortase system-associated acyltransferases (ExoATs, IPR022484), constitute a novel family of *N*-acyl amino acid synthases that is widely distributed amongst the PEP-CTERM/exosortase system-containing *Proteobacteria*.

Table 9: *NAS*-containing eDNA cosmid clones.

Clone	NAS	Amine substrate of NAS	PEP-CTERM (Number)	ExoAT & PrsKRT	Reference
pCSL12	NasY1	tyrosine	Yes (2) [a]		(Brady and Clardy 2000)
pCSL132	NasY2	tyrosine	Yes (3)	Yes	(Brady, Chao et al. 2004)
pCSL144	NasY3	tyrosine			(Brady, Chao et al. 2004)
pCSLC2	FeeM	tyrosine			(Brady, Chao et al. 2002)
pCSLC3	NasY5	tyrosine	Yes (2)	Yes	(Brady, Chao et al. 2004)
pCSLD10	NasY6	tyrosine	Yes (1) [a]		(Brady, Chao et al. 2004)
pCSLF42	NasY7	tyrosine			(Brady, Chao et al. 2004)
pCSLF43	NasY8	tyrosine	Yes (1) [a]		(Brady, Chao et al. 2004)
pCSLG7	NasY9	tyrosine	Yes (1) [a]		(Brady, Chao et al. 2004)
pCSLG10	NasY10	tyrosine	Yes (1) [a]		(Brady, Chao et al. 2004)
pCSL1	NasW	tryptophan			(Brady and Clardy 2005)
pCSL11	NasR	arginine	Yes (2) [a]	Yes	(Brady and Clardy 2005)
pCSL142	NasP1	phenylalanine			(Clardy and Brady 2007)
pEC5	NasP2	phenylalanine			(Craig, Chang et al. 2010)

a. In these six clones, a gene encoding a putative PEP-CTERM motif protein is located directly adjacent to the previously characterized *NAS* gene.

5.3 Results

5.3.1 Many eDNA-derived *NAS* genes are located near genes encoding putative PEP-CTERM motif-containing proteins

N-acyl-tyrosines are the most frequently encountered subset of *N*-acyl amino acids identified from activity-based screens of soil eDNA libraries. The

corresponding NAS enzymes that produce *N*-acyl-tyrosines, referred to as NasYs, can be divided into three groups based on sequence homology - Group 1 (NasY1-2, FeeM, NasY5 and NasY7); Group 2 (NasY6 and NasY8-10); and "others" (NasY3) (Brady, Chao et al. 2004). There are two known *N*-acyl-phenylalanine-producing NASs, NasP1 (NasP) and NasP2 (NasA), that are non-homologous to one another and lack similarity to the NasYs (Clardy and Brady 2007; Craig, Chang et al. 2010). There are also singular examples of *N*-acyl-tryptophan-producing (NasW) and *N*-acyl-arginine-producing (NasR) NASs, neither of which resembles other previously described NASs (Brady and Clardy 2005). The 25-40 kb eDNA inserts on which these 14 *NAS* genes are found were each sequenced, annotated and compared with one another to identify conserved genetic features.

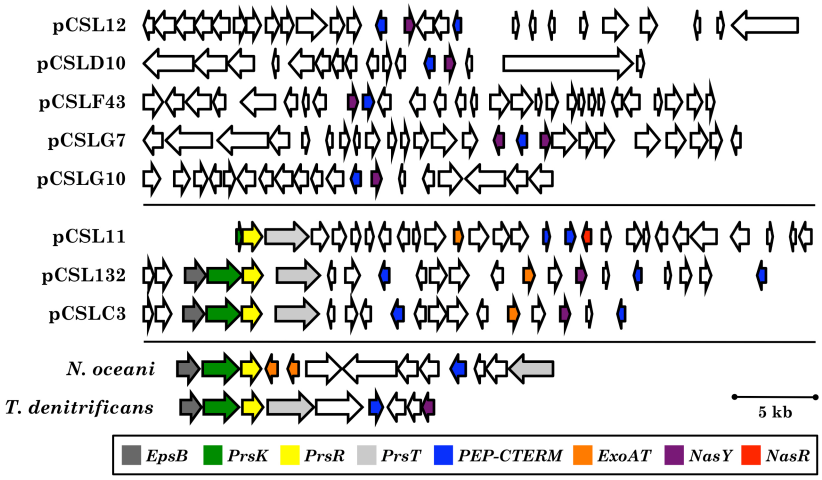
Eight eDNA clones, including seven of the ten *NasY*-containing eDNA clones (Group 1 - 3 of 5; Group 2 - 4 of 4) and the only *NasR*-containing eDNA clone were each found to contain at least one gene encoding a hypothetical protein bearing a PEP-CTERM motif (*VPEP*, Pfam PF07589/IPR011449) (see Table 9 and Figure 26A). The PEP-CTERM motif is a short C-terminal homology domain believed to be analogous to the LPXTG-motif used for protein sorting in Gram-positive bacteria (Schneewind, Model et al. 1992; Schneewind, Mihaylova-Petkov et al. 1993; Mazmanian, Liu et al. 1999; Haft, Paulsen et al. 2006). The N-terminal regions of PEP-CTERM motif proteins are generally predicted to contain signal peptides, but are otherwise variable.

The putative transpeptidase EpsH (exopolysaccharide locus protein H, exosortase, IPR013426) is thought to process all PEP-CTERM motif proteins, most likely for targeting across the outer membrane (Figure 26B) (Haft, Paulsen et al. 2006). In total, 13 genes encoding PEP-CTERM motif proteins were identified from within these eight eDNA clones. The N-terminal regions of these 13 hypothetical proteins are predicted to contain signal peptides, but are not otherwise conserved. Most, however, are acidic proteins of similar size (~200-300 residues) that contain an abundance of residues used for O-linked and N-linked glycosylation (Ser/Thr and Asn, respectively). Within six of these eight clones, at least one gene encoding a putative PEP-CTERM motif protein was found directly adjacent to a previously characterized *NAS* gene.

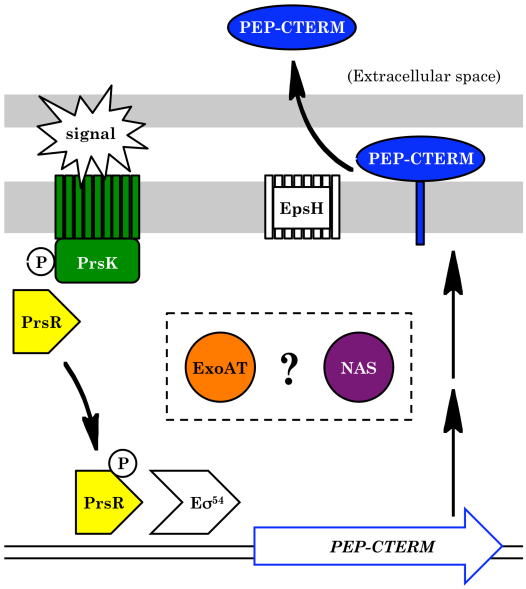
Figure 26: ORF maps of *NAS* and *PEP-CTERM* containing clones.

(A) Open reading frame maps of the eDNA inserts containing previously identified *N*-acyl amino acid synthase genes (either *NasYs* or *NasR*). eDNA insert maps are divided into two sections to separate those that contain whole or partial genes for *PrsT*, *PrsK* and *PrsR* (middle section) from those that do not (top section). For comparison, the bottom section shows similar regions from the genomes of *N. oceani* ATCC 19707™ and *T. denitrificans* ATCC 25259™. (B) In this overview of the proposed PrsK/PrsR-mediated regulation of PEP-CTERM motif protein expression, an unknown signal initiates phosphorelay between PrsK and PrsR. Phosphorylated PrsR then binds to a proposed response-regulator binding motif and activates the alternate σ^{54}-RNAP holoenzyme ($E\sigma^{54}$) in an ATP-dependent process leading to transcription of downstream gene(s) encoding PEP-CTERM motif protein(s). PEP-CTERM motif proteins, anchored to the plasma membrane by their PEP-CTERM motifs, are subsequently cleaved by the transpeptidase EpsH, presumably for targeting across the outer-membrane. The contributions of ExoAT and NAS-like enzymes to this process are currently unknown.

A)


pCSL12
pCSLD10
pCSLF43
pCSLG7
pCSLG10
pCSL11
pCSL132
pCSLC3
N. oceani
T. denitrificans
5 kb
EpsB
PrsK
PrsR
PrsT
PEP-CTERM
ExoAT
NasY
NasR

B)


PEP-CTERM
(Extracellular space)
signal
PEP-CTERM
EpsH
P
PrsK
PrsR
ExoAT
?
NAS
P
PrsR
Eσ54
PEP-CTERM

Bacterial species encoding putative PEP-CTERM motif proteins are found predominantly in sediments, soils and biofilms, possess an outer membrane and invariably contain genes for exopolysaccharide (EPS) production (Haft, Paulsen et al. 2006). In *Proteobacteria*, the genes encoding most PEP-CTERM motif proteins are preceded by intergenic regions that contain a putative *cis*-regulatory site with a σ^{54} binding motif. Haft et al. found that the phylogenetic profile of this presumed regulatory region was identical to that of three conserved protein families; (i) a TPR-repeat protein (PrsT, IPR014266), (ii) a transmembrane histidine kinase (PrsK, IPR014265) and (iii) a σ^{54}-interacting response regulator (PrsR, IPR014264), the latter two of which are believed to regulate the expression of PEP-CTERM motif proteins (Figure 26B). Phylogenetic profile comparisons are used to infer functional relationships between genes and other genetic features through a pattern of co-occurrence across multiple species (Pellegrini, Marcotte et al. 1999). The underlying methodology is premised on the assumption that elements with interrelated functions will be preferentially retained with one another through evolution and lateral transfer. In addition to PrsT, PrsK and PrsR, several other protein families are also associated with the PEP-CTERM/exosortase system, including a family of sugar transferases (EpsB, IPR017464), and an uncharacterized family of hypothetical acyltransferases (PEP-CTERM/exosortase system-associated acyltransferases, IPR022484),

herein referred to as ExoATs, that are members of the acyl-CoA *N*-acyltransferase superfamily (SSF55729) (Figure 26B).

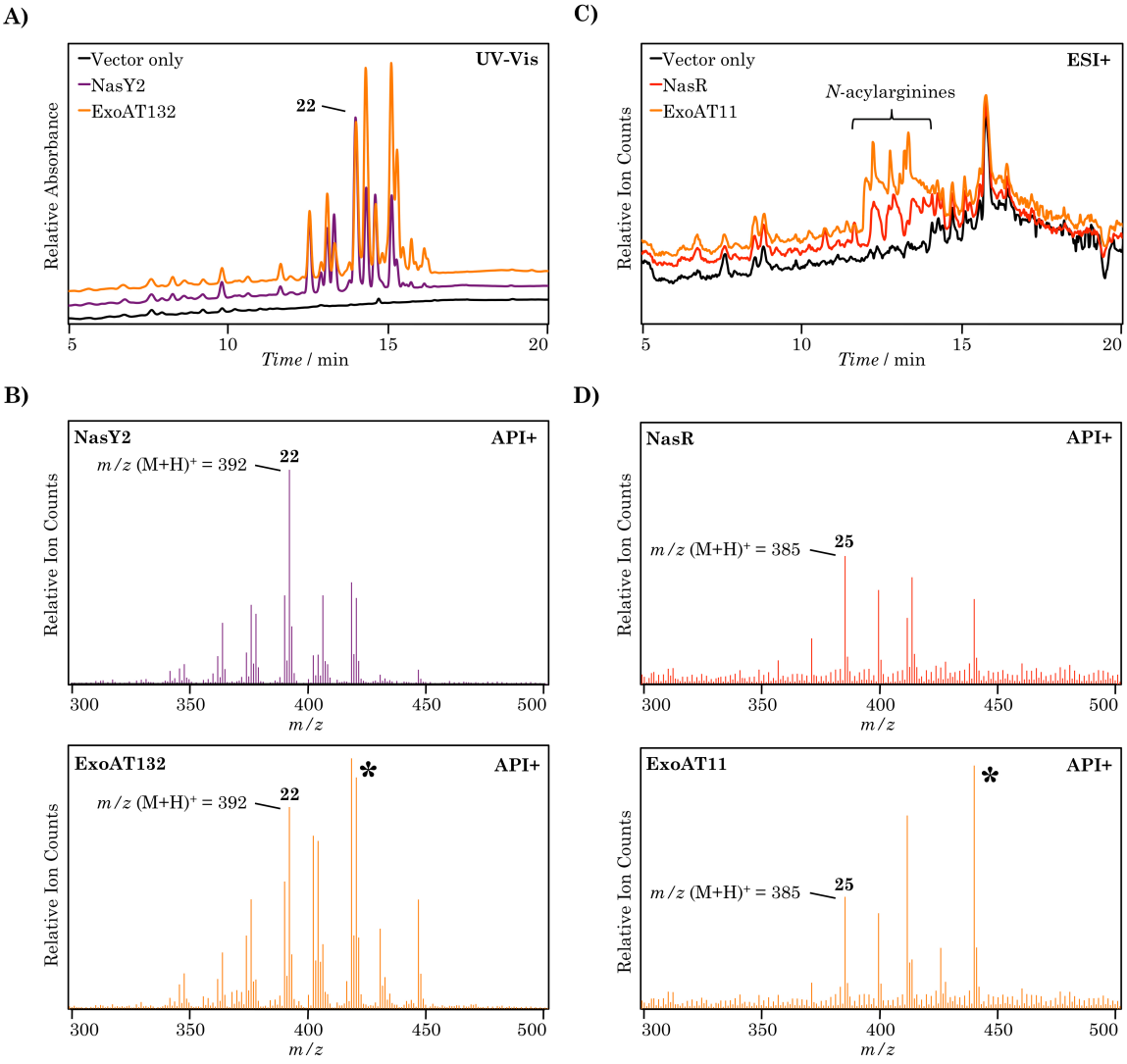
5.3.2 PEP-CTERM/exosortase system-associated acyltransferases (ExoATs) possess *N*-acyl amino acid synthase activity

From within the set of eDNA clones containing gene(s) coding for PEP-CTERM motif proteins, two *NasY*-containing eDNA clones (pCSL132 and pCSLC3) and the *NasR*-containing eDNA clone (pCSL11) were each found to contain an uncharacterized gene encoding for a hypothetical ExoAT (see Table 9 and Figure 26A). As ExoATs were not associated with a specific enzymatic function but appeared to possess biosynthetic potential, the function of these enzymes was investigated by examining their ability to confer the production of organic-extractable small molecules to a heterologous host. Each *ExoAT* gene was cloned into an *E. coli* expression vector downstream of the P_{TAC}-promoter and transformed into *E. coli*. Analysis of oragnic culture broth extracts of all three *ExoAT* expression constructs revealed the presence of long-chain *N*-acyl amino acid mixtures containing compounds identical by mass, retention time and UV-VIS absorption profile to previously reported *N*-acyl amino acids. The two *ExoATs* from the *NasY*-containing clones pCSL132 and pCSLC3 (*ExoAT132* and *ExoATC3*, respectively) enabled *E. coli* to produce *N*-acyl-tyrosine mixtures while the *ExoAT* from the *NasR*-containing clone pCSL11 (*ExoAT11*) afforded the production of *N*-acyl-arginines (Figure 27).

Figure 27: HPLC-MS analyses of ExoAT expression constructs.

(A) UV-Vis chromatograms from the HPLC-MS analysis of organic extracts of *E. coli* heterologous expression constructs containing *NasY2*, *ExoAT132* or empty vector. Prominent signals from *N*-acyl-tyrosines occur between minutes 12-17. (B) API-positive mode ionization data corresponding to minutes 11-18 of the HPLC-MS analysis of *NasY2* (top panel) and *ExoAT132* (bottom panel). The peak with *m/z* $(M+H)^+$ = 392, corresponding to **22** (*N*-tetradecanoyl-tyrosine) is highlighted in both panels. The relatively larger peaks with *m/z* $(M+H)^+$ = 418 and 420 that are marked by a single asterisk in the *ExoAT132* panel correspond to the saturated and monounsaturated 16-carbon *N*-acyl derivatives of tyrosine.

(C) Relative ion-counts (ESI-positive mode) over the course of the HPLC-MS analysis of organic extracts of *E. coli* heterologous expression constructs containing *NasR*, *ExoAT11* or empty vector. (D) API-positive mode ionization data corresponding to minutes 10-18 of the HPLC-MS analysis of *NasR* (top panel) and *ExoAT11* (bottom panel). The peak with *m/z* $(M+H)^+$ = 385, corresponding to **25** (*N*-tetradecanoyl-arginine) is highlighted in both panels. The largest peak in the *ExoAT11* panel with *m/z* $(M+H)^+$ = 439, marked by a single asterisk, corresponds to the monounsaturated 18-carbon *N*-acyl derivative of arginine.


A)
Vector only
NasY2
ExoAT132
UV-Vis
22
Relative Absorbance
5
10
15
20
Time / min
C)
Vector only
NasR
ExoAT11
ESI+
N-acylarginines
Relative Ion Counts
5
10
15
20
Time / min
B)
NasY2
API+
m/z (M+H)+ = 392
22
Relative Ion Counts
300
350
400
450
500
m/z
ExoAT132
API+
m/z (M+H)+ = 392
22
*
Relative Ion Counts
300
350
400
450
500
m/z
D)
NasR
API+
m/z (M+H)+ = 385
25
Relative Ion Counts
300
350
400
450
500
m/z
ExoAT11
API+
m/z (M+H)+ = 385
25
*
Relative Ion Counts
300
350
400
450
500
m/z

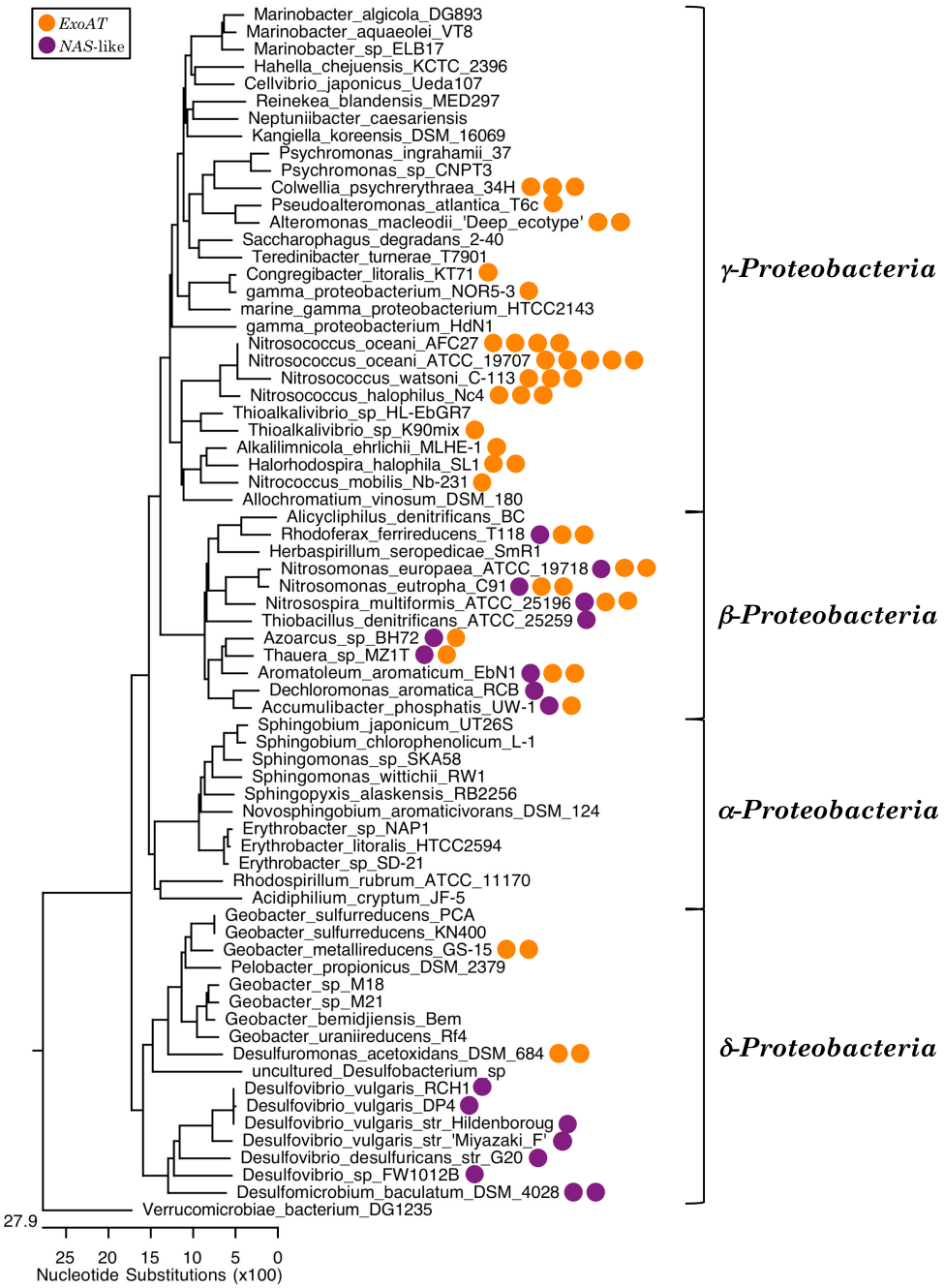
To further investigate the functional relationship between *NAS* and *ExoAT* genes located on the same eDNA clone, *NasY2* from pCSL132 and *NasR* from pCSL11 were heterologously expressed in *E. coli*, and extracts of these cultures were compared directly with extracts of cultures expressing *ExoAT132* and *ExoAT11*, respectively. Organic extracts from cultures expressing *NasY2* and *ExoAT132* both contained complex mixtures of long-chain *N*-acyl-tyrosines, and although the profiles of these mixtures were similar, they were not identical (Figure 27A). For instance, while compound **22** (*N*-tetradecanoyl-tyrosine; C14:0, m/z $(M+H)^+$ = 392) was a major component of both extracts, the *ExoAT132* extract contained relatively greater amounts of the saturated and monounsaturated 16-carbon *N*-acyl derivatives of tyrosine (C16:0, m/z $(M+H)^+$ = 420; C16:1, m/z $(M+H)^+$ = 418) (Figure 27B). Organic extracts from cultures expressing *NasR* and *ExoAT11* both contained complex mixtures of long-chain *N*-acyl-arginines, including compound **25** (*N*-tetradecanoyl-arginine; C14:0, m/z $(M+H)^+$ = 385). The *ExoAT11* extract, however, contained relatively greater amounts of the 16-carbon and 18-carbon monounsaturated *N*-acyl derivatives of arginine (C16:1, m/z $(M+H)^+$ = 411; C18:1, m/z $(M+H)^+$ = 439) (Figure 27C-D). The observed preferences of ExoAT132 and ExoAT11 for slightly longer acyl-substrates than those utilized by NasY2 and NasR were reproducibly observed over multiple trials. The significance of such subtle differences, however, is unclear at this time.

5.3.3 *ExoAT* and *NAS* genes show distinct phylogenetic distribution patterns

Members of the *ExoAT* gene family were first identified through their phylogenetic association with genomes containing the PEP-CTERM/exosortase system. In the case of all three *NAS*-containing eDNA clones from which an *ExoAT* gene conferring the production of *N*-acyl amino acids was identified, whole or partial genes corresponding to the conserved PEP-CTERM/exosortase system proteins PrsT, PrsK and PrsR were also identified (see Table 9 and Figure 26A). To better understand the relationship of the *PrsK/PrsR* two-component system with *ExoATs* and *NAS*-like genes, a search was performed for patterns in the phylogenetic distribution of *ExoATs* and *NAS*-like genes across the 69 species of *Proteobacteria* whose genomes contain both *PrsK* and *PrsR*. Figure 28 illustrates this relationship as a 16S rRNA-based phylogenetic tree to which symbols indicating the presence and number of both putative *ExoAT* and *NAS*-like genes within the genome of each species have been appended.

Figure 28: 16S rRNA phylogram of *PrsKR*$^{+}$ *Proteobacteria*.

16S rRNA phylogram of the 69 species of *Proteobacteria* whose genomes contain *PrsK* and *PrsR* (out of 71 species total). The phylogram was constructed using 16S sequences from the Silva comprehensive ribosomal RNA database "SSU Ref 104", trimmed to the region corresponding to positions 133-1178 of the *E. coli* K-12 MG1655 16S rRNA gene *rrsG*, and aligned using Clustal W. The trimmed 16S rRNA sequence from *Verrucomicrobiae bacterium* DG1235 (a *PrsK*/*PrsR*-containing member of the phylum *Verrumcomicrobia*) was used as an out-group. For each *ExoAT* homolog (IPR022484) found within a genome, an orange dot is appended next to the species name. A purple dot is appended next to the name of each species whose genome contains a putative homolog of the Group 1 *NasYs*.


ExoAT
NAS-like
Marinobacter_algicola_DG893
Marinobacter_aquaeolei_VT8
Marinobacter_sp_ELB17
Hahella_chejuensis_KCTC_2396
Cellvibrio_japonicus_Ueda107
Reinekea_blandensis_MED297
Neptuniibacter_caesariensis
Kangiella_koreensis_DSM_16069
Psychromonas_ingrahamii_37
Psychromonas_sp_CNPT3
Colwellia_psychrerythraea_34H
Pseudoalteromonas_atlantica_T6c
Alteromonas_macleodii_'Deep_ecotype'
Saccharophagus_degradans_2-40
Teredinibacter_turnerae_T7901
Congregibacter_litoralis_KT71
gamma_proteobacterium_NOR5-3
marine_gamma_proteobacterium_HTCC2143
gamma_proteobacterium_HdN1
Nitrosococcus_oceani_AFC27
Nitrosococcus_oceani_ATCC_19707
Nitrosococcus_watsoni_C-113
Nitrosococcus_halophilus_Nc4
Thioalkalivibrio_sp_HL-EbGR7
Thioalkalivibrio_sp_K90mix
Alkalilimnicola_ehrlichii_MLHE-1
Halorhodospira_halophila_SL1
Nitrococcus_mobilis_Nb-231
Allochromatium_vinosum_DSM_180
Alicycliphilus_denitrificans_BC
Rhodoferax_ferrireducens_T118
Herbaspirillum_seropedicae_SmR1
Nitrosomonas_europaea_ATCC_19718
Nitrosomonas_eutropha_C91
Nitrosospira_multiformis_ATCC_25196
Thiobacillus_denitrificans_ATCC_25259
Azoarcus_sp_BH72
Thauera_sp_MZ1T
Aromatoleum_aromaticum_EbN1
Dechloromonas_aromatica_RCB
Accumulibacter_phosphatis_UW-1
Sphingobium_japonicum_UT26S
Sphingobium_chlorophenolicum_L-1
Sphingomonas_sp_SKA58
Sphingomonas_wittichii_RW1
Sphingopyxis_alaskensis_RB2256
Novosphingobium_aromaticivorans_DSM_124
Erythrobacter_sp_NAP1
Erythrobacter_litoralis_HTCC2594
Erythrobacter_sp_SD-21
Rhodospirillum_rubrum_ATCC_11170
Acidiphilium_cryptum_JF-5
Geobacter_sulfurreducens_PCA
Geobacter_sulfurreducens_KN400
Geobacter_metallireducens_GS-15
Pelobacter_propionicus_DSM_2379
Geobacter_sp_M18
Geobacter_sp_M21
Geobacter_bemidjiensis_Bem
Geobacter_uraniireducens_Rf4
Desulfuromonas_acetoxidans_DSM_684
uncultured_Desulfobacterium_sp
Desulfovibrio_vulgaris_RCH1
Desulfovibrio_vulgaris_DP4
Desulfovibrio_vulgaris_str_Hildenboroug
Desulfovibrio_vulgaris_str_'Miyazaki_F'
Desulfovibrio_desulfuricans_str_G20
Desulfovibrio_sp_FW1012B
Desulfomicrobium_baculatum_DSM_4028
Verrucomicrobiae_bacterium_DG1235
γ-Proteobacteria
β-Proteobacteria
α-Proteobacteria
δ-Proteobacteria
27.9
25 20 15 10 5 0
Nucleotide Substitutions (x100)

In total, 32 of the 69 *PrsK*/*PrsR*-containing species of *Proteobacteria* were found to contain at least one *ExoAT* or *NAS*-like gene, and 18 of those 32 species contained at least two such genes, in various combinations. Both *ExoAT* and *NAS*-like genes were seen in *β-Proteobacteria* and *δ-Proteobacteria*. Eight of the 12 *PrsK*/*PrsR*-containing *β-Proteobacteria*, however, contained genes corresponding to both types of enzymes, which was not the case for any of the 17 *PrsK*/*PrsR*-containing *δ-Proteobacteria*. Members of the *γ-Proteobacteria*, the largest group of *PrsK*/*PrsR*-containing *Proteobacteria*, did not contain *NAS*-like genes, but occasionally possessed numerous *ExoAT* genes. Neither type of enzyme was found in the *α-Proteobacteria*. The 45 *ExoAT* homologs represented in Figure 28 constitute 90% of the *ExoAT* genes that have been identified within the sequenced genomes of *Proteobacteria* (50 *ExoAT* sequences are from *Proteobacteria*, out of 53 total). The sequences of all previously described NASs were used as Blastp search queries to identify additional NAS-like homologs, however only searches using query sequences corresponding to Group 1 NasYs contributed to this analysis. In total, these searches returned 18 potential Group 1 NasY homologs, all of which belonged to species of *PrsK*/*PrsR*-containing *Proteobacteria*. Conversely, all four of the Group 2 NasYs showed similarity to just two hypothetical proteins, both of which are predicted NasY homologs present within the genome of *Candidatus Solibacter usitatus* Ellin6076, a member of the phylum *Acidobacteria*.

The promoter regions located immediately upstream of genes encoding PEP-CTERM motif proteins were then examined, revealing that five of seven such promoters from clones harboring Group 1 *NasYs* contained both a putative PrsR response regulator binding motif and a putative σ^{54} binding motif. This was not the case for any of four such promoters found on clones harboring Group 2 *NasYs*. Differences between Group 1 and Group 2 NasYs are likely reflective of the taxonomic division between the *Proteobacteria* and *Acidobacteria*, the latter of which is represented by few sequenced genomes despite being one of the most abundant, diverse and widely distributed bacterial phyla throughout the environment (Quaiser, Ochsenreiter et al. 2003; Ward, Challacombe et al. 2009). Although the PEP-CTERM motif has been found in many divisions of Gram-negative bacteria, the PrsK/PrsR two-component system and the ExoAT-type and Group 1 NasY-type *N*-acyl amino acid synthases have been found almost exclusively in *Proteobacteria*. It is possible that PEP-CTERM/exosortase-containing *Acidobacteria* may regulate PEP-CTERM motif protein expression through factors other than PrsK and PrsR, just as they appear to possess Group 2 NasY-type *N*-acyl amino acid synthases instead of those commonly found in *Proteobacteria*.

5.3.4 Cloning and heterologous expression of the *ExoAT* gene from *P. atlantica* T6c

The newfound awareness of cultured bacterial species containing ExoAT-type *N*-acyl amino acid synthases afforded the opportunity to study *N*-

acyl amino acids in the context of a native *ExoAT*-containing organism. *Pseudoalteromonas atlantica* T6c (ATCC BAA-1087™), a representative of the *PrsK/PrsR*-containing γ-*Proteobacteria* that contains just a single *ExoAT* gene, was chosen for this purpose. *P. atlantica* T6c was the most closely related *ExoAT*-containing species to *E. coli* according to 16S analysis and had demonstrated genetic tractability (Belas, Bartlett et al. 1988). Found in oceans worldwide, strains of *P. atlantica* are considered primary biofilm-forming bacteria that act as pioneer colonizers of solid surfaces through the secretion of exopolysaccharide that serves to concentrate nutrients and act as a growth substrate for more fastidious microorganisms (Corpe 1973). When subcloned into an *E. coli*-based heterologous expression system, the *ExoAT* gene from *P. atlantica* T6c (Patl_1204 *ExoAT*) conferred the production of a mixture of *N*-acyl-tyrosines and *N*-acyl-phenylalanines dominated by 14-carbon and 16-carbon, saturated and monounsaturated *N*-acyl derivatives (Figure 29). Organic extracts of wild-type *P. atlantica* T6c cultures did not contain detectable levels of these metabolites (Figure 29A), suggesting either (i) limited *ExoAT* expression, (ii) rapid hydrolytic turnover of *N*-acyl amino acids, or (iii) the incorporation of *N*-acyl amino acids into components of the cell wall or other non-organic-extractable fractions.

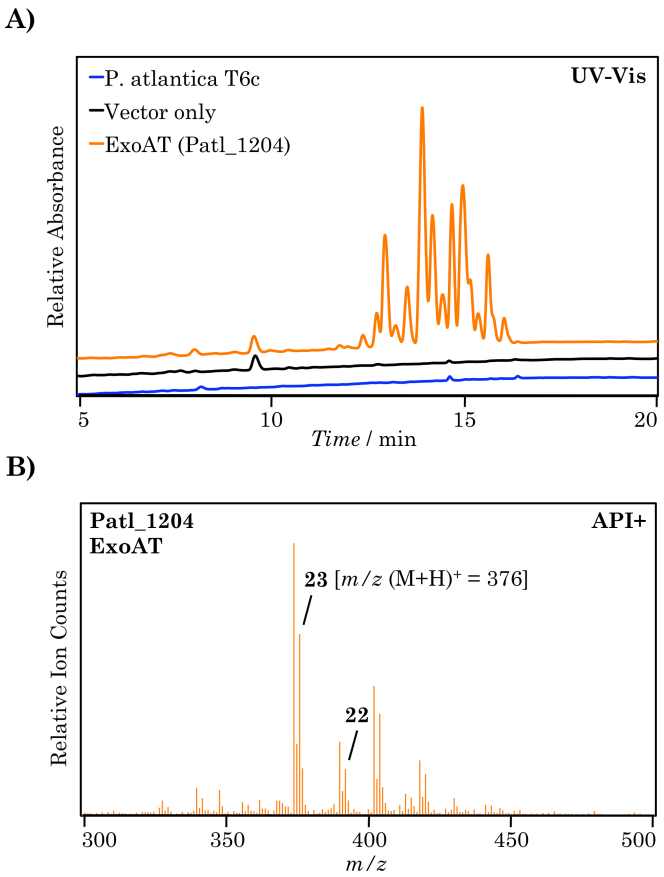


Figure 29: HPLC-MS analysis of Patl_1204 *ExoAT* from *P. atlantica* T6c.
(A) UV-Vis chromatograms from the HPLC-MS analysis of the organic extracts of wild-type *P. atlantica* T6c or the *E. coli* heterologous expression constructs containing the Patl_1204 *ExoAT* gene or empty vector. Prominent signals from *N*-acyl-tyrosines and *N*-acyl-phenylalanines occur between minutes 12-17. (B) API-positive mode ionization data corresponding to minutes 11-18 of the HPLC-MS analysis of the Patl_1204 *ExoAT* expression construct. The peak with m/z $(M+H)^+$ = 376, corresponding to **23** (*N*-tetradecanoyl-phenylalanine) is highlighted, along with the peak corresponding to **22**.

5.3.5 Gene replacement mutagenesis in *P. atlantica* T6c

The close genetic proximity of *NAS* genes to genes encoding PEP-CTERM motif proteins was not limited to NAS-containing eDNA clones. For example, Figure 26A shows similar regions from the genomes of *N. oceani*

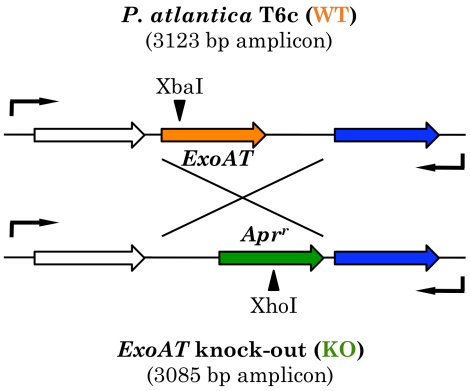
ATCC 19707™ and *T. denitrificans* ATCC 25259™. The lone *ExoAT* gene in the *P. atlantica* T6c genome (Patl_1204) was also found adjacent to a hypothetical PEP-CTERM motif protein (Patl_1203) whose promoter region contained a putative PrsR response regulator binding motif and a putative σ^{54} binding motif. To understand the functional relevance inferred by this genetic relationship, the phenotype of wild-type *P. atlantica* T6c was compared with that of mutant strains that either (i) lack ExoAT-activity (by means of *ExoAT* (Patl_1204) deletion) or (ii) are deficient in the expression of PEP-CTERM motif proteins (by means of *PrsK/PrsR* (Patl_3296/Patl_3295) deletion). Based on previous work by Belas et al., a highly efficient method of gene replacement mutagenesis in *P. atlantica* T6c relying on homologous recombination with a suicide vector was devised (see Section 5.5.7). DNA fragments containing a selectable apramycin resistance marker flanked on each side by sequences homologous to regions of the *P. atlantica* T6c genome targeted for deletion were constructed such that homologous recombination at both homology arms would replace a targeted genomic region with the resistance marker (Thongdee, Gallagher et al. 2008) (Figure 30A). These DNA fragments were then inserted into a mobilizable suicide vector containing the COLA replicon (Zverev and Khmel 1985), which was found to be non-functional in *P. atlantica* T6c. The resulting constructs, pCOLA-ExoAT-KO and pCOLA-PrsKR-KO, were conjugally mated into *P. atlantica* T6c, and double crossover mutants were selected with apramycin (see Section

5.5.6). Using this method, 100% of the resulting apramycin resistant colonies were found to possess the desired genotype (Figure 30B).

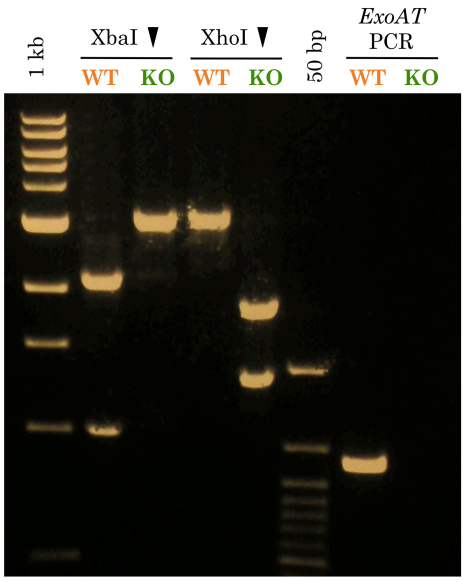
Figure 30: Targeted gene replacement mutagenesis in *P. atlantica* T6c.

(A) Depiction of the targeted homologous recombination event used to produce the *ExoAT* knock-out (KO) strain of *P. atlantica* T6c from wild-type *P. atlantica* T6c (WT). Black bent-arrows represent primers used for PCR amplification of this genomic region. Restriction sites used for PCR amplicon analysis are marked with black wedges. (B) PCR and restriction digest analysis of wild-type *P. atlantica* T6c and the *ExoAT* knock-out strain as depicted in part A. *XbaI* digestion of the 3123 bp *P. atlantica* T6c wild-type amplicon (WT) results in two fragments with lengths of 2117 bp and 1006 bp. *XhoI* digestion of the 3085 bp *ExoAT* knock-out amplicon (KO) results in two fragments with lengths of 1802 bp and 1283 bp. PCR amplification of the 810 bp Patl_1204 *ExoAT* gene (primers 11 and 12; 850 bp amplicon) was successful when using wild-type *P. atlantica* T6c genomic DNA as template but not when using genomic DNA prepared from the *ExoAT* knock-out strain. The lane marked "1 kb" contains 1.0 μg of 1 kb DNA Ladder (NEB®; fragment sizes are 10, 8, 6, 5, 4, 3, 2, 1.5, 1 and 0.5 kb). The lane marked "50 bp" contains 1.0 μg of 50 bp DNA Ladder (NEB®; image truncated - fragment sizes shown are 1350, 916, 766, 700, 650, 600, 550 and 500 bp).

A)


P. atlantica T6c (WT)
(3123 bp amplicon)
XbaI
ExoAT
Apr^r
XhoI
ExoAT knock-out (KO)
(3085 bp amplicon)

B)


1 kb
XbaI
XhoI
50 bp
ExoAT PCR
WT
KO
WT
KO
WT
KO

5.3.6 The expression of *P. atlantica* T6c PEP-CTERM motif proteins is limited under pure culture conditions

Under the most commonly cited conditions used for culturing *P. atlantica* T6c, there is no significant difference between the growth curves of the *ExoAT* or *PrsK/PrsR* knock-out mutants and that of the wild-type strain (Figure 31A). The production of exopolysaccharide, a defining feature of PEP-CTERM/exosortase system-containing species, also appears to be unaffected by *ExoAT* or *PrsK/PrsR* deletion (Figure 31B). As previously noted for cultures of wild-type *P. atlantica* T6c, *N*-acyl amino acids were not detected in the culture broths of either knock-out mutant. To assess the relative expression level of genes under control of the PrsK/PrsR two-component system, GFP-reporter assays were performed, in which the intergenic promoter region corresponding to the region of DNA located immediately upstream of the hypothetical PEP-CTERM motif protein Patl_1203 (P_{VPEP}) was placed upstream of an emerald GFP reporter (Invitrogen™). Cultures of *P. atlantica* T6c transformed with the P_{VPEP}-GFP reporter plasmid yielded fluorescence intensity at the same level as cultures transformed with a non-GFP control plasmid. Positive control cultures transformed with a plasmid containing P_{TAC}-GFP produced a robust fluorescence signal upon induction (Figure 31C). This result suggests that standard laboratory culture conditions may not provide positive stimulation of the PrsK/PrsR two-component system and the subsequent elevation in

expression of proteins containing the PEP-CTERM motif. The biologically relevant factors that serve to activate the PrsK sensor transmembrane histidine kinase are currently unknown. Consequently, additional studies of the orphan PrsK/PrsR two-component system are required to uncover the role PEP-CTERM motif proteins play in the life-cycle of *P. atlantica* T6c and to fully understand the nature of the relationship between PEP-CTERM motif proteins and *N*-acyl amino acid synthases.

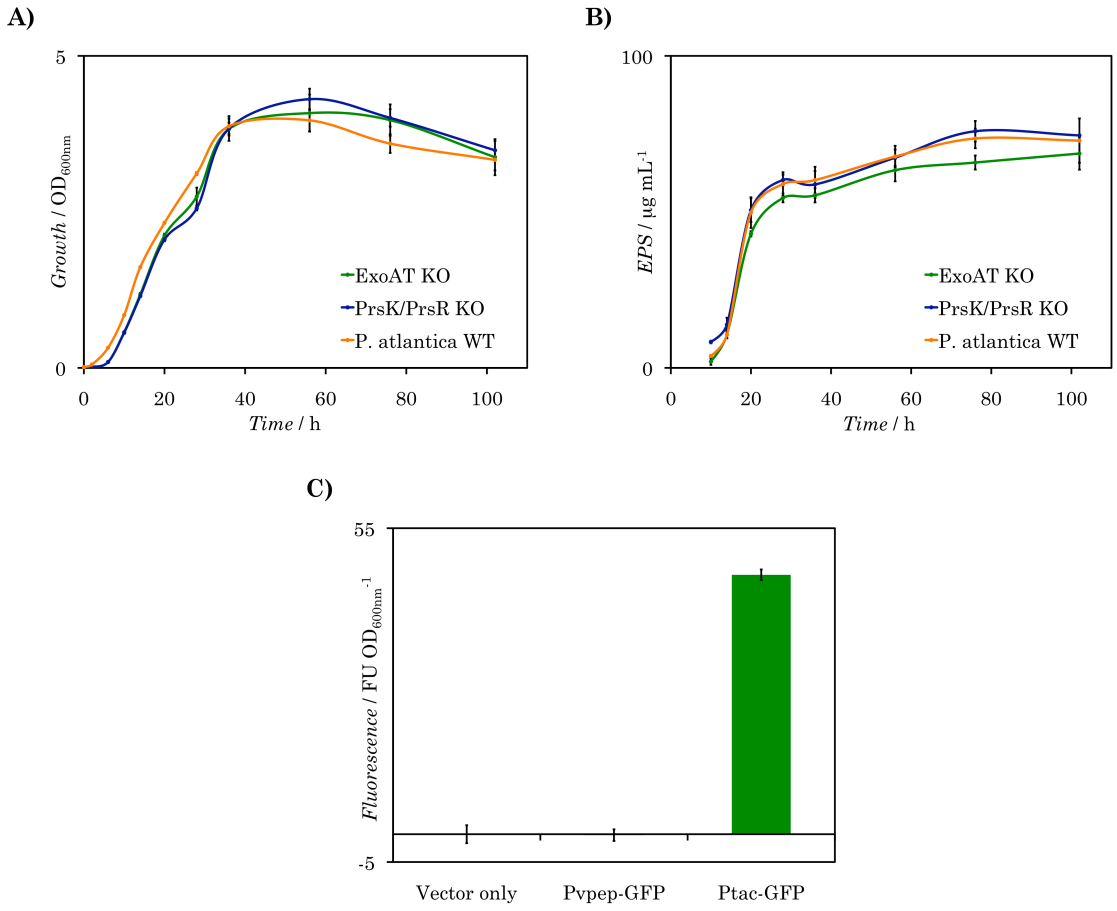


Figure 31: Phenotypic analysis of *ExoAT*- and *PrsKR*- mutant strains of *P. atlantica* T6c.

(A) Growth curves, constructed from OD_{600nm} measurements, and (B) estimates of exopolysaccharide production by wild-type *P. atlantica* T6c, the *ExoAT* knock-out mutant strain and the *PrsK/PrsR* knock-out mutant strain. Error bars represent +/- 1.0 standard deviations of triplicate samples. (C) Relative fluorescence intensities (excitation 487 nm; emission 509 nm) per OD_{600nm} of *P. atlantica* T6c cultures transformed with plasmids containing pEXKm5-P_{VPEP}-GFP (Pvpep-GFP), pEXKm5-P_{TAC}-GFP (Ptac-GFP) or pEXKm5 (Vector only). Error bars represent +/- 1.0 standard deviations of triplicate samples.

5.4 Discussion and Future Directions

Genes with interrelated biological functions are often clustered in bacterial genomes, and this feature can aid in the functional classification of uncharacterized genes and enzymes (Wolf, Rogozin et al. 2001; Rogozin, Makarova et al. 2004). With this in mind, the search for new information pertaining to bacterial *N*-acyl amino acids began with the genetic characterization of *NAS*-containing eDNA clones. Sequence data from 14 unique eDNA clones harboring characterized *NAS* genes was systematically annotated and analyzed to identify conserved genetic elements linked to *NASs*. This analysis indicates that *NAS* genes are commonly found adjacent to genes predicted to encode PEP-CTERM motif-containing proteins (8 of 14 clones examined), an association that was particularly strong amongst both Group 1 and Group 2 *NasY*-containing eDNA clones. This newly discovered link between a known subset of bacterial NASs and PEP-CTERM motif proteins was particularly interesting in light of the association between the PEP-CTERM/exosortase system and a second group of putative biosynthetic enzymes, the PEP-CTERM/exosortase system-associated acyltransferases, which have been referred to as ExoATs.

Three of the NAS-containing eDNA clones were found to contain *ExoAT* genes that conferred the production of long-chain *N*-acyl amino acids to *E. coli*, thus indicating that the ExoAT family of enzymes (IPR022484) is a new class of bacterial *N*-acyl amino acid synthase found frequently in PEP-

CTERM/exosortase system-containing *Proteobacteria*. In the case of all three eDNA clones that encoded a previously characterized NAS and a previously uncharacterized ExoAT, both the NAS and the ExoAT were found to synthesize the same type of *N*-acyl amino acid with similar distributions of acyl-chain lengths and saturations. *NasY2*, a member of the Group 1 *NasYs*, and *ExoAT132*, a bona-fide IPR022484-family member, are located less than 2.5 kb apart from one another on clone pCSL132. These genes show little sequence similarity (Clustal W; 14% translated and aligned identity, 34% similarity), making their apparent functional equivalence all the more striking.

Examination of the taxonomic distributions of *ExoAT* and *NAS*-like genes across the *PrsK/PrsR*-containing *Proteobacteria* highlighted the potential of using *Pseudoalteromonas atlantica* T6c for functional studies of ExoAT and PrsK/PrsR based on single gene knock-out mutagenesis experiments. In studies with mutant strains of *P. atlantica* T6c lacking either *ExoAT* or *PrsK/PrsR*, neither mutant demonstrated an altered growth profile from that of the wild-type parent. Both knock-out mutants were also capable of producing exopolysaccharide at a level similar to that of wild-type *P. atlantica* T6c, and like the wild-type strain, *N*-acyl amino acids were not detected in the organic culture broth extracts of either mutant. The phenotypic equivalence of the *ExoAT* knock-out mutant and the *PrsK/PrsR* knock-out mutant with wild-type *P. atlantica* T6c was in some ways more

understandable upon consideration of the minimal expression of PEP-CTERM motif proteins suggested by GFP reporter-based assays. These results may indicate that the PEP-CTERM/exosortase system, and in particular the PrsK/PrsR two-component system, is minimally active under the culture conditions most commonly used for cultivating *P. atlantica* T6c. Once suitable activating stimuli are identified, however, the strategy for the genetic manipulation of *P. atlantica* T6c described in Section 5.5.7 could easily be extended to address other outstanding issues related to the PEP-CTERM/exosortase system, which has yet to be described in the same detail as the analogous LPXTG/sortase-mediated protein-sorting system of Gram-positive bacteria (Schneewind, Model et al. 1992; Schneewind, Mihaylova-Petkov et al. 1993; Mazmanian, Liu et al. 1999).

Although *N*-acyl amino acid-producing eDNA-clones have been identified by virtue of the antibacterial-activity bestowed upon them through the production of *N*-acyl amino acids, the relatively high concentrations of *N*-acyl amino acids required to exert this effect have generally evoked caution when discussing the biological purpose of these metabolites within an environmental context. Chemical-signaling, on the other hand, has been one of the prevailing hypotheses regarding the more plausible alternative functions of *N*-acyl amino acids. The association between ExoAT and NAS-type biosynthetic enzymes and the orphan PrsK/PrsR two-component system was therefore intriguing. Although not every *PrsK/PrsR*-containing genome

contains an *ExoAT* or *NAS*-like gene, additional yet undiscovered families of *N*-acyl amino acid synthases might exist to account for this discrepancy. There is also evidence that some bacterial species contain genes for small molecule signaling receptors but not for the corresponding biosynthetic enzymes used in the production of cognate small molecules; Presumably such bacteria rely on interspecies communication (Patankar and Gonzalez 2009). It is easy to imagine a scenario where *N*-acyl amino acids are used as ligands for the multi-pass transmembrane sensor domains of the histidine kinase components of two-component signal transduction systems. Based on this type of scenario and the putative roles of PrsK and PrsR, the primary function of *N*-acyl amino acids would be to regulate expression of defined sets of PEP-CTERM motif proteins in a quorum-dependent manner. The work presented in this Chapter (Chapter 5) was intended, in part, to specifically address this possibility. The absence of a defined phenotype associated with either the *ExoAT* or *PrsK/PrsR* knock-out mutant, however, suggests that one or more of these possible signaling factors may be inactive in the context of this experimental design. Under these circumstances it is not possible to provide support for or against the hypothesis that *N*-acyl amino acids act as signaling molecules in *P. atlantica* T6c or other PEP-CTERM/exosortase system-containing organisms.

5.5 Materials and Methods

5.5.1 Sequencing and analysis of eDNA clones

14 cosmid eDNA clones containing previously characterized *NAS* genes were fully sequenced using a combination of 454 pyrosequencing and standard Sanger-type sequencing (Table 9). Sequences from each respective eDNA insert were analyzed for putative open reading frames (ORFs) using GLIMMER v3.02 (Delcher, Harmon et al. 1999) and SoftBerry FGENESB: Bacterial Operon and Gene Prediction Program (Softberry, Inc., Mount Kisco, NY; http://linux1.softberry.com/berry.phtml?topic=fgenesb&group=programs&subgroup=gfindb). Predicted ORFs from each respective eDNA insert were annotated using the entire suite of software applications available through InterProScan (Zdobnov and Apweiler 2001; McDowall and Hunter 2011). These sequence data have been submitted to the GenBank database under Accession Numbers JF429409 (CSL12), JF429410 (CSL132), JF429412 (CSL144), JF429413 (CSLC2), JF429414 (CSLC3), JF429415 (CSLD10), JF429416 (CSLF42), JF429417 (CSLF43), JF429404 (CSLG7), JF429405 (CSLG10), JF429407 (CSL1), JF429408 (CSL11), JF429411 (CSL142) and JF429406 (EC5).

5.5.2 Bacterial Cultures, organic extraction and HPLC analysis

50 mL cultures of *E. coli* transformed with *NAS* or *ExoAT* expression constructs were grown in Luria-Bertani medium containing 200 μg mL^{-1} ampicillin for 72 hours at 30°C. Individual 30 mL aliquots were then extracted 1:1 with ethyl acetate, dried under vacuum, and resuspended in methanol. Each sample was then subjected to analytical HPLC-MS using the following conditions: (Waters XBridge C_{18} 5 mm column [4.6 by 150 mm], 1.5 mL $minute^{-1}$ flow rate); 3 minutes at 50:50 H_2O:methanol with 0.1% formic acid, followed by a linear gradient from 50:50 H_2O:methanol with 0.1% formic acid to 100% methanol with 0.1% formic acid over 12 minutes, followed by 100% methanol with 0.1% formic acid for 5 minutes.

5.5.3 GFP-reporter assays

Overnight cultures of *P. atlantica* T6c were grown at RT to an OD_{600nm} = ~2.0 and then diluted with fresh medium (Difco™ Marine Broth 2216) to an OD_{600nm} = 1.0. GFP fluorescence measurements were collected in a 1.0 cm quartz cuvette at 509 nm using an excitation wavelength of 487 nm.

5.5.4 Construction of *E. coli*-based heterologous expression constructs

Cosmids pCSL11, pCSL132 and pCSLC3 were used as PCR templates for the amplification of *NAS* and *ExoAT* genes. Primer names, sequences and restriction sites are located in Table 10. PCR reactions were carried out under the following conditions: 100 μL reactions, 1.25x Thermopol Buffer

(NEB®), 500 μM dNTP mix, primers at 0.4 μM each, 100 ng template DNA, 0.5 μL per reaction of both Phusion Hot Start DNA Polymerase and Dynazyme II DNA Polymerase (NEB® - Finnzymes), 1 cycle at 95°C for 120 seconds, followed by 31 cycles of 30 seconds at 97°C/ 30 seconds at 64°C/ 50 seconds at 72°C, followed by 120 seconds at 72°C. Similarly, 250 ng of *Pseudoalteromonas atlantica* T6c genomic DNA (ATCC BAA-1087D-5™) and primers 11 and 12 were used for amplification of the Patl_1204 *ExoAT* gene. All PCR amplicons were subsequently digested with either NdeI/PstI or NcoI/SbfI in preparation for cloning into the *E. coli* expression vector p4R-P_{TAC}, prepared by reciprocal enzymatic digestion and CIP-treatment (NEB®) dephosphorylation of p4R-P_{TAC}-GFP (see Table 11). Insert and vector samples were ligated (Fast-Link™ DNA Ligation kit, Epicentre®), transformed into Transformax™ Electrocompetent *E. coli* EC100™ (Epicentre®) and selected on Luria-Bertani agar (LB, LB-agar) containing 200 μg μL^{-1} ampicillin. See Table 11 for complete descriptions of all plasmid constructs.

Table 10: PCR primers discussed in Chapter 5.

#	Primer name	Sequence (restriction sites are highlighted in bold font)
1	CSL132-ExoAT-NdeI-F	GATCTA**CATATG**AAAAGCCCGCATCGACTTTGCGTC
2	CSL132-ExoAT-PstI-R	GATCTT**CTGCAG**TTTTTATTTAGTCACACCGGTTTGCAGCACGCTGTCC
3	CSLC3-ExoAT-NcoI-F	GATC**CCATGG**AAAGCCCGCATCGACTTTGCGTCGCCAAGAG
4	CSLC3-ExoAT-SbfI-R	CTAG**CCTGCAGG**GCTCAAAGTTCCTCGCTTTTTATTTAGTCACG
5	CSL11-ExoAT-NdeI-F	GATCTA**CATATG**ATTGTGCCCGATAAGCCGCCACAG
6	CSL11-ExoAT-PstI-R	GATCTA**CTGCAG**TCACGAGCCGCGAAACCACGTCAGCAG
7	CSL132-NasY2-NdeI-F	GATCTT**CATATG**TCTCTACCTGCTTACCACTCGAATCC
8	CSL132-NasY2-PstI-R	GATCTA**CTGCAG**CCAATCAAGACAAAGCAATAGCAAACCTTGTTC
9	CSL11-NasR-NdeI-F	GATCAT**CATATG**CAGCCAGAGATCTTCGCGCTTCGTTATG
10	CSL11-NasR-PstI-R	GATCTT**CTGCAG**TCTCAGTCGCTCACATCTCGCCGCGGAAC
11	Patl_1204-ExoAT-NcoI-F	GAACAA**CCATGG**AACAACTTCGAAAGTACGCTGATCTCC
12	Patl_1204-ExoAT-HindIII-R	GATCTT**AAGCTT**GCAAGCCCAAGTTTTGAGCTTCAGTTG
13	Patl-ExoAT-2F	CATATTGGCGTGTTCGATGCTTTACAGAAG
14	Patl-ExoAT-2R	AGATTCATTTCTTTAAAACGGTAGTCCACATC
15	Patl-ExoAT-3F	ATGTCAAACTTAGTTAAGAAAACAAAATTAG
16	Patl-ExoAT-3R	GAGTAATTAACTCTTATTTTTTAAATCTTCTAC
17	ExoAT-2Rrc-Apr^{r}-1F[a]	GGAACTAGATGTGGACTACCGTTTTAAAGAAATGAATCT*TTGGATACAC CAAGGAAAGTC*
18	ExoAT-3Frc-Apr^{r}-1R[a]	GGCTAGTGTTACTAATTTTGTTTTCTTAACTAAGTTTGACAT*AACGTCAT CTCGTTCTCC*
19	Patl-ExoAT-4F	CTTAATAGGTCCGGCGAAGGATAAGCAAGG
20	Patl-ExoAT-5R	CTACACAACATCAGACCTGCGAATGTAAATGG
21	Patl_3296-PrsK-NcoI-F	GATGAT**CCATGG**TAGCAGTTTGGGGATATGCAGCCAGTGCTTTG
22	Patl_3295-PrsR-BamHI-R	GATCTT**GGATCC**TTAGGAACCAGTTTTAATGTCCTCTAGATGGTAC
23	Patl-PrsK-2F	GCTGAAGATATCGACGTAGTGATTGTATCTTGTG
24	Patl-PrsK-2R	AATGGTGTTGTCCCTGACGTGATTGCTATTGTG
25	Patl-PrsR-3F	TGTGAACAACGTGTGTGAAGGGTTAC
26	Patl-PrsR-3R	CAATGGACTTGTAATCGACTTGGTCAGCAAG
27	PrsK-2Rrc-Apr^{r}-1F[a]	GAAAAACACAATAGCAATCACGTCAGGGACAACACCATT*TTGGATACAC CAAGGAAAGTC*
28	PrsR-3Frc-Apr^{r}-1R[a]	CTTAGCCGAACGTCGGGTAACCCTTCACACACGTTGTTCACA*AACGTCA TCTCGTTCTCC*
29	Patl-PrsKR-4F	GTGAAGAAGAATTATCTGATCAGGCCGAAATGG
30	Patl-ExoAT-5R	GAATACCGCCAATATGTTGACTGGTCTCAATCG
31	Patl_1203-P_{VPEP}-XbaI-F	GATCAC**TCTAGA**AGCTCAAAACTTGGGCTTGCCGGCTG
32	Patl_1203-P_{VPEP}-NcoI-R[b]	GATCAC**CCATGG**AAACCTCCTTTTAGACCAAAAGATGAGACCAAAG

a. The regions of these primers corresponding to sequences within cosmid pOJ436 are italicized.

b. The seven bases closest to the Patl_1203 start codon (AAGGA*ATATCAC***ATG**) were not conserved in order to introduce an optimal ribosomal binding site. The corresponding sequence from the PCR product of primers 31/32 reads, from 5' to 3', AAGGA*GGTTTCC***ATG**.

Table 11: Strains, clones and plasmids discussed in Chapter 5.

Item	Description	Source or reference
Strain		
E. coli EC100	*γ-Proteobacteria* host, cloning strain	Epicentre® Biotechnologies
P. atlantica T6c	Marine *γ-Proteobacteria*; *ExoAT*$^+$, *PrsKR*$^+$	(Belas, Bartlett et al. 1988), ATCC BAA-1087™
Clone or plasmid		
p4R-P_{TAC}-GFP	Derivative of a custom DNA synthesis construct (Epoch Biolabs, Inc., Sugarland, Texas) containing a codon-optimized version of the *emGFP* gene found in pRSET/EmGFP (V353-20, Invitrogen™) cloned downstream of the IPTG-inducible P_{tac} promoter and *E. coli* RBS found in the pTAC-MAT-Tag-2 expression vector (Sigma-Aldrich E5405) and sandwiched between the transcriptional terminator regions present in the CloneSmart series of plasmids (Lucigen®, e.g. pSMART-LCKan); *E. coli lacI* repressor gene; Amp^r	(Craig, Cherry et al. 2011)
p4R-P_{TAC}-ExoAT132	p4RpTAC-based expression construct containing the PEP/exosortase associated acyltransferase gene from cosmid pCSL132 (*ExoAT132*); Amp^r	(Craig, Cherry et al. 2011)
p4R-P_{TAC}-ExoATC3	p4RpTAC-based expression construct containing the PEP/exosortase associated acyltransferase gene from cosmid pCSLC3 (*ExoATC3*), Amp^r	(Craig, Cherry et al. 2011)
p4R-P_{TAC}-ExoAT11	p4RpTAC-based expression construct containing the PEP/exosortase associated acyltransferase gene from cosmid pCSL11 (*ExoAT11*); Amp^r	(Craig, Cherry et al. 2011)
p4R-P_{TAC}-NasY2	p4RpTAC-based expression construct containing the *NasY2* gene from pCSL132; Amp^r	(Craig, Cherry et al. 2011)
p4R-P_{TAC}-NasR	p4RpTAC-based expression construct containing the *NasR* gene from pCSL11; Amp^r	(Craig, Cherry et al. 2011)
p4R-P_{TAC}-Patl-ExoAT	p4RpTAC-based expression construct containing the PEP/exosortase associated acyltransferase gene from the genome of *Pseudoalteromonas atlantica* T6c (Patl_1204 *ExoAT*); Amp^r	(Craig, Cherry et al. 2011)
pEXKm5	Conjugative vector (RK2 *oriT*) used for allele replacement in *B. pseudomallei*; *pMB9* origin of replication (*ColE1*); *E. coli gusA* reporter gene; *sacB*; Kan^r	(Lopez, Rholl et al. 2009)
pOJ436	Conjugative cosmid (RK2 *oriT*) vector for the transfer of DNA from *E. coli* to *Streptomyces* spp; *pMB1* origin of replication (*ColE1*); *aac(3)IV* gene conferring apramycin resistance; Apr^r	(Bierman, Logan et al. 1992)
pCOLADuet-1	*E. coli*-based expression vector utilizing the COLA replicon from pColA (Zverev and Khmel 1985); Kan^r	Novagen®
pCOLA-ExoAT-KO	Conjugative suicide vector for replacement of the PEP/exosortase associated acyltransferase gene of *Pseudoalteromonas atlantica* T6c (Patl_1204 *ExoAT*) with the apramycin resistance cassette from pOJ436 (Bierman, Logan et al. 1992); COLA replicon derived from pCOLADuet-1; RK2 *oriT* and *gusA* from pEXKm5; *aac(3)IV* gene (Apr^r, pOJ436) flanked by DNA regions homologous to those flanking the Patl_1204 *ExoAT* gene; Apr^r and Kan^r	(Craig, Cherry et al. 2011)
pCOLA-PrsKR-KO	Conjugative suicide vector for replacement of the *PrsK* and *PrsR* genes of *Pseudoalteromonas atlantica* T6c (Patl_3296 and Patl_3295, respectively) with the apramycin resistance cassette from pOJ436 (Bierman, Logan et al. 1992); COLA replicon derived from pCOLADuet-1; RK2 *oriT* and *gusA* from pEXKm5; *aac(3)IV* gene (Apr^r, pOJ436) flanked by DNA regions homologous to those flanking the *PrsKR* operon; Apr^r and Kan^r	(Craig, Cherry et al. 2011)
pEXKm5-P_{TAC}-GFP	pEXKm5 containing a codon-optimized version of the *emGFP* gene found in pRSET/EmGFP (V353-20, Invitrogen™) cloned under the control of the P_{TAC} promoter; Kan^r	(Craig, Cherry et al. 2011)
pEXKm5-P_{VPEP}-GFP	pEXKm5 containing a codon-optimized version of the *emGFP* gene found in pRSET/EmGFP (V353-20, Invitrogen™) cloned under the control of the 376 bp intergenic promoter region containing conserved DNA sequence motifs (I and II) corresponding to putative PrsK/PrsR/sigma-54-controlled cis-regulatory elements (Haft, Paulsen et al. 2006) and located directly upstream of the *Pseudoalteromonas atlantica* T6c gene Patl_1203, encoding a putative PEP-CTERM motif-containing protein; Kan^r	(Craig, Cherry et al. 2011)

5.5.5 Construction of GFP-reporter plasmids for *P. atlantica* T6c

P. atlantica T6c genomic DNA and primers 31 and 32 (see Table 10) were used for PCR amplification of the intergenic promoter region corresponding to the nucleotides from position -376 to -8 upstream of the hypothetical PEP-CTERM motif protein Patl_1203 (P_{VPEP}). PCR conditions were the same as above except the annealing temperature was 60°C and the extension time was 30 seconds. The P_{VPEP} PCR amplicon was XbaI/NcoI-digested, then ligated to reciprocally-digested and CIP-treated p4R-P_{TAC}-GFP. This ligation reaction was transformed into *E. coli* EC100™ and selected on LB-agar containing 200 μg mL^{-1} ampicillin. Successful ligation transformants were identified by colony PCR and restriction mapping. The resulting construct, p4R-P_{VPEP}-GFP, and p4R-P_{TAC}-GFP were XhoI-digested, blunt-end repaired (End-It™ DNA End-Repair Kit, Epicentre®) and ligated into SmaI-digested and CIP-treated pEXKm5 vector. These ligation reactions were transformed into *E. coli* EC100™ and selected on LB-agar containing 50 μg mL^{-1} kanamycin. One clone corresponding to pEXKm5-P_{VPEP}-GFP, one to pEXKm5-P_{TAC}-GFP and one to unmodified pEXKm5 were then transformed into electrocompetent donor *E. coli* S17.1 (ATCC 47055™ (Simon, Priefer et al. 1983), prepared according to Sharma and Schimke (Sharma and Schimke 1996)) and introduced into *Pseudoalteromonas atlantica* T6c (ATCC BAA-1087™) by conjugal mating.

5.5.6 Conjugal transfer of plasmid DNA into *P. atlantica* T6c

A 20 mL culture of donor *E. coli* S17.1 was grown overnight at 37°C in LB medium containing 25 μg mL^{-1} spectinomycin and 15 μg mL^{-1} trimethoprim along with plasmid-specific antibiotic(s); 50 μg mL^{-1} kanamycin and/or 100 μg mL^{-1} apramycin were routinely used as selections for both *E. coli* and *P. atlantica* T6c. A 50 mL overnight culture of *P. atlantica* T6c was simultaneously grown in marine broth (Difco™ Marine Broth 2216) containing 20 μg mL^{-1} tetracycline for 14-24 hours at RT or 30°C to an OD_{600nm} = 2.0-2.5. A 1 mL aliquot of the *E. coli* donor, diluted to an OD_{600nm} = 1.0, was harvested by centrifugation (2000 rcf, 2.5 minutes) in a microcentrifuge tube. The *E. coli* pellet was washed once with fresh LB and centrifuged a second time, after which the supernatant was removed by pipette. An 8 mL aliquot of *P. atlantica* T6c was harvested by centrifugation (4000 rcf, 15 minutes) in a 13 mL conical tube, after which the supernatant was carefully removed by decantation. 200 μL of fresh marine broth was then added to the *P. atlantica* T6c pellet, which was vortexed to achieve complete resuspension. The resuspended *P. atlantica* T6c pellet was transferred to a microcentrifuge tube containing the *E. coli* pellet, and the donor/recipient sample was then mixed by vortex. The entire donor/recipient mixture was plated directly onto a 100 mm marine-agar double-selection plate containing 60 μg mL^{-1} tetracycline and plasmid-specific antibiotic(s).

Resistant *P. atlantica* T6c colonies were detectable after 2-3 days and appeared fully mature within 6 days.

5.5.7 Optimization of gene replacement mutagenesis in *P. atlantica* T6c

This section describes the development of a highly efficient method for gene replacement mutagenesis in *P. atlantica* T6c, which was subsequently used to construct mutant strains lacking either *ExoAT* (Patl_1204) or *PrsK/PrsR* (Patl_3296/Patl_3295). Belas et al. previously reported the construction of a *P. atlantica* mutant defective in agarase production by using a gene replacement strategy based on homologous recombination with a suicide vector (Belas, Bartlett et al. 1988). This strategy was reported to be rather inefficient, however, most likely due to its reliance on natural *pMB1*-type (*ColE1*) plasmid segregation. While *ColE1*-type replicons are widely considered to be *E. coli* specific, they are also maintained at low copy number in several additional species (Engleberg, Cianciotto et al. 1988; Keenleyside and Whitfield 1995; Mikiewicz, Wrobel et al. 1997). Plasmid minipreps of *P. atlantica* T6c transformed with the *pMB1*-replicon (*ColE1*-type) plasmid pOJ436 (Bierman, Logan et al. 1992) indicated that this vector was stably maintained under selection in a self-replicative state. Optimally efficient selection of suicide vector-based homologous recombination events requires the use of an origin of replication that is non-functional in the recipient strain. It was therefore necessary to find an alternative to the *pMB1*-type

(*ColE1*) origin of replication that operates well in *E. coli* but not in *P. atlantica* T6c. Nine plasmids, each containing a unique origin of replication known to function in *E. coli*, were linearized with blunt-end producing restriction enzymes and ligated to the DraI-fragment of pOJ436 containing the RK2 *oriT* and *aac(3)IV* gene encoding apramycin resistance (Apr^r) (Table 12). Each pOJ436-derivative plasmid was then conjugally mated into *P. atlantica* T6c (using either *E. coli* S17.1 (ATCC 47055™) (Simon, Priefer et al. 1983) or ET12567 [pUZ8002] *E. coli* (MacNeil, Gewain et al. 1992; Paget, Chamberlin et al. 1999) as donor strains) and selected for on marine-agar containing 100 μg mL^{-1} apramycin and 60 μg mL^{-1} tetracycline. The pOJ436-derivative of pCOLADuet-1 (Novagen®), containing the *COLA* replicon (Zverev and Khmel 1985), was the only plasmid that failed to produce colonies of *P. atlantica* T6c exconjugants. This result implied that *P. atlantica* T6c gene replacement mutants would be most efficiently identified through the use of COLA replicon-based suicide vectors.

Table 12: Plasmid replicons retrofitted with RK2 *oriT* and *aac(3)IV*.

Plasmid	Replicon (source)	Selection	Linear Site	Source/Reference
pUC19	ColE1 (pBR322) point mutant	Amp	SmaI	New England Biolabs® Inc.
pACYCDuet-1	P15A (pACYC184)	Cam	EcoRV	Novagen®
pCDFDuet-1	CloDF13	Spec	EcoRV	Novagen®
pRSFDuet-1	RSF1030	Kan	EcoRV	Novagen®
pCOLADuet-1	COLA (pColA)	Kan	HpaI	Novagen®
pCC1BAC	F-factor (pBeloBAC11)	Cam	HpaI	Epicentre® Biotechnologies
pBBR1MCS-5	pBBR1	Gent	SmaI	(Kovach, Elzer et al. 1995)
pJWC1	RK2 (IncP1-a)	Tet	ScaI	(Craig, Chang et al. 2010)
pMMB67EH	RSF1010 (IncQ)	Amp	SmaI	ATCC 37622™

5.5.8 Vector construction for gene replacement mutagenesis in *P. atlantica* T6c

For construction of the suicide vectors required to create *ExoAT* and *PrsK/PrsR* knock-out mutants of *P. atlantica* T6c, DNA fragments carrying the selectable apramycin resistance marker from pOJ436 flanked on each side by sequences homologous to those flanking the regions of the *P. atlantica* T6c genome targeted for deletion were constructed according to the two-step PCR protocol described by Thongdee et al. (Thongdee, Gallagher et al. 2008), such that homologous recombination at both homology arms replaces a targeted genomic region with the apramycin resistance marker. The following provides a full description of the construction of pCOLA-ExoAT-KO. The construction of pCOLA-PrsKR-KO was identical except for the use of alternate primers (primers 21-30 substitute for primers 11-20, respectively, see Table 10 for primer details). The DNA regions flanking each side of the Patl_1204 *ExoAT* gene were PCR amplified using primers 13 and 14 (820 bp PCR product) and primers 15 and 16 (850 bp PCR product), respectively. PCR reactions were carried out under the following conditions: 100 μL reactions, 1.25x Thermopol Buffer (NEB®), 500 μM dNTP mix, primers at 0.4 μM each, 250 ng of *P. atlantica* T6c genomic DNA as template (ATCC BAA-1087D-5™), 0.5 μL per reaction of both Phusion Hot Start II DNA Polymerase and Dynazyme II DNA Polymerase (NEB® - Finnzymes), 1 cycle at 95°C for 120 seconds, followed by 36 cycles of 30 seconds at 97°C/ 30

seconds at 57°C/ 60 seconds at 72°C, followed by 120 seconds at 72°C. The apramycin resistance cassette from pOJ436 flanked on each side by ~40 bp of end sequence identical to the adjacent region fragments flanking the Patl_1204 *ExoAT* gene was PCR amplified using primers 17 and 18. PCR conditions were the same as above, except 50 ng of pOJ436 cosmid DNA was used as template and the extension time was increased to 75 seconds per cycle. PCR products were analyzed by 1.4% agarose gel electrophoresis and purified using QIAquick® PCR Purification Kits (Qiagen®). A second-round PCR reaction containing 5 ng of each of the three PCR products described above (as template DNA) was then used to create a 2818 bp *ExoAT* knock-out cassette (for producing an *ExoAT*::*Apr*^{r} genotype). Second-round PCR conditions were the same as above, except primers 13 and 16 were included at 0.5 μM each and the extension time was increased to 180 seconds per cycle. The *ExoAT* knock-out cassette was then analyzed by 1.0% agarose gel electrophoresis, excised from the gel with a sterile razorblade and extracted using a QIAquick® Gel Extraction Kit (Qiagen®).

The gel-purified *ExoAT* knock-out cassette was then blunt-end repaired (End-It™ DNA End-Repair Kit, Epicentre®) and ligated into the vector pEXKm5 (Fast-Link™ DNA ligation kit, Epicentre®) (Lopez, Rholl et al. 2009), which had previously been SmaI-digested and CIP-treated (NEB®). This ligation reaction was then transformed directly into Transformax™ Electrocompetent *E. coli* EC100™ (Epicentre®) and plated onto LB-agar

containing 50 μg mL^{-1} kanamycin and 100 μg mL^{-1} apramycin. Ligation transformants were analyzed by restriction digestion (EcoRV and EcoRI) to verify the presence of both an intact insert and vector and to determine the relative orientations of each. One clone, corresponding to pEXKm5 with the *ExoAT* knock-out cassette and having an orientation such that the *aac(3)IV* apramycin resistance gene was on the opposite strand as the *nptII* kanamycin resistance gene, was then digested with *XbaI* and *EcoRI* (NEB®) to excise the *pMB9* (*ColE1*) origin of replication, blunt-end repaired, dephosphorylated and purified using a QIAquick® PCR Purification Kit (Qiagen®). This purified sample was then ligated into the pCOLADuet-1 vector, which had previously been digested with *HpaI* (NEB®). Ligation reactions were transformed into electrocompetent donor *E. coli* S17.1 prepared according to Sharma and Schimke (Sharma and Schimke 1996) and plated onto LB-agar containing 25 μg mL^{-1} kanamycin and 50 μg mL^{-1} apramycin. Ligation transformants were analyzed by restriction digestion (EcoRV) to verify the presence of both an intact insert and vector and to determine the relative orientations of each. One clone, corresponding to pCOLA-ExoAT-KO and having an orientation such that the kanamycin resistance gene from the pCOLADuet-1 vector was on the opposite strand as the pEXKm5 kanamycin resistance gene, was subsequently transferred to *P. atlantica* T6c by conjugal mating.

5.5.9 Analysis of candidate gene replacement mutants

P. atlantica T6c genomic DNA samples were prepared using a modified protocol involving components of QIAprep® Spin Miniprep Kits (Qiagen®). Briefly, cell pellets were resuspended in 400 μL of P1 resuspension buffer, and then mixed with 400 μL of P2 lysis buffer. Lysis reactions were heated at 50°C for 5 minutes using a covered heating block, then cooled to room temperature and vortexed vigorously for 20 seconds. 600 μL of N3 neutralization buffer was then added to each lysis reaction, which were subsequently mixed by inversion. Each sample was then centrifuged (>20,000 rcf, 20 minutes), after which supernatants were applied to QIAprep® Spin Columns, bound, washed and eluted according to standard Qiagen® protocols. Purified genomic DNA preparations from candidate mutants were used as template for two separate PCR reactions designed for the verification of recombinant identity. The first PCR reaction was used to confirm the presence of the appropriate insertion (primers 19 and 20) and the second to confirm the loss of the Patl_1204 *ExoAT* gene targeted for deletion (primers 11 and 12). PCR conditions were the same as above, except 500 ng of genomic DNA was used as template, 63°C was used as the annealing temperature and the extension time was increased to 240 seconds per cycle. PCR reactions were subsequently purified using QIAquick® PCR Purification Kits (Qiagen®). PCR products corresponding to the ~3 kb amplicons produced by primers 19 and 20 were digested separately with *XbaI* (site

internal to the Patl_1204 *ExoAT* gene) and *XhoI* (site internal to the *aac(3)IV* apramycin resistance gene) then analyzed by 1.0% agarose gel electrophoresis to confirm the presence and absence of bands of the appropriate sizes (Figure 30).

5.5.10 Estimation of exopolysaccharide production

Exopolysaccharide (EPS) production was estimated according to established protocols (Mancuso Nichols, Garon et al. 2004; Saravanan and Jayachandran 2008). Briefly, 0.9 mL of cell-free supernatant was mixed with 0.6 mL of absolute ethanol and incubated overnight at -20°C. Each crude EPS sample was then harvested by centrifugation (16,000 rcf, 20 minutes) at 4°C, washed twice with 70% ethanol, air-dried for 2 hours at RT and dissolved in 360 μL of distilled H_2O. The polysaccharide content of each sample was then determined with the phenol-sulfuric acid method (Dubois, Gilles et al. 1956). Glucose was used as the reference standard.

CHAPTER 6

6 Future Directions

6.1 Additional Gram-negative host species

In conjunction with the validation of additional IncP1-α compatible hosts from the *α*-, *β*- and *γ-Proteobacteria*, the efficacy of functional metagenomic screening strategies could be substantially increased by the establishment of diverse Gram-negative hosts from other bacterial divisions, such as the *δ-Proteobacteria*, *Acidobacteria*, *Bacteroidetes* and *Verrucomicrobia*. There is currently just a single report describing the transfer and stable replication of an IncP1-α group vector within a member of the *Acidobacteria* (*Acidobacterium capsulatum* 161) (Quentmeier and Friedrich 1994). The results of this study have received zero follow-up, however, indicating that this single observation may have been artifactual. Several reports have described the integration of IncP1-α group vectors into the genomes of representative *δ-Proteobacteria*, but stable replication of IncP1-α group vectors has never been observed in this subphylum (Glomp, Saulnier et al. 1988; Saulnier, Hanquier et al. 1988; Jaoua, Neff et al. 1992). Attempts to transfer IncP1-α group vectors to members of the *Bacteroidetes* have not been successful, and no attempts have been made to establish

IncP1-α group vectors in cultured members of the *Verrucomicrobia* (Shoemaker and Salyers 1987; Wexler and Johnston 2010).

6.2 Gram-positive host species

Several reports have described the transfer and stable replication of IncP1-α group vectors to Gram-positive bacterial species, however, such findings constitute a substantial deviation from the typical host-range exhibited by the most well-studied IncP1-α group plasmids (Metzler, Zhang et al. 1992; Musovic, Oregaard et al. 2006). Consequently, an expansion of the list of hosts available for functional metagenomic studies to include notable Gram-positive bacterial species will likely require the use of additional vector systems. Perhaps the two most important Gram-positive divisions of bacteria to explore as hosts for functional metagenomics are the *Actinobacteria* and the *Firmicutes*, which are the two most prevalent Gram-positive bacterial phyla throughout most terrestrial environments (Nemergut, Costello et al. 2011). Along with *E. coli* and *P. putida* (both members of the *γ-Proteobacteria*), *Bacillus subtilis* (a member of the *Firmicutes*) and several species of the genus *Streptomyces* (members of the *Actinobacteria*; e.g. *S. lividans*, *S. coelicolor* and *S. avermitilis*) are cited as the most heavily relied upon heterologous hosts for natural product production (Zhang, Wang et al. 2008).

6.2.1 *Bacillus subtilis*

Like *E. coli*, *B. subtilis* is considered a safe organism for routine laboratory experimentation, and the genomes of several strains of *B. subtilis* have been fully sequenced. *B. subtilis* has long been considered the "Gram-positive model-system equivalent of *E. coli*", and as such, a number of powerful molecular biology tools are available for the genetic and metabolic engineering of this species (Errington 1988; Schumann 2007). *B. subtilis* demonstrates natural competence, which permits foreign DNA to be introduced and recombinantly inserted within the native chromosome. The convenient growth properties of *B. subtilis* facilitate rapid fermentation protocols and allow for the rapid optimization of natural product production schemes. Various strains of *B. subtilis* have been reported to produce a variety of natural products, including nonribosomal peptides, polyketides, and terpenoids (Zhang, Wang et al. 2008). As a result, *B. subtilis* is believed to possess many of the essential components needed for heterlogous natural product production, including the sfp 4'-phosphopantetheinyltransferase, which has been used successfully for metabolic engineering studies in *E. coli* (Mootz, Finking et al. 2001). The use of *B. subtilis* for the heterologous production of natural products is limited, however, by the lack of stable plasmid vectors for the transfer and replication of large natural product gene clusters. Chromosomal integration is a potential solution to this problem, but one that limits the functional gene dosage of foreign genes and operons.

6.2.2 *Streptomyces* spp.

A vast array of important secondary metabolites, including antibacterials, antifungals, immunosuppressants and anticancer agents, are produced by soil *Actinobacteria*. Within the *Actinobacteria* phylum, *Streptomyces* is the largest antibiotic producing genus, providing more than 60% of all characterized antibiotics of microbial origin (~6000 in total). Examples of natively produced *Streptomyces* natural products include diverse polyketides, non-ribosomal peptides and terpenoids: erythromycin A, tacrolimus and pikromycin (macrolides), amphotericin B and nystatin (polyene antibiotics), tetracycline and daunorubicin (type II polyketides), vancomycin and bleomycin (non-ribosomal peptides) and terpentecin and phenalinolactones (isoprenoids) (McGuire, Bunch et al. 1952; Devoe, Renfroe et al. 1963; Jarai, Jozsa et al. 1964; Linke, Mechlinski et al. 1974; Isshiki, Tamamura et al. 1985; Tanaka, Kuroda et al. 1987; Durr, Schnell et al. 2006).

Several *Streptomyces* spp. and related actinomycetes have already been established as successful hosts for the heterologous production of secondary metabolites. These efforts were initiated, in part, because of the prolific biosynthetic capacity of the *Streptomyces* genus, which led researchers to conclude that many *Streptomyces* spp. possess (i) the metabolic background needed for ample production of biosynthetic precursors, (ii) auxiliary enzymes needed for the post-translational modification of biosynthetic enzymes and (iii) robust resistance mechanisms that provide

protection from newly synthesized bioactive compounds. Several *Streptomyces* species, such as *S. coelicolor*, *Saccharopolyspora erythraea* and *S. avermitilis* have been sequenced and experimentally characterized (Bentley, Chater et al. 2002; Ikeda, Ishikawa et al. 2003; Oliynyk, Samborskyy et al. 2007).

Streptomyces spp. are also nonpathogenic, and many are associated with established fermentation protocols. However, when compared to other popular hosts for natural product production (*E. coli*, *P. putida* and *B. subtilis*), most *Streptomyces* spp. demonstrate significantly slower growth rates and more complex growth profiles. *Streptomyces* spp. are also associated with more difficult protocols for routine genetic manipulations (e.g. PEG-assisted protoplast DNA transfer). The polyphasic growth characteristics of *Streptomyces* spp. frequently necessitate extended batch runs, require complex media formulations, and increase the risk of competing microbial contamination. As a result of these challenges, large-scale functional metagenomic screening in *Streptomyces* has never been achieved. Nevertheless, the establishment of robust *Streptomyces*-based systems for functional metagenomic screening is one of the highest priorities for natural product researchers interested in the chemistries of uncultured bacteria.

6.3 Combined selection and screening methods for enhanced detection of natural product gene clusters

In addition to using an expanded host-range, the efficacy of functional metagenomic screening strategies would likely be increased by the addition of upstream selection protocols that enriched for clones containing natural product biosynthetic gene clusters. From a practical standpoint, combined selection and screening protocols would serve to limit the number of clones that required subsequent screening and would increase the amount of attention that could be paid to the relatively smaller number of clones that survived the initial selection processes. To this end, several academic labs have embraced a selection scheme based on the action of sfp-type 4'-phosphopantetheinyltransferase enzymes (PPTs), which for many species are necessary for survival under iron-limiting conditions due to their role in siderophore production (Hider and Kong 2010).

Non-ribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) holoenzymes are modified post-translationally by sfp-type PPTs, which catalyze the covalent attachment of 4'-phosphopantetheinyl groups to single conserved serine residues found in peptidyl-carrier-protein and acyl-carrier-protein domains, respectively (Lambalot, Gehring et al. 1996). Because of their functional necessity, genes encoding sfp-type PPTs are frequently found within or directly adjacent to NRPS and PKS containing gene clusters. Therefore, a simple selection scheme can be performed using "PPT knock-out"

mutant strains of *E. coli* or other heterologous hosts, in which bacterial growth under iron-limiting conditions is dependent on the heterologous complementation of endogenous siderophore biosynthetic pathways by eDNA-derived sfp-type PPT homologs. eDNA clones that are selected for in this manner will usually contain, in addition to various sfp-type PPT homologs, whole or partial gene clusters encoding NRPSs and/or PKSs. Thus, this selective enrichment strategy targets two of the most important classes of secondary metabolic enzymes.

This strategy was first optimized using standard cloning strains of *E. coli*, which rely a on a single NRPS-derived siderophore (enterobactin) for iron acquisition (Blattner, Plunkett et al. 1997; Braun and Braun 2002; Raymond, Dertz et al. 2003). The enterobactin biosynthetic gene cluster contains a sfp-type PPT-encoding gene, *entD*, whose product is required for the post-translational modification of entB and entF (Grossman, Tuckman et al. 1993). Deletion of *entD* from wild-type *E. coli* completely abolishes enterobactin production, and *entD* knock-out strains of *E. coli* are unable to survive on iron-deficient media, making them suitable for the selection scheme described above. Upon transfer of an eDNA library into *entD* knock-out *E. coli*, selection for eDNA clones that are capable of complementing the remainder of the enterobactin biosynthetic pathway is likely to yield eDNA-inserts that are significantly enriched in small molecule biosynthetic genes. These clones are then available to be screened for small molecule production

using a combination of standard phenotypic criteria and organic extract analysis.

6.4 Follow-up studies of the PEP-CTERM/exosortase system

The genetic link between bacterial *N*-acyl amino acid synthase enzymes (NASs and ExoATs) and the putative PEP-CTERM/exosortase protein sorting system was described in Chapter 5, along with the establishment of *Pseudoalteromonas atlantica* T6c as a model system for studying various PEP-CTERM/exosortase system components. Figure 26B depicts the currently proposed model for the transcriptional regulation and post-translational processing of PEP-CTERM motif containing proteins. This model is applicable to members of the *Proteobacteria* that contain the full set of conserved PEP-CTERM/exosortase system elements, including PrsK, PrsR, and EpsH (Haft, Paulsen et al. 2006). The PEP-CTERM/exosortase system has thus far only been described bioinformatically, meaning that there has been no experimental verification of the model proposed in Figure 26B. While the lack of prior functional studies makes experimental design more difficult, it also indicates that a wealth of information remains to be learned concerning all aspects of PEP-CTERM/exosortase system function.

Along with uncovering the contributions of NAS and ExoAT enzymes to the PEP-CTERM/exosortase system, four distinct aspects of the PEP-CTERM/exosortase system represent interesting and readily approachable

areas of study: (i) verification that the PrsK/PrsR two-component system controls the level of transcription of PEP-CTERM motif proteins *in vivo*, and if so, how temporal constraints impact this regulation (e.g. growth phase), (ii) identification of the PrsK activating stimulus that initiates phosphorelay between PrsK and PrsR and leads to subsequent increases in PEP-CTERM motif protein expression, (iii) verification that the EpsH enzyme (exosortase) actually fuctions as a transpeptidase that targets proteins containing the PEP-CTERM motif, and if so, precise identification of the site(s) of cleavage within PEP-CTERM motif containing proteins (i.e. defining site(s) of proteolytic cleavage by EpsH), and (iv) determination of the biological fate(s) of PEP-CTERM motif containing proteins - for example, some may become incorporated into the cell wall (similar to sortase-processed LPXTG-containing proteins in Gram-positive bacteria) while others are released into the extracellular space or used as proteinaceous modifications to secreted exopolysaccharide.

APPENDIX

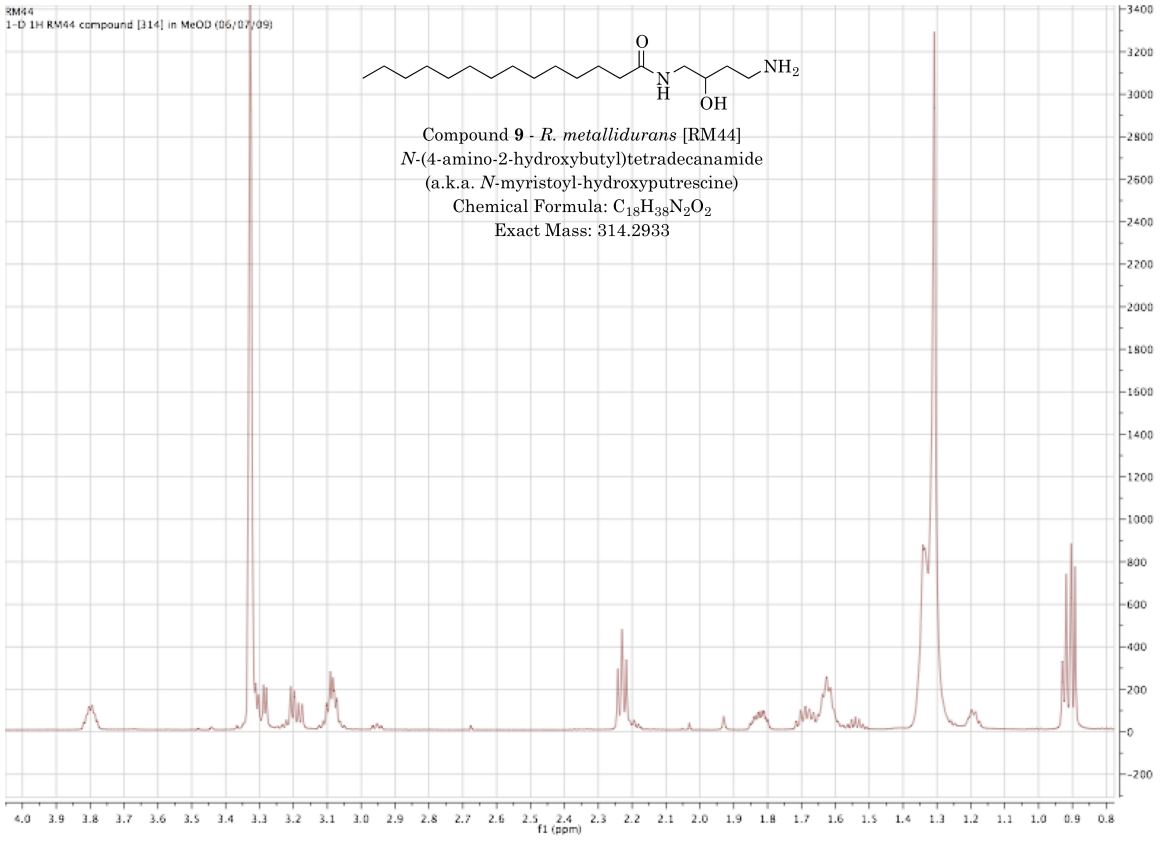


Figure 32: ^{1}H NMR spectrum of compound 9.

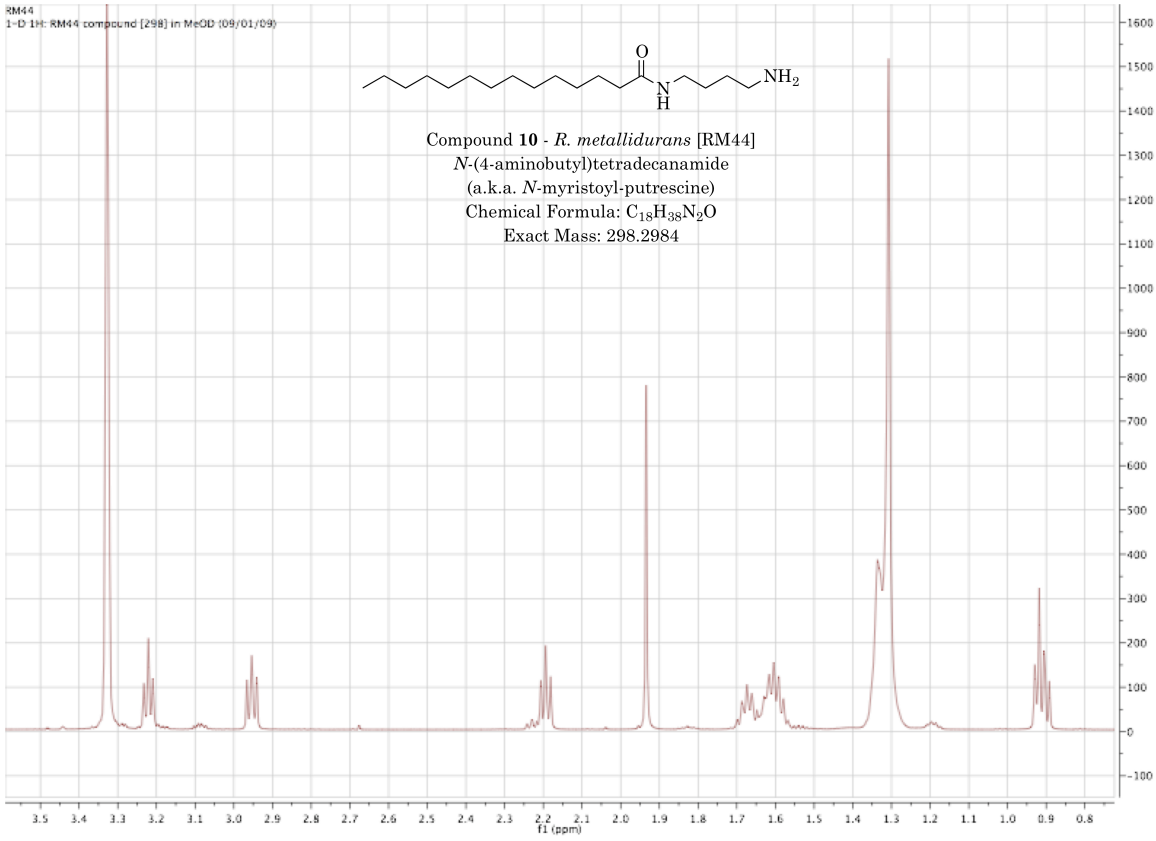


Figure 33: ^{1}H NMR spectrum of compound 10.

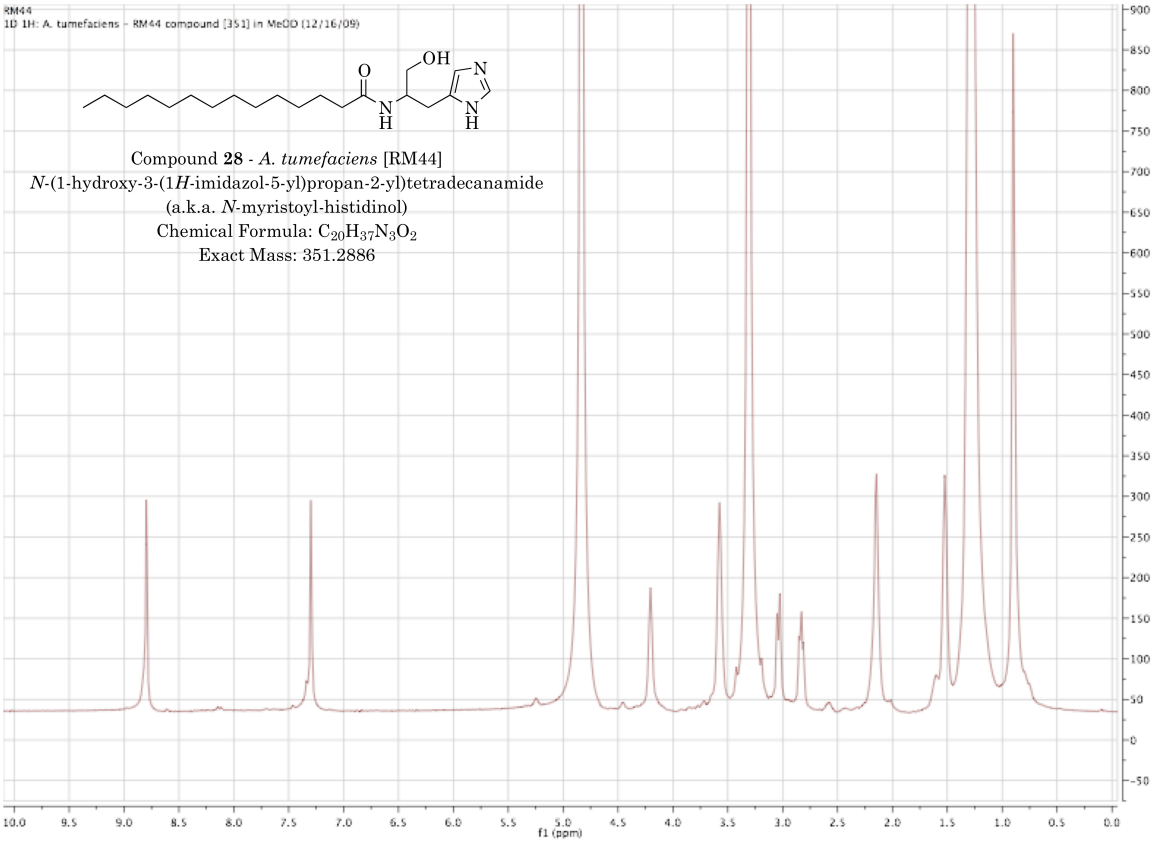


Figure 34: ^{1}H NMR spectrum of compound 28.

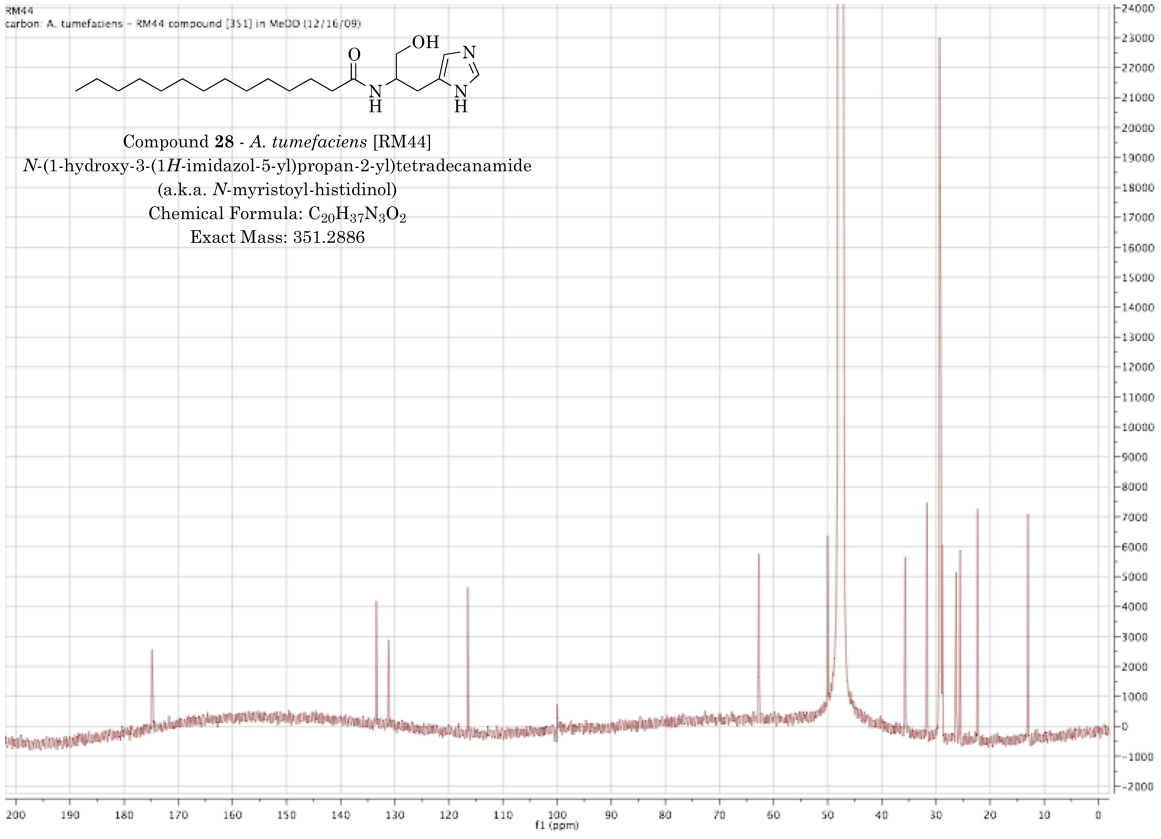


Figure 35: ^{13}C NMR spectrum of compound 28.

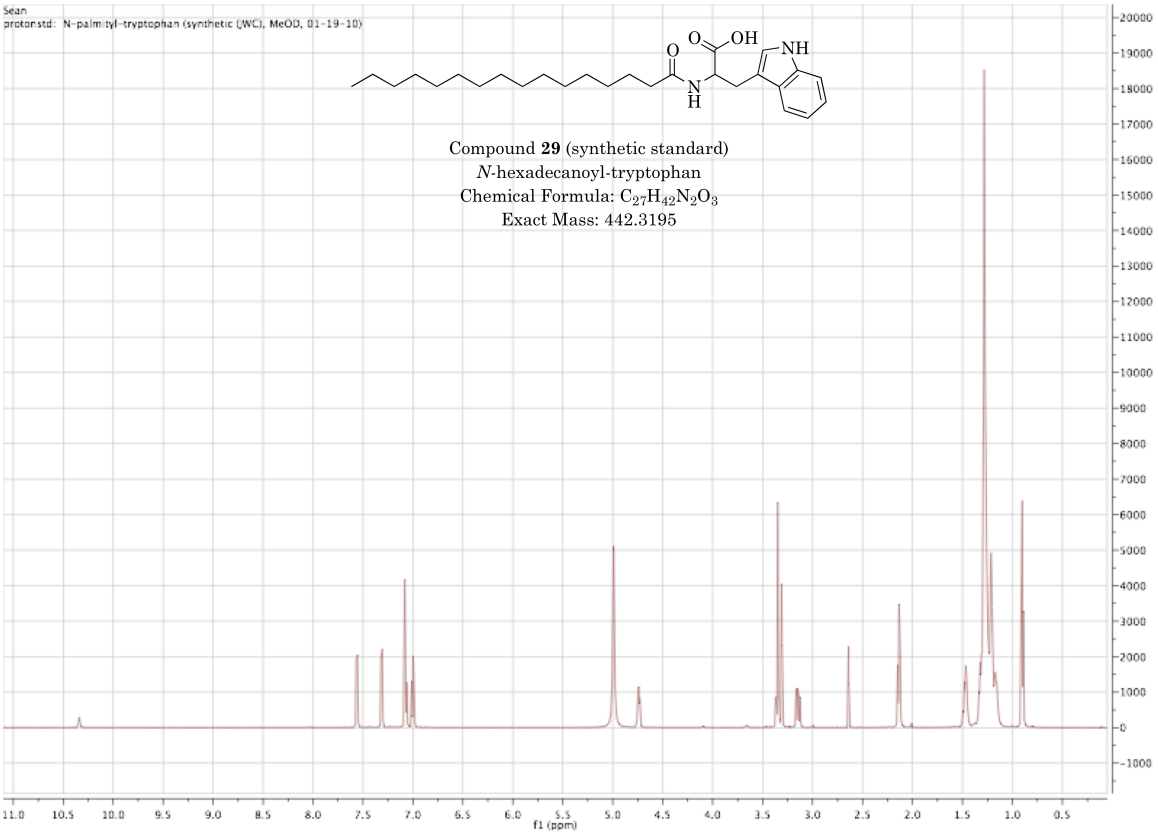


Figure 36: ^{1}H NMR spectrum of compound 29.

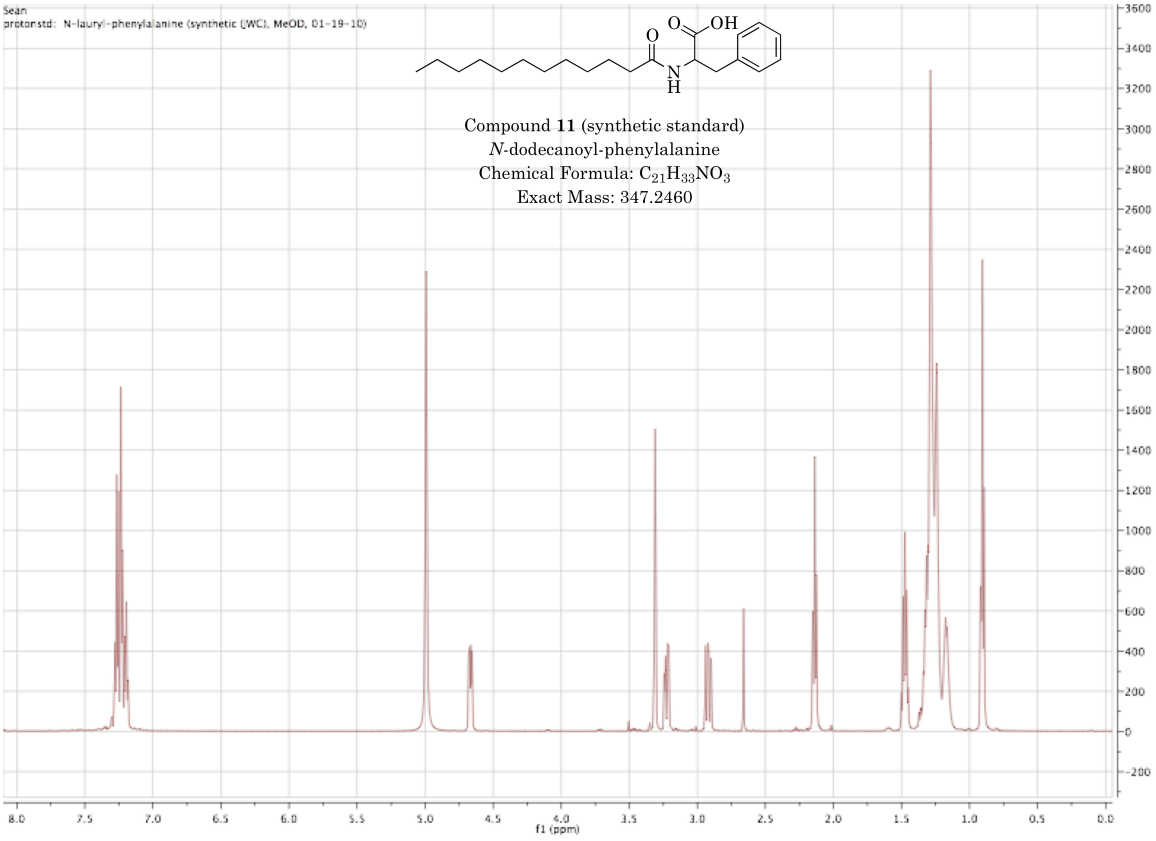


Figure 37: ^{1}H NMR spectrum of compound 11.

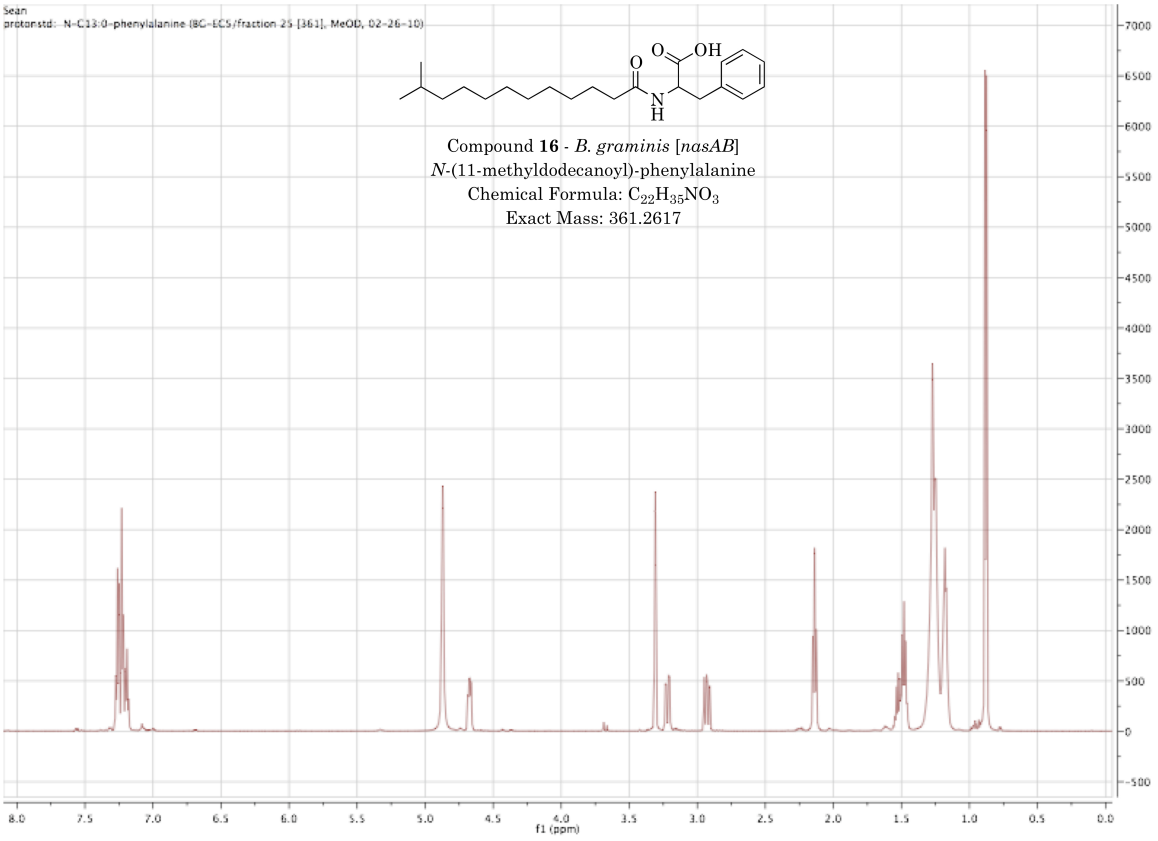


Figure 38: ^{1}H NMR spectrum of compound 16.

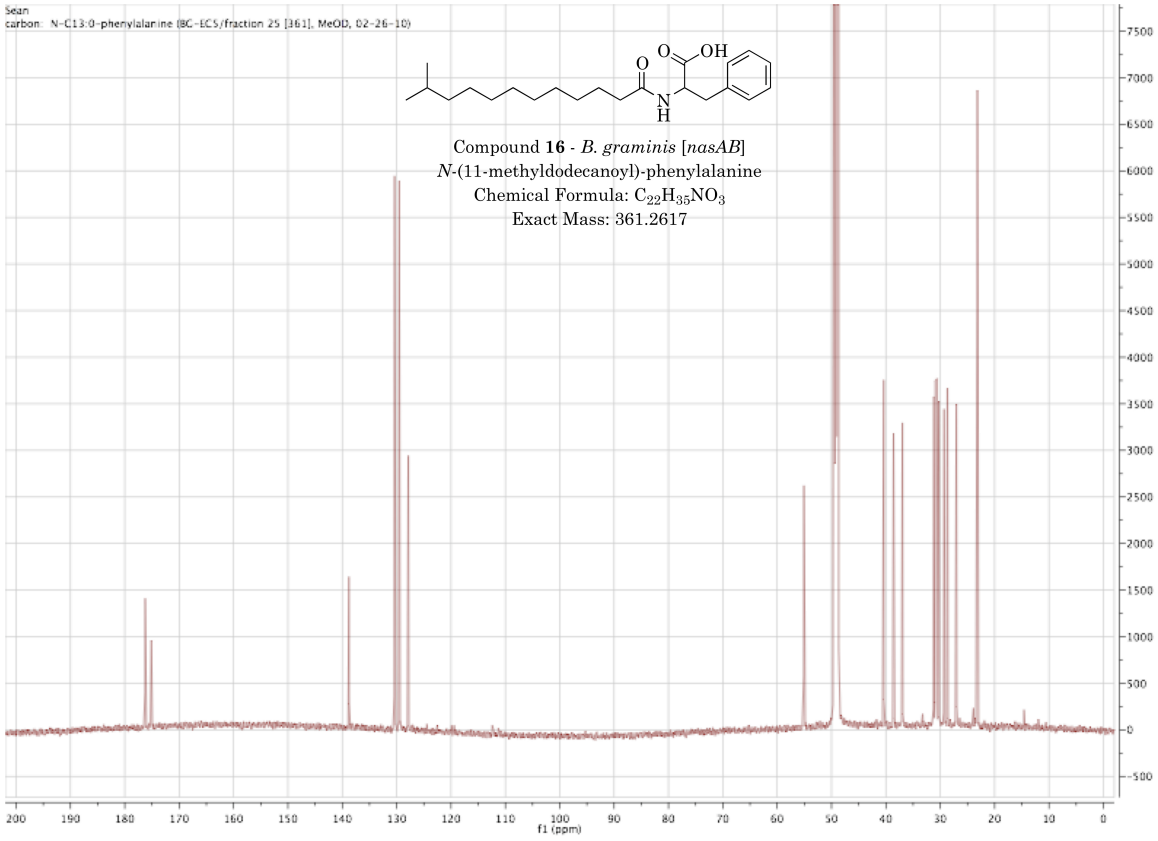


Figure 39: ^{13}C NMR spectrum of compound 16.

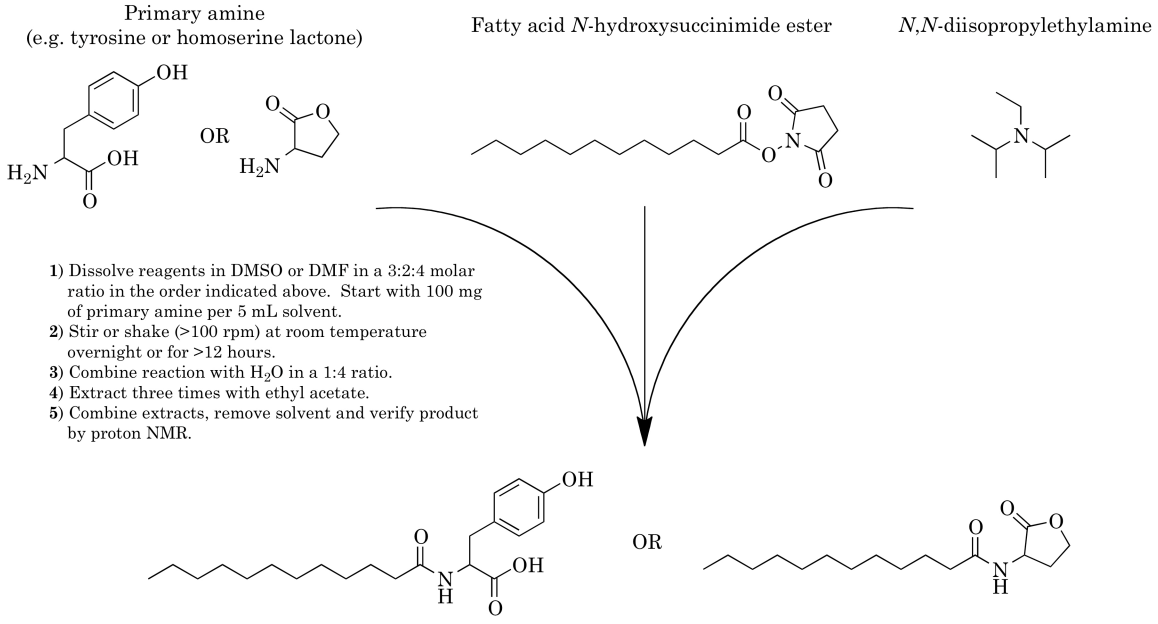


Figure 40: Total chemical synthesis of *N*-acyl amino acids.

REFERENCES

Aakvik, T., K. F. Degnes, et al. (2009). "A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species." FEMS Microbiol Lett **296**(2): 149-58.

Abbanat, D., M. Leighton, et al. (1998). "Cell wall active antifungal compounds produced by the marine fungus hypoxylon oceanicum LL-15G256. I. Taxonomy and fermentation." J Antibiot (Tokyo) **51**(3): 296-302.

Adamczyk, M. and G. Jagura-Burdzy (2003). "Spread and survival of promiscuous IncP-1 plasmids." Acta Biochim Pol **50**(2): 425-453.

Anantharaman, V. and L. Aravind (2003). "Application of comparative genomics in the identification and analysis of novel families of membrane-associated receptors in bacteria." BMC Genomics **4**(1): 34.

Ansari, M. Z., G. Yadav, et al. (2004). "NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases." Nucleic Acids Res **32**(Web Server issue): W405-13.

Audit, B. and C. A. Ouzounis (2003). "From genes to genomes: universal scale-invariant properties of microbial chromosome organisation." J Mol Biol **332**(3): 617-33.

Barns, S. M., S. L. Takala, et al. (1999). "Wide distribution and diversity of members of the bacterial kingdom Acidobacterium in the environment." Appl. Environ. Microbiol. **65**(4): 1731-7.

Bassler, B. L. and R. Losick (2006). "Bacterially speaking." Cell **125**(2): 237-46.

Beja, O., M. T. Suzuki, et al. (2000). "Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage." Environ Microbiol **2**(5): 516-29.

Belas, R., D. Bartlett, et al. (1988). "Cloning and Gene Replacement Mutagenesis of a Pseudomonas atlantica Agarase Gene." Appl Environ Microbiol **54**(1): 30-37.

Bentley, S. D., K. F. Chater, et al. (2002). "Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)." Nature **417**(6885): 141-7.

Bergquist, P. L. (1987). Incompatibility. Plasmids—a practical approach. K. G. Hardy. Oxford, United Kingdom, IRL Press.

Beringer, J. E. (1974). "R factor transfer in Rhizobium leguminosarum." J Gen Microbiol **84**(1): 188-98.

Bertrand, H., F. Poly, et al. (2005). "High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction." J Microbiol Methods **62**(1): 1-11.

Bierman, M., R. Logan, et al. (1992). "Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp." Gene **116**(1): 43-9.

Blattner, F. R., G. Plunkett, 3rd, et al. (1997). "The complete genome sequence of Escherichia coli K-12." Science **277**(5331): 1453-62.

Brady, S. F. (2007). "Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules." Nat. Protoc. **2**(5): 1297-305.

Brady, S. F., C. J. Chao, et al. (2002). "New natural product families from an environmental DNA (eDNA) gene cluster." J. Am. Chem. Soc. **124**(34): 9968-9.

Brady, S. F., C. J. Chao, et al. (2004). "Long-chain N-acyltyrosine synthases from environmental DNA." Appl Environ Microbiol **70**(11): 6865-70.

Brady, S. F. and J. Clardy (2000). "Long-chain N-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA." J. Am. Chem. Soc. **122**: 12903-12904.

Brady, S. F. and J. Clardy (2004). "Palmitoylputrescine, an antibiotic isolated from the heterologous expression of DNA extracted from bromeliad tank water." J Nat Prod **67**(8): 1283-6.

Brady, S. F. and J. Clardy (2005). "Cloning and heterologous expression of isocyanide biosynthetic genes from environmental DNA." Angew. Chem. Int. Ed. Engl. **44**(43): 7063-5.

Brady, S. F. and J. Clardy (2005). "N-acyl derivatives of arginine and tryptophan isolated from environmental DNA expressed in Escherichia coli." Org Lett **7**(17): 3613-6.

Braun, V. and M. Braun (2002). "Iron transport and signaling in Escherichia coli." FEBS Lett **529**(1): 78-85.

Breitbart, M., P. Salamon, et al. (2002). "Genomic analysis of uncultured marine viral communities." Proc Natl Acad Sci U S A **99**(22): 14250-5.

Brulc, J. M., D. A. Antonopoulos, et al. (2009). "Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases." Proc Natl Acad Sci U S A **106**(6): 1948-53.

Chen, I. C., W. D. Lin, et al. (2009). "Isolation and characterization of a novel lysine racemase from a soil metagenomic library." Appl Environ Microbiol **75**(15): 5161-6.

Chen, W., W. Tang, et al. (2006). "Pregnane, coumarin and lupane derivatives and cytotoxic constituents from Helicteres angustifolia." Phytochemistry **67**(10): 1041-7.

Choi, K. H., R. J. Heath, et al. (2000). "beta-ketoacyl-acyl carrier protein synthase III (FabH) is a determining factor in branched-chain fatty acid biosynthesis." J Bacteriol **182**(2): 365-70.

Chooi, Y. H. and Y. Tang (2010). "Adding the lipo to lipopeptides: do more with less." Chem Biol **17**(8): 791-3.

Chu, X., H. He, et al. (2008). "Identification of two novel esterases from a marine metagenomic library derived from South China Sea." Appl Microbiol Biotechnol **80**(4): 615-25.

Chung, E. J., H. K. Lim, et al. (2008). "Forest soil metagenome gene cluster involved in antifungal activity expression in Escherichia coli." Appl Environ Microbiol **74**(3): 723-30.

Clardy, J. and S. F. Brady (2007). "Cyclic AMP directly activates NasP, an N-acyl amino acid antibiotic biosynthetic enzyme cloned from an uncultured beta-proteobacterium." J Bacteriol **189**(17): 6487-9.

Claros, M. G. and G. von Heijne (1994). "TopPred II: an improved software for membrane protein structure predictions." Comput Appl Biosci **10**(6): 685-6.

Cole, J. R., B. Chai, et al. (2007). "The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data." Nucleic Acids Res **35**(Database issue): D169-72.

Cole, J. R., Q. Wang, et al. (2009). "The Ribosomal Database Project: improved alignments and new tools for rRNA analysis." Nucleic Acids Res **37**(Database issue): D141-5.

Collins, J. and B. Hohn (1978). "Cosmids: a type of plasmid gene-cloning vector that is packageable in vitro in bacteriophage lambda heads." Proc Natl Acad Sci U S A **75**(9): 4242-6.

Corey, E. J. and W. D. Li (1999). "Total synthesis and biological activity of lactacystin, omuralide and analogs." Chem Pharm Bull (Tokyo) **47**(1): 1-10.

Corpe, W. A. (1973). Microfouling: The role of primary film forming marine bacteria. Third International Congress on Marine Corrosion and Fouling . Evanston, Illinois, Northwestern U.P.

Courtois, S., C. M. Cappellano, et al. (2003). "Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products." Appl. Environ. Microbiol. **69**(1): 49-55.

Couturier, M., F. Bex, et al. (1988). "Identification and classification of bacterial plasmids." Microbiol Rev **52**(3): 375-95.

Craig, J. W. and S. F. Brady (2011). "Discovery of a metagenome-derived enzyme that produces branched-chain acyl-(acyl-carrier-protein)s from branched-chain α-keto acids." Chembiochem **In Press.**

Craig, J. W. and S. F. Brady (2011). Expanding small molecule functional metagenomics through parallel screening of broad host-range cosmid environmental DNA libraries in diverse Proteobacteria. Handbook of Molecular Microbial Ecology, Volume II: Metagenomics in Different Habitats. F. J. de Bruijn, John Wiley & Sons, Inc. **2:** 507-515.

Craig, J. W., F.-Y. Chang, et al. (2010). "Expanding small molecule functional metagenomics through parallel screening of broad host-range cosmid environmental DNA libraries in diverse Proteobacteria." Appl Environ Microbiol **76**(5): 1633-1641.

Craig, J. W., F. Y. Chang, et al. (2009). "Natural products from environmental DNA hosted in Ralstonia metallidurans." ACS Chem Biol **4**(1): 23-8.

Craig, J. W., M. A. Cherry, et al. (2011). "Bacterial N-acyl amino acid synthases are linked to the putative PEP-CTERM/exosortase protein-sorting system." J Bacteriol: Manuscript submitted.

Datta, N. (1979). Plasmid classification: incompatibility grouping. Plasmids of medical, environmental and commercial importance. K. N. Timmis and A. Puhler. Amsterdam, The Netherlands, Elsevier/North Holland Publishing Co.**:** 3-12.

De Mot, R. and A. H. Parret (2002). "A novel class of self-sufficient cytochrome P450 monooxygenases in prokaryotes." Trends Microbiol. **10**(11): 502-8.

Delcher, A. L., D. Harmon, et al. (1999). "Improved microbial gene identification with GLIMMER." Nucleic Acids Res **27**(23): 4636-41.

DeSantis, T. Z., P. Hugenholtz, et al. (2006). "Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB." Appl Environ Microbiol **72**(7): 5069-72.

DeSantis, T. Z., Jr., P. Hugenholtz, et al. (2006). "NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes." Nucleic Acids Res **34**(Web Server issue): W394-9.

Devoe, S. E., H. B. Renfroe, et al. (1963). "Production of Picromycin by Cultures Deposited as Methymycin Producers." Antimicrob Agents Chemother (Bethesda) **161**: 125-9.

Dickschat, J. S. (2010). "Quorum sensing and bacterial biofilms." Nat Prod Rep **27**(3): 343-69.

Dixit, A. K., J.-J. Chen, et al. (2003). "Alkyl phenols from the leaves of Formosan Ardisia virens." Chin. Pharm. J. (Taipei, Taiwan) **55**(4): 273-278.

Dower, W. J., J. F. Miller, et al. (1988). "High efficiency transformation of E. coli by high voltage electroporation." Nucleic Acids Res **16**(13): 6127-45.

Dubois, M., K. A. Gilles, et al. (1956). "Colorimetric method for determination of sugars and related substances." Analytical Chemistry **28**: 350-356.

Durr, C., H. J. Schnell, et al. (2006). "Biosynthesis of the terpene phenalinolactone in Streptomyces sp. Tu6071: analysis of the gene cluster and generation of derivatives." Chem Biol **13**(4): 365-77.

Elend, C., C. Schmeisser, et al. (2006). "Isolation and biochemical characterization of two novel metagenome-derived esterases." Appl Environ Microbiol **72**(5): 3637-45.

Engleberg, N. C., N. Cianciotto, et al. (1988). "Transfer and maintenance of small, mobilizable plasmids with ColE1 replication origins in Legionella pneumophila." Plasmid **20**(1): 83-91.

Entcheva, P., W. Liebl, et al. (2001). "Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia." Appl Environ Microbiol **67**(1): 89-99.

Errington, J. (1988). "Generalized cloning vectors for Bacillus subtilis." Biotechnology **10**: 345-62.

Feinstein, L. M., W. J. Sul, et al. (2009). "Assessment of bias associated with incomplete extraction of microbial DNA from soil." Appl Environ Microbiol **75**(16): 5428-33.

Fierer, N., M. A. Bradford, et al. (2007). "Toward an ecological classification of soil bacteria." Ecology **88**(6): 1354-64.

Finn, R. D., J. Mistry, et al. (2010). "The Pfam protein families database." Nucleic Acids Res **38**(Database issue): D211-22.

Finn, R. D., J. Tate, et al. (2008). "The Pfam protein families database." Nucleic Acids Res **36**(Database issue): D281-8.

Friedman, A. M., S. R. Long, et al. (1982). "Construction of a broad host range cosmid cloning vector and its use in the genetic analysis of Rhizobium mutants." Gene **18**(3): 289-96.

Frostegard, A., S. Courtois, et al. (1999). "Quantification of bias related to the extraction of DNA directly from soils." Appl Environ Microbiol **65**(12): 5409-20.

Funabashi, M., N. Funa, et al. (2008). "Phenolic lipids synthesized by type III polyketide synthase confer penicillin resistance on Streptomyces griseus." J. Biol. Chem. **283**(20): 13983-91.

Gabor, E. M., W. B. Alkema, et al. (2004). "Quantifying the accessibility of the metagenome by random expression cloning techniques." Environ Microbiol **6**(9): 879-86.

Gabor, E. M., E. J. de Vries, et al. (2003). "Efficient recovery of environmental DNA for expression cloning by indirect extraction methods." FEMS Microbiol Ecol **44**(2): 153-63.

Gans, J., M. Wolinsky, et al. (2005). "Computational improvements reveal great bacterial diversity and high metal toxicity in soil." Science **309**(5739): 1387-90.

Gao, J. L., B. Weissenmayer, et al. (2004). "Identification of a gene required for the formation of lyso-ornithine lipid, an intermediate in the biosynthesis of ornithine-containing lipids." Mol Microbiol **53**(6): 1757-70.

Gay, P., D. Le Coq, et al. (1985). "Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria." J Bacteriol **164**(2): 918-21.

Geiger, O., N. Gonzalez-Silva, et al. (2010). "Amino acid-containing membrane lipids in bacteria." Prog Lipid Res **49**(1): 46-60.

Gilchrist, A. and J. Smit (1991). "Transformation of freshwater and marine caulobacters by electroporation." J Bacteriol **173**(2): 921-5.

Gillespie, D. E., S. F. Brady, et al. (2002). "Isolation of antibiotics turbomycin a and B from a metagenomic library of soil microbial DNA." Appl Environ Microbiol **68**(9): 4301-6.

Glomp, I., P. Saulnier, et al. (1988). "Transfer of IncP plasmids into Stigmatella aurantiaca leading to insertional mutants affected in spore development." Mol Gen Genet **214**(2): 213-7.

Grossman, T. H., M. Tuckman, et al. (1993). "Isolation and characterization of Bacillus subtilis genes involved in siderophore biosynthesis:

relationship between B. subtilis sfpo and Escherichia coli entD genes." J Bacteriol **175**(19): 6203-11.

Guiney, D. G. and E. Yakobson (1983). "Location and nucleotide sequence of the transfer origin of the broad host range plasmid RK2." Proc Natl Acad Sci U S A **80**(12): 3595-8.

Haft, D. H., I. T. Paulsen, et al. (2006). "Exopolysaccharide-associated protein sorting in environmental organisms: the PEP-CTERM/EpsH system. Application of a novel phylogenetic profiling heuristic." BMC Biol **4**: 29.

Hamann, M. T., C. S. Otto, et al. (1996). "Kahalalides: Bioactive Peptides from a Marine Mollusk Elysia rufescens and Its Algal Diet Bryopsis sp.(1)." J Org Chem **61**(19): 6594-6600.

Handelsman, J., M. R. Rondon, et al. (1998). "Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products." Chem Biol **5**(10): R245-9.

Harrison, D. E., R. Strong, et al. (2009). "Rapamycin fed late in life extends lifespan in genetically heterogeneous mice." Nature **460**(7253): 392-5.

Healy, F. G., R. M. Ray, et al. (1995). "Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose." Appl Microbiol Biotechnol **43**(4): 667-74.

Heinemann, J. A. and G. F. Sprague, Jr. (1989). "Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast." Nature **340**(6230): 205-9.

Henne, A., R. Daniel, et al. (1999). "Construction of environmental DNA libraries in Escherichia coli and screening for the presence of genes conferring utilization of 4-hydroxybutyrate." Appl Environ Microbiol **65**(9): 3901-7.

Herring, P. (1972). "Porphyrin Pigmentation in Deep-sea Medusae." Nature **238**: 276-277.

Hider, R. C. and X. Kong (2010). "Chemistry and biology of siderophores." Nat Prod Rep **27**(5): 637-57.

Howitz, K. T., K. J. Bitterman, et al. (2003). "Small molecule activators of sirtuins extend Saccharomyces cerevisiae lifespan." Nature **425**(6954): 191-6.

Huang, Y., X. Lai, et al. (2009). "Characterization of a deep-sea sediment metagenomic clone that produces water-soluble melanin in Escherichia coli." Mar Biotechnol (NY) **11**(1): 124-31.

Ikeda, H., J. Ishikawa, et al. (2003). "Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis." Nat Biotechnol **21**(5): 526-31.

Isshiki, K., T. Tamamura, et al. (1985). "The structure of a new antibiotic, terpentecin." J Antibiot (Tokyo) **38**(12): 1819-21.

Jannasch, H. W. and G. E. Jones (1959). "Bacterial Populations in sea water determined by different methods of enumeration." Limnol Oceanography **4**(4): 128-139.

Jaoua, S., S. Neff, et al. (1992). "Transfer of mobilizable plasmids to Sorangium cellulosum and evidence for their integration into the chromosome." Plasmid **28**(2): 157-65.

Jarai, M., G. Jozsa, et al. (1964). "Biochemical Studies on Streptomyces Aureofaciens. Iv. Studies on the Biosynthesis of Chlortetracycline." Acta Microbiol Acad Sci Hung **11**: 203-10.

Jeon, J. H., J. T. Kim, et al. (2009). "Characterization and its potential application of two esterases derived from the arctic sediment metagenome." Mar Biotechnol (NY) **11**(3): 307-16.

Jordan, S. W. and J. E. Cronan, Jr. (1997). "A new metabolic link. The acyl carrier protein of lipid synthesis donates lipoic acid to the pyruvate dehydrogenase complex in Escherichia coli and mitochondria." J Biol Chem **272**(29): 17903-6.

Joseph, S. J., P. Hugenholtz, et al. (2003). "Laboratory cultivation of widespread and previously uncultured soil bacteria." Appl. Environ. Microbiol. **69**(12): 7210-5.

Kaeberlein, T., K. Lewis, et al. (2002). "Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment." Science **296**(5570): 1127-9.

Kaneda, T. (1963). "Biosynthesis of branched chain fatty acids. II. Microbial synthesis of branched long chain fatty acids from certain short chain fatty acid substrates." J Biol Chem **238**: 1229-35.

Kaneda, T. (1991). "Iso- and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance." Microbiol Rev **55**(2): 288-302.

Keenleyside, W. J. and C. Whitfield (1995). "Lateral transfer of rfb genes: a mobilizable ColE1-type plasmid carries the rfbO:54 (O:54 antigen biosynthesis) gene cluster from Salmonella enterica serovar Borreze." J Bacteriol **177**(18): 5247-53.

Keller, M. and K. Zengler (2004). "Tapping into microbial diversity." Nat Rev Microbiol **2**(2): 141-50.

Kim, J. S., H. K. Lim, et al. (2009). "Production of porphyrin intermediates in Escherichia coli carrying soil metagenomic genes." FEMS Microbiol Lett **295**(1): 42-9.

Kim, U. J., H. Shizuya, et al. (1992). "Stable propagation of cosmid sized human DNA inserts in an F factor based vector." Nucleic Acids Res **20**(5): 1083-5.

Knietsch, A., S. Bowien, et al. (2003). "Identification and characterization of coenzyme B12-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures." Appl Environ Microbiol **69**(6): 3048-60.

Kolatka, K., S. Kubik, et al. (2010). "Replication and partitioning of the broad-host-range plasmid RK2." Plasmid **64**(3): 119-34.

Kudla, G., A. W. Murray, et al. (2009). "Coding-sequence determinants of gene expression in Escherichia coli." Science **324**(5924): 255-8.

Kwon, S. J., M. Y. Lee, et al. (2007). "High-throughput, microarray-based synthesis of natural product analogues via in vitro metabolic pathway construction." ACS Chem. Biol. **2**(6): 419-25.

Lambalot, R. H., A. M. Gehring, et al. (1996). "A new enzyme superfamily - the phosphopantetheinyl transferases." Chem Biol **3**(11): 923-36.

Lange, B. M., T. Rujan, et al. (2000). "Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes." Proc Natl Acad Sci U S A **97**(24): 13172-7.

Lee, C. M., Y. S. Yeo, et al. (2008). "Identification of a novel 4-hydroxyphenylpyruvate dioxygenase from the soil metagenome." Biochem Biophys Res Commun **370**(2): 322-6.

Li, Y., M. Wexler, et al. (2005). "Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of Rhizobium leguminosarum and of Escherichia coli reveals different classes of cloned trp genes." Environ Microbiol **7**(12): 1927-36.

Liles, M. R., B. F. Manske, et al. (2003). "A census of rRNA genes and linked genomic sequences within a soil metagenomic library." Appl Environ Microbiol **69**(5): 2684-91.

Lin, J.-J. (1994). "Optimization of the transformation efficiency of Agrobacterium tumefaciens cells using electroporation." Plant Science **101**: 11-15.

Linke, H. A., W. Mechlinski, et al. (1974). "Production of amphotericin B-14C by Streptomyces nodosus fermentation, and preparation of the amphotericin B-14C-methyl-ester." J Antibiot (Tokyo) **27**(3): 155-60.

Lopez, C. M., D. A. Rholl, et al. (2009). "Versatile dual-technology system for markerless allele replacement in Burkholderia pseudomallei." Appl Environ Microbiol **75**(20): 6496-503.

Lowbury, E. J., H. A. Lilly, et al. (1969). "Sensitivity of Pseudomonas aeruginosa to antibiotics: emergence of strains highly resistant to carbenicillin." Lancet **2**(7618): 448-52.

Lu, Y. J., Y. M. Zhang, et al. (2004). "Product diversity and regulation of type II fatty acid synthases." Biochem Cell Biol **82**(1): 145-55.

MacNeil, D. J., K. M. Gewain, et al. (1992). "Analysis of Streptomyces avermitilis genes required for avermectin biosynthesis utilizing a novel integration vector." Gene **111**(1): 61-8.

Mancuso Nichols, C. A., S. Garon, et al. (2004). "Production of exopolysaccharides by Antarctic marine bacterial isolates." J Appl Microbiol **96**(5): 1057-66.

Margalit, D. N., L. Romberg, et al. (2004). "Targeting cell division: small-molecule inhibitors of FtsZ GTPase perturb cytokinetic ring assembly and induce bacterial lethality." Proc Natl Acad Sci U S A **101**(32): 11821-6.

Martinez, A., S. J. Kolvek, et al. (2004). "Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts." Appl Environ Microbiol **70**(4): 2452-63.

Mattevi, A., G. Obmolova, et al. (1993). "Crystallographic analysis of substrate binding and catalysis in dihydrolipoyl transacetylase (E2p)." Biochemistry **32**(15): 3887-901.

Mazmanian, S. K., G. Liu, et al. (1999). "Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall." Science **285**(5428): 760-3.

McDowall, J. and S. Hunter (2011). "InterPro protein classification." Methods Mol Biol **694**: 37-47.

McGuire, J. M., R. L. Bunch, et al. (1952). "[Ilotycin, a new antibiotic]." Schweiz Med Wochenschr **82**(41): 1064-5.

Mergeay, M., S. Monchy, et al. (2003). "Ralstonia metallidurans, a bacterium specifically adapted to toxic metals: towards a catalogue of metal-responsive genes." FEMS Microbiol. Rev. **27**(2-3): 385-410.

Metzler, M. C., Y. P. Zhang, et al. (1992). "Transformation of the gram-positive bacterium Clavibacter xyli subsp. cynodontis by electroporation with plasmids from the IncP incompatibility group." J Bacteriol **174**(13): 4500-3.

Meyer, R., D. Figurski, et al. (1975). "Molecular vehicle properties of the broad host range plasmid RK2." Science **190**(4220): 1226-8.

Miao, V., M.-F. Coëffet-Legal, et al. (2005). "Daptomycin biosynthesis in Streptomyces roseosporus: cloning and analysis of the gene cluster and revision of peptide stereochemistry." Microbiology **151**(Pt 5): 1507-23.

Mikiewicz, D., B. Wrobel, et al. (1997). "Isolation and characterization of a ColE1-like plasmid from Enterobacter agglomerans with a novel variant of rom gene." Plasmid **38**(3): 210-9.

Miller, J. R., R. W. Busby, et al. (2000). "Escherichia coli LipA is a lipoyl synthase: in vitro biosynthesis of lipoylated pyruvate dehydrogenase complex from octanoyl-acyl carrier protein." Biochemistry **39**(49): 15166-78.

Mirete, S., C. G. de Figueras, et al. (2007). "Novel nickel resistance genes from the rhizosphere metagenome of plants adapted to acid mine drainage." Appl Environ Microbiol **73**(19): 6001-11.

Moore, B. S. and C. Hertweck (2002). "Biosynthesis and attachment of novel bacterial polyketide synthase starter units." Nat Prod Rep **19**(1): 70-99.

Mootz, H. D., R. Finking, et al. (2001). "4'-phosphopantetheine transfer in primary and secondary metabolism of Bacillus subtilis." J Biol Chem **276**(40): 37289-98.

More, M. I., L. D. Finger, et al. (1996). "Enzymatic synthesis of a quorum-sensing autoinducer through use of defined substrates." Science **272**(5268): 1655-8.

Morino, T., A. Masuda, et al. (1994). "Stevastelins, novel immunosuppressants produced by Penicillium." J Antibiot (Tokyo) **47**(11): 1341-3.

Musovic, S., G. Oregaard, et al. (2006). "Cultivation-independent examination of horizontal transfer and host range of an IncP-1 plasmid among gram-positive and gram-negative bacteria indigenous to the barley rhizosphere." Appl Environ Microbiol **72**(10): 6687-92.

Namatame, I., H. Tomoda, et al. (1999). "Beauveriolides, specific inhibitors of lipid droplet formation in mouse macrophages, produced by Beauveria sp. FO-6979." J Antibiot (Tokyo) **52**(1): 1-6.

Nemergut, D. R., E. K. Costello, et al. (2011). "Global patterns in the biogeography of bacterial taxa." Environ Microbiol **13**(1): 135-44.

Newman, D. J. and G. M. Cragg (2004). "Marine natural products and related compounds in clinical and advanced preclinical trials." J Nat Prod **67**(8): 1216-38.

Newman, D. J. and G. M. Cragg (2007). "Natural products as sources of new drugs over the last 25 years." J Nat Prod **70**(3): 461-77.

Newman, D. J., G. M. Cragg, et al. (2003). "Natural products as sources of new drugs over the period 1981-2002." J Nat Prod **66**(7): 1022-37.

Nicolaou, K. C., S. M. Dalby, et al. (2009). "Total synthesis of (+)-haplophytine." Angew Chem Int Ed Engl **48**(41): 7616-20.

Nougayrède, J.-P., S. Homburg, et al. (2006). "Escherichia coli induces DNA double-strand breaks in eukaryotic cells." Science **313**(5788): 848-51.

Novick, R. P. (1987). "Plasmid incompatibility." Microbiol Rev **51**(4): 381-95.

Ogram, A., G. S. Sayler, et al. (1987). "DNA extraction and purification from sediments." J Microbiological Methods **7**: 57-66.

Oku, H. and T. Kaneda (1988). "Biosynthesis of branched-chain fatty acids in Bacillus subtilis. A decarboxylase is essential for branched-chain fatty acid synthetase." J Biol Chem **263**(34): 18386-96.

Oliynyk, M., M. Samborskyy, et al. (2007). "Complete genome sequence of the erythromycin-producing bacterium Saccharopolyspora erythraea NRRL23338." Nat Biotechnol **25**(4): 447-53.

Omura, S., H. Ikeda, et al. (2001). "Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites." Proc Natl Acad Sci U S A **98**(21): 12215-20.

Ono, A., R. Miyazaki, et al. (2007). "Isolation and characterization of naphthalene-catabolic genes and plasmids from oil-contaminated soil by using two cultivation-independent approaches." Appl Microbiol Biotechnol **74**(2): 501-10.

Osborn, A. M. and C. J. Smith (2005). Molecular Microbial Ecology. Abingdon [England], New York, Taylor & Francis.

Paget, M. S., L. Chamberlin, et al. (1999). "Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in Streptomyces coelicolor A3(2)." J Bacteriol **181**(1): 204-11.

Parsley, L. C., E. J. Consuegra, et al. (2010). "Identification of diverse antimicrobial resistance determinants carried on bacterial, plasmid, or viral metagenomes from an activated sludge microbial assemblage." Appl Environ Microbiol **76**(11): 3753-7.

Patankar, A. V. and J. E. Gonzalez (2009). "Orphan LuxR regulators of quorum sensing." FEMS Microbiol Rev **33**(4): 739-56.

Pellegrini, M., E. M. Marcotte, et al. (1999). "Assigning protein functions by comparative genome analysis: protein phylogenetic profiles." Proc Natl Acad Sci U S A **96**(8): 4285-8.

Perham, R. N. (1991). "Domains, motifs, and linkers in 2-oxo acid dehydrogenase multienzyme complexes: a paradigm in the design of a multifunctional protein." Biochemistry **30**(35): 8501-12.

Plonka, P. M. and M. Grabacka (2006). "Melanin synthesis in microorganisms--biotechnological and medical aspects." Acta Biochim Pol **53**(3): 429-43.

Pohlmann, A., W. F. Fricke, et al. (2006). "Genome sequence of the bioplastic-producing "Knallgas" bacterium Ralstonia eutropha H16." Nat. Biotechnol. **24**(10): 1257-62.

Quaiser, A., T. Ochsenreiter, et al. (2003). "Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics." Mol Microbiol **50**(2): 563-75.

Quentmeier, A. and C. G. Friedrich (1994). "Transfer and expression of degradative and antibiotic resistance plasmids in acidophilic bacteria." Appl Environ Microbiol **60**(3): 973-8.

Rappe, M. S. and S. J. Giovannoni (2003). "The uncultured microbial majority." Annu. Rev. Microbiol. **57**: 369-94.

Raymond, K. N., E. A. Dertz, et al. (2003). "Enterobactin: an archetype for microbial iron transport." Proc Natl Acad Sci U S A **100**(7): 3584-8.

Reed, L. J. (1974). "Multienzyme Complexes." Acc Chem Res **7**: 40-46.

Riesenfeld, C. S., R. M. Goodman, et al. (2004). "Uncultured soil bacteria are a reservoir of new antibiotic resistance genes." Environ Microbiol **6**(9): 981-9.

Rock, C. O. and J. E. Cronan, Jr. (1985). Biochemistry of Lipids and Membranes. Menlo Park, CA, Benjamin/Cummings Publishing Co.

Rogozin, I. B., K. S. Makarova, et al. (2004). "Computational approaches for the analysis of gene neighbourhoods in prokaryotic genomes." Brief Bioinform **5**(2): 131-49.

Rondon, M. R., P. R. August, et al. (2000). "Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms." Appl Environ Microbiol **66**(6): 2541-7.

Salanoubat, M., S. Genin, et al. (2002). "Genome sequence of the plant pathogen Ralstonia solanacearum." Nature **415**(6871): 497-502.

Sampson, B. A., R. Misra, et al. (1989). "Identification and characterization of a new gene of Escherichia coli K-12 involved in outer membrane permeability." Genetics **122**(3): 491-501.

San Feliciano, A., J. M. Miguel del Corral, et al. (1990). "3,4-Dihydroisocoumarins from Ononis natrix." Phytochemistry **29**(3): 945-8.

Saravanan, P. and S. Jayachandran (2008). "Preliminary characterization of exopolysaccharides produced by a marine biofilm-forming bacterium Pseudoalteromonas ruthenica (SBT 033)." Lett Appl Microbiol **46**(1): 1-6.

Saulnier, P., J. Hanquier, et al. (1988). "Utilization of IncP-1 plasmids as vectors for transposon mutagenesis in myxobacteria." J Gen Microbiol **134**(11): 2889-95.

Savile, C. K., J. M. Janey, et al. (2010). "Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture." Science.

Schaefer, A. L., D. L. Val, et al. (1996). "Generation of cell-to-cell signals in quorum sensing: acyl homoserine lactone synthase activity of a purified Vibrio fischeri LuxI protein." Proc Natl Acad Sci U S A **93**(18): 9505-9.

Schatz, A. and S. A. Waksman (1944). "Effect of streptomycin and other antibiotic substances upon Mycobacterium tuberculosis and related organisms." Proceedings of the Society for Experimental Biology and Medicine (New York, N.Y.) **57**: 244-248.

Schlingmann, G., L. Milne, et al. (1998). "Cell wall active antifungal compounds produced by the marine fungus Hypoxylon oceanicum LL-15G256. II. Isolation and structure determination." J Antibiot (Tokyo) **51**(3): 303-16.

Schloss, P. D. and J. Handelsman (2003). "Biotechnological prospects from metagenomics." Curr Opin Biotechnol **14**(3): 303-10.

Schmidhauser, T. J. and D. R. Helinski (1985). "Regions of broad-host-range plasmid RK2 involved in replication and stable maintenance in nine species of gram-negative bacteria." J Bacteriol **164**(1): 446-55.

Schneewind, O., D. Mihaylova-Petkov, et al. (1993). "Cell wall sorting signals in surface proteins of gram-positive bacteria." EMBO J **12**(12): 4803-11.

Schneewind, O., P. Model, et al. (1992). "Sorting of protein A to the staphylococcal cell wall." Cell **70**(2): 267-81.

Schumann, W. (2007). "Production of recombinant proteins in Bacillus subtilis." Adv Appl Microbiol **62**: 137-89.

Schwarzer, D., R. Finking, et al. (2003). "Nonribosomal peptides: from genes to products." Nat Prod Rep **20**(3): 275-87.

Sharma, R. C. and R. T. Schimke (1996). "Preparation of electrocompetent E. coli using salt-free growth medium." Biotechniques **20**(1): 42-4.

Shizuya, H., B. Birren, et al. (1992). "Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector." Proc Natl Acad Sci U S A **89**(18): 8794-7.

Shoemaker, N. B. and A. A. Salyers (1987). "Facilitated transfer of IncP beta R751 derivatives from the chromosome of Bacteroides uniformis to Escherichia coli recipients by a conjugative Bacteroides tetracycline resistance element." J Bacteriol **169**(7): 3160-7.

Siddique, A. and D. H. Figurski (2002). "The active partition gene incC of IncP plasmids is required for stable maintenance in a broad range of hosts." J Bacteriol **184**(6): 1788-93.

Simon, C. and R. Daniel (2009). "Achievements and new knowledge unraveled by metagenomic approaches." Appl Microbiol Biotechnol **85**(2): 265-76.

Simon, C., J. Herath, et al. (2009). "Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice." Appl Environ Microbiol **75**(9): 2964-8.

Simon, R., U. Priefer, et al. (1983). "A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in gram negative bacteria." Nature Bio/Technology **1**: 784-791.

Slater, S. C., B. S. Goldman, et al. (2009). "Genome sequences of three agrobacterium biovars help elucidate the evolution of multichromosome genomes in bacteria." J Bacteriol **191**(8): 2501-11.

Smirnova, N. and K. A. Reynolds (2001). "Branched-chain fatty acid biosynthesis in Escherichia coli." J Ind Microbiol Biotechnol **27**(4): 246-51.

Smith, D., J. H. Wang, et al. (2006). "Variations on a theme: diverse N-acyl homoserine lactone-mediated quorum sensing mechanisms in gram-negative bacteria." Sci Prog **89**(Pt 3-4): 167-211.

Sobecky, P. A., T. J. Mincer, et al. (1998). "Isolation of broad-host-range replicons from marine sediment bacteria." Appl Environ Microbiol **64**(8): 2822-30.

Sogin, M. L., H. G. Morrison, et al. (2006). "Microbial diversity in the deep sea and the underexplored "rare biosphere"." Proc Natl Acad Sci U S A **103**(32): 12115-20.

Staley, J. T. and A. Konopka (1985). "Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats." Annu Rev Microbiol **39**: 321-46.

Stein, J. L., T. L. Marsh, et al. (1996). "Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon." J Bacteriol **178**(3): 591-9.

Suenaga, H., T. Ohnuki, et al. (2007). "Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds." Environ Microbiol **9**(9): 2289-97.

Sylvia, D. M. (2005). Principles and Applications of Soil Microbiology. Upper Saddle River, NJ, Pearson Prentice Hall.

Taghavi, S., D. van der Lelie, et al. (1994). "Electroporation of Alcaligenes eutrophus with (mega) plasmids and genomic DNA fragments." Appl. Environ. Microbiol. **60**(10): 3585-91.

Tanaka, H., A. Kuroda, et al. (1987). "Physicochemical properties of FK-506, a novel immunosuppressant isolated from Streptomyces tsukubaensis." Transplant Proc **19**(5 Suppl 6): 11-6.

Tankere, S. P., D. G. Bourne, et al. (2002). "Microenvironments and microbial community structure in sediments." Environ Microbiol **4**(2): 97-105.

Tebbe, C. C. and W. Vahjen (1993). "Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast." Appl Environ Microbiol **59**(8): 2657-65.

Thiel, V., B. Kunze, et al. (2009). "New structural variants of homoserine lactones in bacteria." Chembiochem **10**(11): 1861-8.

Thomas, C. M. and D. R. Helinski (1989). Vegetative replication and stable inheritance of IncP plasmids. Promiscuous Plasmids in Gram-Negative Bacteria. C. M. Thomas. London, Academic Press**:** 1-25.

Thongdee, M., L. A. Gallagher, et al. (2008). "Targeted mutagenesis of Burkholderia thailandensis and Burkholderia pseudomallei through natural transformation of PCR fragments." Appl Environ Microbiol **74**(10): 2985-9.

Torsvik, V., J. Goksoyr, et al. (1990). "High diversity in DNA of soil bacteria." Appl. Environ. Microbiol. **56**(3): 782-7.

Torsvik, V. and L. Ovreas (2002). "Microbial diversity and function in soil: from genes to ecosystems." Curr Opin Microbiol **5**(3): 240-5.

Torsvik, V., L. Ovreas, et al. (2002). "Prokaryotic diversity--magnitude, dynamics, and controlling factors." Science **296**(5570): 1064-6.

Torsvik, V., K. Salte, et al. (1990). "Comparison of phenotypic diversity and DNA heterogeneity in a population of soil bacteria." Appl Environ Microbiol **56**(3): 776-81.

Tringe, S. G., C. von Mering, et al. (2005). "Comparative metagenomics of microbial communities." Science **308**(5721): 554-7.

Trost, B. M. and G. Dong (2008). "Total synthesis of bryostatin 16 using atom-economical and chemoselective approaches." Nature **456**(7221): 485-8.

Tsai, Y. L. and B. H. Olson (1991). "Rapid method for direct extraction of DNA from soil and sediments." Appl Environ Microbiol **57**(4): 1070-4.

Tyson, G. W., J. Chapman, et al. (2004). "Community structure and metabolism through reconstruction of microbial genomes from the environment." Nature **428**(6978): 37-43.

Urlacher, V. B., S. Lutz-Wahl, et al. (2004). "Microbial P450 enzymes in biotechnology." Appl. Microbiol. Biotechnol. **64**(3): 317-25.

Val, D. L. and J. E. Cronan, Jr. (1998). "In vivo evidence that S-adenosylmethionine and fatty acid synthesis intermediates are the substrates for the LuxI family of autoinducer synthases." J Bacteriol **180**(10): 2644-51.

Van Wagoner, R. M. and J. Clardy (2006). "FeeM, an N-acyl amino acid synthase from an uncultured soil microbe: structure, mechanism, and acyl carrier protein binding." Structure **14**(9): 1425-35.

Venter, J. C., K. Remington, et al. (2004). "Environmental genome shotgun sequencing of the Sargasso Sea." Science **304**(5667): 66-74.

Viallard, V., I. Poirier, et al. (1998). "Burkholderia graminis sp. nov., a rhizospheric Burkholderia species, and reassessment of [Pseudomonas] phenazinium, [Pseudomonas] pyrrocinia and [Pseudomonas] glathei as Burkholderia." Int J Syst Bacteriol **48 Pt 2**: 549-63.

Villegas, A. and A. M. Kropinski (2008). "An analysis of initiation codon utilization in the Domain Bacteria - concerns about the quality of bacterial genome annotation." Microbiology **154**(Pt 9): 2559-661.

Voget, S., C. Leggewie, et al. (2003). "Prospecting for novel biocatalysts in a soil metagenome." Appl Environ Microbiol **69**(10): 6235-42.

Wang, C., D. J. Meek, et al. (2006). "Isolation of poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants." Appl Environ Microbiol **72**(1): 384-91.

Wang, G. Y., E. Graziani, et al. (2000). "Novel natural products from soil DNA libraries in a streptomycete host." Org. Lett. **2**(16): 2401-4.

Wang, M., W. Liu, et al. (1988). "Structure elucidation of heliclactone." Huaxue Xuebao **46**(8): 768-71.

Ward, N. L., J. F. Challacombe, et al. (2009). "Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils." Appl Environ Microbiol **75**(7): 2046-56.

Warren, R. L., J. D. Freeman, et al. (2008). "Transcription of foreign DNA in Escherichia coli." Genome Res **18**(11): 1798-805.

Watanabe, K., A. P. Praseuth, et al. (2007). "A comprehensive and engaging overview of the type III family of polyketide synthases." Curr. Opin. Chem. Biol. **11**(3): 279-86.

Waters, V. L. (2001). "Conjugation between bacterial and mammalian cells." Nat Genet **29**(4): 375-6.

Webster, G., L. Yarram, et al. (2007). "Distribution of candidate division JS1 and other Bacteria in tidal sediments of the German Wadden Sea using targeted 16S rRNA gene PCR-DGGE." FEMS Microbiol Ecol **62**(1): 78-89.

Weinstein, M., R. C. Roberts, et al. (1992). "A region of the broad-host-range plasmid RK2 causes stable in planta inheritance of plasmids in Rhizobium meliloti cells isolated from alfalfa root nodules." J. Bacteriol. **174**(22): 7486-9.

Wexler, M., P. L. Bond, et al. (2005). "A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase." Environ Microbiol **7**(12): 1917-26.

Wexler, M. and A. W. Johnston (2010). "Wide host-range cloning for functional metagenomics." Methods Mol Biol **668**: 77-96.

Wilkinson, D., T. Jeanicke, et al. (2002). "Efficient molecular cloning of environmental DNA from geothermal sediments." Biotechnology Letters **24**(2): 155-161.

Willecke, K. and A. B. Pardee (1971). "Fatty acid-requiring mutant of bacillus subtilis defective in branched chain alpha-keto acid dehydrogenase." J Biol Chem **246**(17): 5264-72.

Winson, M. K., M. Camara, et al. (1995). "Multiple N-acyl-L-homoserine lactone signal molecules regulate production of virulence determinants

and secondary metabolites in Pseudomonas aeruginosa." Proc Natl Acad Sci U S A **92**(20): 9427-31.

Wolf, Y. I., I. B. Rogozin, et al. (2001). "Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context." Genome Res **11**(3): 356-72.

Wolfgang, M. C., B. R. Kulasekara, et al. (2003). "Conservation of genome content and virulence determinants among clinical and environmental isolates of Pseudomonas aeruginosa." Proc Natl Acad Sci U S A **100**(14): 8484-9.

Wyatt, M. A., W. Wang, et al. (2010). "Staphylococcus aureus nonribosomal peptide secondary metabolites regulate virulence." Science **329**(5989): 294-6.

Zaehner, D. and F. P. Fiedler (1999). "Fifty years of antimicrobials: past perspectives and future trends in The Need for New Antibiotics: Possibly Ways Forward " 53rd Symposium of the Society for General Microbiology **98**: 67-84.

Zdobnov, E. M. and R. Apweiler (2001). "InterProScan--an integration platform for the signature-recognition methods in InterPro." Bioinformatics **17**(9): 847-8.

Zengler, K., G. Toledo, et al. (2002). "Cultivating the uncultured." Proc Natl Acad Sci U S A **99**(24): 15681-6.

Zhang, H., Y. Wang, et al. (2008). "Bacterial hosts for natural product production." Mol Pharm **5**(2): 212-25.

Zheng, Y., B. P. Anton, et al. (2005). "Phylogenetic detection of conserved gene clusters in microbial genomes." BMC Bioinformatics **6**: 243.

Zhou, J., M. A. Bruns, et al. (1996). "DNA recovery from soils of diverse composition." Appl Environ Microbiol **62**(2): 316-22.

Zverev, V. V. and I. A. Khmel (1985). "The nucleotide sequences of the replication origins of plasmids ColA and ColD." Plasmid **14**(3): 192-9.