

POLYKETIDE SYNTHASE PATHWAY DISCOVERY FROM
SOIL METAGENOMIC LIBRARIES

Except where reference is made to the work of others, the work described in this thesis is my own or was done in collaboration with my advisory committee. This thesis does not include proprietary or classified information.

Ann Marie Goode

Certificate of Approval:

Paul Cobine
Assistant Professor
Biological Sciences

Mark R. Liles, Chair
Assistant Professor
Biological Sciences

Laura S. Suh
Assistant Professor
Biological Sciences

George T. Flowers
Dean
Graduate School

POLYKETIDE SYNTHASE PATHWAY DISCOVERY FROM
SOIL METAGENOMIC LIBRARIES

Ann Marie Goode

A Thesis

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Master of Science

Auburn, Alabama
August 10, 2009

POLYKETIDE SYNTHASE PATHWAY DISCOVERY FROM
SOIL METAGENOMIC LIBRARIES

Ann Marie Goode

Permission is granted to Auburn University to make copies of this thesis at its discretion, upon request of individuals or institutions and at their expense.
The author reserves all publication rights.

Signature of Author

Date of Graduation

VITA

Ann Marie Goode, daughter of Richard and Theresea Goode, was born August 8, 1982, in Decatur, Alabama. She was an Honor's graduate from Decatur High School in 2001. She attended Lipscomb University in Nashville, Tennessee, and graduated in 2005 with a Bachelor of Science degree in Biochemistry and a minor in Biology. In August, 2006, she entered the Department of Biological Sciences at Auburn University to pursue a Master of Science in Microbiology where she was a graduate teaching assistant and graduate research assistant.

THESIS ABSTRACT

POLYKETIDE SYNTHASE PATHWAY DISCOVERY FROM
SOIL METAGENOMIC LIBRARIES

Ann Marie Goode

Master of Science, August 10, 2009
(B.S., Lipscomb University, 2005)

91 Typed Pages

Directed by Mark R. Liles

Polyketides are structurally diverse bacterial secondary metabolites, many of which have antibiotic or anti-cancer activity. Modular polyketide synthase (PKS) enzymatic complexes contain conserved ketoacyl synthase (KS) domains, and most PKS biosynthetic pathways exceed 30 kb in size. A fosmid metagenomic library constructed from soil at the Hancock Agricultural Research Station in Hancock, WI, (18,432 clones, average insert of 42 kb) was spotted onto a nylon membrane. The macroarray was screened using a degenerate DNA probe targeting the KS domain. Thirty-four clones containing KS domains were identified by Southern hybridization; however, only 21 of the 34 PKS-positive clones produced a PCR product. Interestingly, most of the clones (8 out of 13) that were PCR-negative were nonetheless KS-positive by Southern blot hybridization. DNA sequences from a KS-containing clone that was consistently PCR-negative have revealed a biosynthetic pathway that is divergent from known pathways,

and is hypothesized to have an origin from the newly described bacterial division Acidobacteria that is prevalent within soils yet has few cultured representatives.

ACKNOWLEDGMENTS

I would like to thank Dr. Mark Liles for his professional guidance, patience, and scientific knowledge, my committee members, Dr. Laura Suh and Dr. Paul Cobine for their direction and support while I was conducting research and through the writing process. I would like to acknowledge Larissa Parsley for her invaluable opinions and editing skills. I also sincerely thank Nancy Capps, Erin Consuegra, Molli Newman, and Kavita Kakirde for their assistance. A special thank you goes to Dr. Al Brown, Dr. Linda Phipps, Dr. Kent Clinger, and Eric Lockert for their academic and professional advice. Finally, I graciously thank my parents, Richard and Theresea Goode, and my grandmother, Geraldine Fromhold, for their endless love and encouragement, without which I could not have made it to this point.

Style manual or journal used: Journal of Microbiology

Computer software used: Microsoft Word 2000, EndNote X1, ChromasPro 1.4.1.0,
Sequencher 4.1.4, Mauve 2.3.0, SWAAP 1.0.3

TABLE OF CONTENTS

| | |
|---|-----|
| LIST OF TABLES..... | xi |
| LIST OF FIGURES..... | xii |
| CHAPTER 1: LITERATURE REVIEW..... | 1 |
| ANTIBIOTICS..... | 2 |
| METAGENOMICS..... | 5 |
| POLYKETIDE SYNTHASE..... | 13 |
| SUMMARY..... | 21 |
| CHAPTER 2: IDENTIFICATION OF FOSMID CLONES CONTAINING A TYPE I POLYKETIDE SYNTHASE PATHWAY FROM A SOIL METAGENOMIC LIBRARY..... | 22 |
| INTRODUCTION..... | 23 |
| MATERIALS AND METHODS..... | 24 |
| RESULTS..... | 29 |
| DISCUSSION..... | 35 |
| CHAPTER 3: CHARACTERIZATION OF CLONES EXPRESSING ANTIMICROBIAL ACTIVITY..... | 38 |
| INTRODUCTION..... | 39 |
| MATERIALS AND METHODS..... | 40 |
| RESULTS..... | 42 |
| DISCUSSION..... | 44 |
| FUTURE DIRECTIONS..... | 46 |

| | |
|-----------------------|----|
| CONCLUSIONS..... | 48 |
| LITERATURE CITED..... | 49 |

LIST OF TABLES

| | |
|---|----|
| 2.1 PCR primers for design of KS probe..... | 59 |
| 3.1 Activity of BAC clones tested against <i>Alcaligenes faecalis</i> and <i>Pseudomonas</i> <i>aeruginosa</i> | 60 |

LIST OF FIGURES

| | |
|---|----|
| 1.1 Modular arrangement of type I PKS pathways..... | 61 |
| 2.1 Electrophoresis of multiple primer sets..... | 62 |
| 2.2 Detection of type I KS-domain positive fosmid clones..... | 63 |
| 2.3 Maximum parsimony tree of KS domains recovered from fosmid clones..... | 64 |
| 2.4 Restriction Fragment Length Polymorphism analysis of fosmid clones..... | 65 |
| 2.5 Southern Blot analysis of fosmid clones negative by PCR..... | 66 |
| 2.6 Contig derived from fosmid 454 sequencing with homology to a PKS sequence..... | 67 |
| 2.7 Identification of fosmid clone A12 containing a PKS pathway sequence..... | 68 |
| 2.8 Alignment of <i>Solibacter usitatus</i> genome with A12 contig sequences..... | 69 |
| 2.9 BLASTx analysis of <i>Solibacter usitatus</i> genome with A12 contig sequences..... | 70 |
| 2.10 Similarity and G + C content of A12 contig sequences..... | 73 |
| 3.1 Growth inhibition of <i>Alcaligenes faecalis</i> | 79 |

CHAPTER 1

LITERATURE REVIEW

ANTIBIOTICS

Alexander Fleming made a serendipitous discovery in the early 1900s that would forever change the fields of medicine and microbiology. By discovering that the mold *Penicillium* produced a substance that was lethal to bacteria, Fleming unknowingly opened a door to the broad world of antibiotics. In 1945 Selman Waksman, a Ukrainian biochemist and microbiologist, originally proposed that an antibiotic was “a chemical substance of microbial origin that possesses antibiotic powers” (Waksman 1961). This concept has been expanded over time to include plant and animal products, as well as synthetic and semisynthetic compounds used to inhibit the growth of microorganisms (Yim et al. 2006).

Today, antibiotics are generally considered to be low-molecular weight secondary metabolites that are synthesized by many diverse microorganisms, and at low concentrations will inhibit the growth of other microorganisms (Lancini et al. 1993). They are distinguishable from primary metabolites in that they are not products of cellular metabolism, but are rather products of biosynthetic pathways that may require many separate enzymatic activities, requiring the coordinate regulation of multiple genetic loci. Antibiotics may contribute to the fitness of bacterial populations by their antagonistic effects on competing microbial cells. While certain phylogenetic groups (e.g., *Streptomyces* spp.) are well-known producers of antibiotic compounds, secondary metabolites with the ability to inhibit the growth and/or viability of other microorganisms are known to be produced by every major microbial phylum (Lancini et al. 1993).

Though our knowledge and understanding of antibiotics are still developing, antibiotics remain the primary recourse for treatment of bacterial infections (Davies 2006). It is now known that bacteria have the ability in their natural environments to exchange genetic information and confer genetic antibiotic resistance to phylogenetically diverse recipients. The emergence of multidrug-resistant bacteria is a phenomenon that is of serious concern to the medical and pharmaceutical industries. This developing resistance among specific pathogens is also the major cause of failure in the treatment of infectious disease (Davies 1994). While more than 350 antimicrobial agents are available today, the need for novel antibiotic compounds is unmet by efforts to discover novel antimicrobial compounds, particularly by the pharmaceutical industry. Despite the \$10 billion annual market for antimicrobial compounds, the advances that have been made in combinatorial chemistry and biosynthesis, metabolic pathway engineering, and gene rearrangement combined with an antibiotic rediscovery rate exceeding 90%, diminishing financial returns for novel antibiotics, and the high costs associated with clinical trials have caused major drug companies to withdraw from investments in antibiotic discovery (Davies 1994; Bull et al. 2000; Staunton et al. 2001; Newman et al. 2003; Demain et al. 2005).

Most pharmaceutical companies have shifted their interest and investment into the drug markets for depression, heartburn, and erectile dysfunction (Zaehner et al. 1995; Demain et al. 2005). Even in these target areas, natural products (i.e., bioactive compounds synthesized from a living entity) have continued to be an excellent resource for drug discovery. More than 60% of approved and new drug application candidates are

either natural products or are related to them, making natural products the most important anti-infective agents (Demain et al. 2005). These natural products have structural novelty and complexity in the form of chirality centers, rings, bridges, and functional groups, which make them far superior to synthetic compounds (Bull et al. 2000).

Traditionally, new antibiotics are discovered by the cultivation of microorganisms from the environment, with soil being the most abundant source of microorganisms producing antibiotics that are in use today (Zaehner et al. 1995). Most soil microorganisms have evolved as oligotrophs, and do not readily grow on laboratory media that are rich in nutrients.

Several modern cultivation techniques have been employed to exploit antimicrobial compounds produced by soil organisms. One of these methods included the use of a diffusion chamber that allowed growth of previously uncultured microorganisms from marine sediment in a simulated natural environment. Organisms were placed in diffusion chambers and incubated the chambers in an aquarium that simulated these organisms' natural environment. The membranes allow exchange of chemicals between the chamber and the environment but restrict movement of cells. This diffusion chamber method bypasses the limitations associated with culturing microorganisms in an unnatural environment (Kaeberlein et al. 2002).

In addition to diffusion chambers, a method used to cultivate previously uncultured microorganisms developed includes the combination of encapsulation of cells in gel microdroplets followed by flow cytometry to detect microcolonies within the droplets. Soil samples were diluted with appropriate buffer and mixed with agarose then

emulsified. Flow cytometry was used to determine exact cell numbers within the microdroplet. This technology allows for microbial cultivation under low nutrient conditions (Zengler et al. 2002).

Even with these advances in culturing the previously uncultured soil microflora, there are limitations to cultivation-dependent methods that at best may achieve cultivation of approximately 10-20% of the extant soil microbial taxa (Hugenholtz et al. 1998). Advances in microbiology have risen to the challenge of antibiotic discovery by, for example, developing culture-independent methods for natural product discovery (Molnar et al. 2000).

METAGENOMICS

It is generally accepted that all small molecules produced by microbes have specific biological functions. Genome sequencing of *Streptomyces* showed the genetic capacity to produce more than 25 different small molecules; however, the majority of compounds were not detected under laboratory conditions (Nystrom 2004; Curtis et al. 2005). This limited recovery is usually attributed to the foreign environment of nutrient-rich liquid medium with constant oxygen supply and temperature (Demain et al. 1999) that exists in the laboratory but not necessarily in the environment. Soil environments are generally nutritionally depleted, reducing the generation time of the microbe 50-100 times. Under such unnatural laboratory conditions cellular metabolism is disturbed, causing normal functions to be over- or under-expressed, respectively (Davies 2006).

For over one hundred and fifty years, the ability to cultivate microorganisms has allowed us to better understand the microbial world around us. In every natural environment explored by the use of cultivation and microscopy, a consistent disparity exists between the large number of microorganisms observed by microscopy (typically $>10^9$ cells per g of soil) and the much lower ($\sim 10^6$ cells per g of soil) recovered by laboratory cultivation, which has been described as the “Great Plate Count Anomaly” (Staley et al. 1985). Many of the microorganisms in soil may be physiologically dormant, and exist in a viable but not culturable (VBNC) state. Other microorganisms may not be cultured in the laboratory due to co-metabolic requirements unmet by laboratory media. The question of whether the vast majority of microorganisms that did not grow on laboratory media represented unique phyla, or simply previously described

microbial phyla in a VBNC state, was answered through the application of molecular phylogenetics. The use of 16S rRNA gene sequences (and other phylogenetically informative genes) that was pioneered by Dr. Carl Woese and colleagues has revealed many novel bacterial divisions, expanding the recognized bacterial divisions to over 40 monophyletic groups (Woese et al. 1990; Hugenholtz et al. 1998)

Metagenomics, also known as “environmental genomics” (Stein et al. 1996; Hallam et al. 2004) or “community genomics” (Allen et al. 2005; Tyson et al. 2005) is the collective genomic analysis of a group of organisms (Handelsman et al. 2002). Metagenomics is a method of analysis that utilizes direct extraction and cloning of DNA from an assemblage of microorganisms, allowing scientists to overcome the hurdles involved in culture-based methods (Handelsman 2004). Libraries containing DNA extracted directly from the environment can be used to provide phylogenetic, functional, and genomic sequence information. Metagenomic libraries have been constructed from a variety of environmental sources including seawater (Stein et al. 1996; Cottrell et al. 1999; Beja et al. 2000; Beja et al. 2000), soil (Henne et al. 1999; Henne et al. 2000; Rondon et al. 2000; Entcheva et al. 2001), and marine sponges (Schleper et al. 1998), and recent efforts have been devoted to also exploring the human commensal microflora using a culture-independent approach (Gill et al. 2006).

Metagenomic libraries may be screened by sequence-based or functional methods to identify recombinant clones of interest. When a metagenomic library is screened for a specific gene target, conserved phylogenetic markers such as 16S rRNA or *recA* genes can provide insight into the origin of the DNA and identify additional genes from this

microorganisms contained within the same genome fragment (Schleper et al. 1997; Eisen 1998; Vergin et al. 1998; Sandler et al. 1999; Beja et al. 2000; Rondon et al. 2000; Beja et al. 2002; Beja et al. 2002). One example of this was a study that identified four fosmid clones from *Pirellula* and Planctomycetales taxa by identifying 16S rRNA-containing recombinant clones and their associated functional genes (Vergin et al. 1998). This approach of linking phylogenetic and functional information from large-insert metagenomic clones was one way in which functions of previously uncultivated microorganisms could be predicted, yet a significant limitation of this methodology is that the functional genes identified from any targeted prokaryotic taxa are dependent upon being adjacent to a recognized phylogenetic marker.

Novel genes and gene products discovered using sequence-based screening of metagenomic libraries include the first bacteriorhodopsin of bacterial origin (Beja et al. 2000), and new members of families of known proteins, such as an Na⁽⁺⁾(Li⁽⁺⁾)/H⁽⁺⁾ antiporters, RecA proteins, DNA polymerases, and antibiotic resistance determinants (Handelsman 2004). With the advent of new sequencing technologies, such as pyrosequencing by 454 Life Sciences (Roche, Inc.), deep sequencing of microbial communities has allowed a more complete sequence-based assessment of microbial community genomic diversity. Reassembly of multiple bacterial genomes using a metagenomic approach has provided insight into energy and nutrient cycling within the community, genome structure, gene function, population genetics, and lateral gene transfer among members of uncultured bacterial communities (Handelsman 2004).

More recently, other methods, including the cultivation of previously uncultured taxa (Kaeberlein et al. 2002), genetic re-engineering of existing pathways (Pfeifer et al. 2001), and direct cloning and expression of metagenomic DNA from natural environments (Rondon et al. 2000) have been used to identify bioactive compounds produced by soil microorganisms. Isolation of high molecular weight DNA to construct a large-insert, metagenomic library is particularly advantageous because genes for biosynthetic pathways are typically clustered, making it feasible to clone entire pathways. Furthermore, natural products that are potentially harmful to prokaryotes are often linked to genes for resistance to the natural product, so that the heterologous bacterial host (e.g., *E. coli*) in which the pathway is expressed may potentially be protected (Chater et al. 1985; Distler et al. 1987; Handelsman et al. 1998).

There have been significant efforts to identify novel PKS pathways contained within metagenomic libraries using a metagenomic approach. To access potentially expressed PKS pathways derived from *Streptomyces* spp., one study constructed a soil metagenomic library in an *Escherichia coli*-*Streptomyces lividans* shuttle cosmid vector (pOS700I) and screened recombinant clones for phylogenetic content and PKS pathways (Courtois et al. 2003). The phylogenetic content of the DNA library was extremely diverse, containing various microorganisms that have not been previously cultured. The library was screened by PCR for sequences similar to parts of type I PKS genes using two different sets of PKS-specific primers and subsequently tested for the expression of new molecules by screening of live colonies and cell extracts. The results revealed new PKS genes in at least eight clones. In addition, five additional clones were confirmed to

express PKS pathways and generate polyketide products revealed by high-pressure liquid chromatography analysis and/or from their biological activity (Courtois et al. 2003).

Schirmer et al. also explored PKS domains by PCR amplification of ketosynthase domains of type I modular PKS from the microbial community of the marine sponge, *Discodermia dissolute*, revealing great diversity and a new group of sponge-specific PKS ketosynthase domains. Metagenomic libraries totaling more than 4 Gbp of bacterial genomes were screened for type I modular PKS gene clusters, and 0.7% contained PKS genes. Most of the PKS clones carried small PKS clusters of one to three modules, although some clones encoded large multimodular PKSs (more than five modules). The most abundant large modular PKS appeared to be encoded by a bacterial symbiont that comprised < 1% of the sponge community. Sequencing of this PKS revealed 14 modules that, if expressed and active, are predicted to produce a multimethyl-branched fatty acid resembling that of mycobacterial lipid components (Schirmer et al. 2005).

To further access PKS gene diversity from soil, Wawrick and colleagues developed degenerate PCR primers specific to actinomycete type II ketosynthase genes. Twenty-one soil samples were collected from diverse sources and their bacterial PCR products were generated using bacterial 16S rRNA gene primers (27F and 1525R), as well as an actinomycete-specific forward primer. PCR products were cloned, and seven novel clades of KS genes were identified. The nucleotide sequences were between 74% and 81% identical to known sequences in GenBank. One cluster of sequences was most similar to the KS domain involved in ardacin, a glycopeptide antibiotic, produced by *Kibdelosporangium aridum*, an Actinobacteria. The remaining sequences showed greatest

similarity to the KS genes in pathways producing the angucycline-derived antibiotics simocyclinone, pradimicin, and jasomycin (Wawrik et al. 2005).

Other academic and industrial research groups have identified PKS pathways from environmental sources using a culture-independent methodology. Numerous PKS domain sequences have been recovered from soil microbial DNA, and these PKS domain sequences suggest diverse and novel structures for their polyketide products (Courtois et al. 2003; Ginolhac et al. 2004; Wawrik et al. 2005; Parsley et al. 2006). Other investigators have isolated intact PKS pathways, including a study that identified and heterologously expressed PKS genes using a *Streptomyces* shuttle vector (McDaniel et al. 1993). PKS pathways have also been isolated by macroarray hybridization, the first and only study to publish on this was a limited study that generated a small fosmid library from east China Sea marine sediment, screened 500 fosmid clones using a single KS domain as a probe, and demonstrated that one fosmid clone contained a functional PKS domain (Jiao et al. 2008). In addition to these efforts, Kosan Biosciences is investigating the use of gene manipulation to alter or enhance microbial PKS genes in order to increase potency, correct limitations, and potentially improve large-scale production (Hutchinson 2005).

To get an estimate of the percentage of cultured soil bacteria that contain the biosynthetic capacity for PKS expression, Khatun et al. screened cultured bacterial isolates from different soil types for PKS genes via PCR amplification of Type I PKS genes with a degenerate oligonucleotide primer set (Khatun et al. 2002). It was found that the percent of bacteria testing PKS-positive ranged from 1.4% to 18.5%, depending

upon the soil type (a clay-rich soil gave the highest percentage). These data can be used to provide a very rough estimate, given certain assumptions, of the percentage of recombinant clones that may contain a PKS pathway using a culture-independent approach. If one assumes that cultured soil bacteria have the same relative frequency of encoding PKS pathways as do as-yet-uncultured bacteria, and that the average bacterial genome size is 4×10^6 bp (~ 1 *E. coli* genome equivalent), and that the bacterial taxa represented within the metagenomic library will have an approximately equal relative abundance, it would be predicted that the frequency of detecting soil metagenomic library fosmid clones (with an average insert size of 42 kb) that contain a PKS pathway using the PCR method would range between 1 in every 515 clones (at a frequency of 18.5% of bacterial genomes that are PKS-positive) to 1 out of every 6,802 clones (at a frequency of 1.4% of bacterial genomes that are PKS-positive). From a fosmid soil metagenomic library with 9,984 clones (42 kb avg. insert size), it would therefore be predicted that the number of PKS-positive clones discovered would be somewhere in the range of 1 clone (1.4% PKS frequency) to 18 clones, depending upon the soil type used and a number of potentially misleading assumptions.

Previous methods of detecting KS domains were likely biased due to the use of degenerate primer sets that were developed based on the existing database of KS domains from cultured microorganisms. Since soil metagenomic libraries contain bacterial genome fragments from diverse bacterial species, many of which represent as-yet-uncultured bacterial phyla (Handelsman et al. 1998), there may be many PKS pathways that would be undetected by using a primer set developed solely from PKS pathways

studied in cultured microorganisms; therefore, in this study, PKS pathways contained within large-insert clones were identified by medium-stringency Southern blot hybridization with a heterogenous KS domain probe. By identifying previously unknown PKS biosynthetic pathways, this study will contribute to our understanding of PKS gene diversity in environmental microorganisms. Future studies will build upon this work to explore the heterologous expression of these cloned PKS pathways and the natural product chemistry that these pathways may express in their native or heterologous host.

POLYKETIDE SYNTHASES

Microbial metabolites have long been a primary source of therapeutically important drugs. One promising lead for new drug therapy is the development of polyketides. Polyketides are a diverse family of natural products found in bacteria, fungi, and plants (Shen 2003); included in this family are the antibiotics erythromycin, tylosin, rifamycin, tetracyclines, immunosuppressants FK506 and rapamycin, and antitumor compounds doxorubicin and mithramycin (McDaniel et al. 2005).

Polyketides were accidentally discovered in 1893 at London University by James Collie, who was attempting to determine the structure of dehydroacetic acid by boiling it with barium hydroxide and acid. He was surprised to find orcinol, an aromatic compound, as one of the products. Collie further hypothesized the mechanism of orcinol formation, proposing that a polyketone intermediate could be formed from the γ -pyrone of dehydroacetic acid by addition of water and ring opening. Loss of water would lead to the formation of the orcinol ring structure. Collie went on to hypothesize that these polyketone intermediates might be generated and produced by living cells. Unfortunately, other scientists of the day lacked Collie's insight and did not agree with his ideas. In 1917, Robert Robinson's findings re-affirmed Collie's original belief that polyphenols are produced from polyketones. Arthur Birch revived the scientific community's interest in polyketides in the 1950s by experimentally proving that polyketones could be generated from acetate units by repeated condensation reactions, which later became known as the Collie-Birch polyketide hypothesis (Birch et al. 1955). The Collie-Birch theory holds that polyketide biosynthesis is similar to fatty acid synthesis.

Both fatty acids and polyketides are assembled from acetyl-coenzyme A, or related acyl-coenzymes, in a series of repeated a head-to-tail linkage until the desired chain length is reached (Hutchinson et al. 1995; Staunton et al. 2001). Experiments with isotope labeling further confirm the Collie-Birch theory proving that the mechanisms of polyketide synthesis fall within the four types of reactions used to make fatty acids (O'Hagan 1991). Although the biosynthetic pathways are similar, there is one critical difference between fatty acid synthesis and polyketide synthase synthesis. Fatty acid biosynthetic pathways reduce and dehydrate each resulting β -keto carbon, producing an inert hydrocarbon. Polyketide systems modify or omit some reactions, preserving some chemical reactivity along portions of the polyketide chain. Still other polyketide enzymes selectively promote internal cyclization and π -bond rearrangements to produce a diverse amount of products (Austin et al. 2003).

Polyketide synthases can be divided into three groups: type I, type II, and type III. Type I contain multifunctional proteins that are each responsible for a different active site for catalytic reactions in polyketide chain assembly (Moore et al. 2001). The actinorhodins, or type II PKSs, are composed of individual proteins that perform one enzymatic activity to catalyze the formation of aromatic polyketides (Moore et al. 2001; Shen 2003). Type III PKSs, previously thought to be found only in plants, participate in the assembly of small aromatic compounds (Moore et al. 2001).

Type I PKSs are large multifunctional proteins organized into modules, each of which contains an individual enzyme responsible for the catalysis of polyketide chain elongation. Modular PKS are demonstrated by 6-deoxyerythromycin B synthase (DEBS)

for the biosynthesis of reduced polyketides such as erythromycin A (Shen 2003), seen in figure 1.1 (Staunton et al. 2001). Three open reading frames (ORFs), *eryAI*, *eryAII*, *eryAIII*, make up the structural genes responsible for the synthesis of multienzyme polypeptides 1, 2, and 3, respectively. Each ORF is further organized into two modules. Each individual module contains three domains to catalyze one cycle of chain extension, ketosynthase (KS), acyltransferase (AT), and acyl carrier protein (ACP). Some modules carry variable domains, such as ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER), which are responsible for keto group modification (Donadio et al. 1991; Bevitt et al. 1992). DEBS 1 contains a loading domain that accepts the propionate from propionyl-CoA, while DEBS 3 terminates with a thioesterase that catalyzes the cyclization of the 6-dEB (Staunton et al. 2001).

The manipulation and development of PKS pathways offers vast opportunities for the production of novel compounds. The modular organization of type I PKSs present diverse possibilities to determine if new metabolites can be found from different combinations of the modules, using a combination of genes from the same or different organisms. If these rearrangements are successful, this method would provide another source of chemical diversity in the quest for new drugs (Hutchinson et al. 1995). Research is ongoing to mutate the *eryAI*, *II*, and *III* genes to produce potentially new metabolites, but so far, attempts to find these novel metabolites have been unsuccessful (Hutchinson et al. 1995).

The structure of type II PKS is comprised of enzymes similar to Type I: KS, ACP, chain length factor (CLF) and cyclases (McDaniel et al. 2005). These three subunits

constitute the “minimal PKS” found in all type II PKSs (Austin et al. 2003). The CLF controls the chain length while the ACP/*act* I assembles the polyketone chain from one unit of acetyl-CoA and seven units of chain-extending malonyl-CoA. The *actIII* and cyclases act in cyclization and aromatization (Staunton et al. 2001; McDaniel et al. 2005). More specifically, after decarboxylation of malonyl-ACP, the acetyl group is transferred to the KS α active site where the first condensation step will occur, resulting in acetoacetyl-ACP. The ketoester is transferred from the ACP back to KS α where another cycle of chain extension can occur (Staunton et al. 2001).

Type I and type II PKSs differ in the formation of cyclic aromatic compounds that do not require the dehydration and reduction steps required in modular synthesis. Type II PKSs were actually discovered several years before type I PKSs; however, a complete set of type II PKS enzymes has not been found (Hutchinson et al. 1995). The first genes encoding type II PKSs were discovered by studying the genetics of actinorhodin biosynthesis in *Streptomyces coelicolor* and tetracenomycin biosynthesis in *Streptomyces glaucescens* (Malpartida et al. 1984; Motamedi et al. 1987). *S. coelicolor* produces a blue-pigmented polyketide, actinorhodin (Wakil 1989). For actinorhodin biosynthesis, an aromatic polyketide gene cluster encodes the PKS genes involved in cyclization, aromatization, and tailoring enzymes. Further sequence analysis of the PKS-encoding DNA encoding granaticin, tetracenomycin, and actinorhodin genes in *Streptomyces violaceoruber* was used to confirm the basic design of type II PKS (Summers et al. 1992; Summers et al. 1993)

Previous studies involving type II PKS pathways have revealed new products from new combinations of heterologous bacterial genes. The exact function of the *act I*, *act IV*, and *act VII* is yet to be determined and could provide more insight into the potential for novel PKS metabolites.

Type III PKSs are small homodimeric proteins that are similar to chalcone synthase (CHS) and stilbene synthase (STS) families that were thought to be found only in plants (Schroder 1999; Shen 2003). Their sequence homology and vague evolutionary history led some to believe that a number of bacteria had acquired the CHS-like genes via horizontal gene transfer from plants. However, the recent boom in genome sequences has revealed only about 25% amino acid homology with CHS, confirming that type III PKSs are indeed a novel system found in bacteria (Austin et al. 2003).

Type III PKSs differ from CHS and STS families in their starter molecules, the number of acetyl additions they catalyze, and their mechanism of chain termination (Austin et al. 2003). This type of PKS generally uses a malonyl-CoA starter unit, continues through malonyl-CoA decarboxylation, polyketide chain elongation, and cyclization, yielding a chalcone or stilbene. There are currently 12 known type III PKS pathways attributed to plants and three from bacteria. (Austin et al. 2003)

The first demonstration of the enzymatic production of an aromatic PKS product by a type III polyketide synthase from a non-plant was demonstrated from the Chinese club moss, *Huperzia serrata*. These findings imply that type III PKS enzymes accept a larger starter substrate and possibly contain a larger starter-substrate binding site at the active site than chalcone synthase pathways found in plants (Wanibuchi et al. 2007).

Furthermore, the type III enzyme RppA offers broad substrate specificity allows for diversity in final end products, leading some to believe that RppA-like enzymes are involved in the biosynthesis of a wide variety of metabolites in various different hosts (Funa et al. 2002).

Currently, two-thirds of known bioactive polyketide natural products originate from actinomycetes. Other microbial sources of polyketides include myxobacteria, cyanobacteria and fungi (Pfeifer et al. 2001). While the exact function of many polyketides is unknown, it is thought that in many cases they act as a secondary metabolite to out-compete other microbes in times of nutrient limitation. Polyketides have several distinguishing characteristics that set them apart from other enzymatic pathways. They have a large size, usually 100 to 10,000 kDa, and are soluble cytosolic multienzyme systems that require no eukaryotic intracellular substructure or organelle to maintain activity. The possibilities of joining various purified protein components *in vitro* to yield novel PKS activity makes these pathways ideal for expression in heterologous hosts (Pfeifer et al. 2001).

Heterologous expression can be employed to produce a desired protein in specific quantities and can be used to overcome limitations associated with manipulating the larger modular arrangement of PKS in native microorganisms. *Streptomyces coelicolor* and *Escherichia coli* have been engineered in such a way to be able to express PKS pathways that might otherwise not be characterized.

The macrocyclic core of the antibiotic erythromycin, 6-deoxyerythronolide B (6dEB), is a complex natural product produced by the soil bacterium *Saccharopolyspora*

erythraea through a PKS reaction. A derivative of *Escherichia coli* has been genetically engineered to convert exogenous propionate into 6dEB.

6-deoxyerythronolide B synthase 3 (dEB 3) is proposed to catalyze the fifth and sixth condensation cycles in the assembly of the polyketide 6-dEB, the first intermediate in the biosynthesis of erythromycin A. The gene encoding dEB 3 has been engineered into a pT7-based expression system for over-expression in *Escherichia coli*. (Roberts et al. 1993).

As organisms evolve and become more resistant to existing drug therapy, the need for innovative treatment will become paramount. The cornerstones of this work have been and will continue to be the known type I, II, and III pathways. Developments in cloning biosynthetic gene clusters and advancements in technology for DNA sequencing and bioinformatics have provided many new avenues to search for unique biosynthetic pathways (Shen 2003). The field of PKS research is a rapidly developing field of scientific study that will surely continue to attract the attention of scientists and the medical community.

SUMMARY

Evolutionary processes have resulted in an incredible diversity of chemical moieties with antibiotic activity. To access a greater degree of the biosynthetic capacity of environmental microorganisms the following studies used a culture-independent methodology for antibiotic discovery, combining recent advances in the isolation and cloning of high molecular weight genomic DNA from diverse soil microorganisms with both sequence-based and function-based screening of soil community genomic “metagenomic” libraries to identify recombinant clones either encoding or expressing the biosynthetic machinery for microbial natural products.

CHAPTER 2

IDENTIFICATION OF FOSMID CLONES CONTAINING A TYPE I

POLYKETIDE SYNTHASE PATHWAY FROM A SOIL METAGENOMIC

LIBRARY

INTRODUCTION

Isolation of secondary metabolites produced by soil microorganisms has historically been an effective means of identifying novel antibiotics. Soil environments contain a staggering $\sim 10^{16}$ microbial cells per ton with only 1% being cultured in a laboratory on, for example, soil extract agar medium (Hugenholtz et al. 1998; Curtis et al. 2005). While the small percentage of soil microorganisms cultured in the laboratory are still phylogenetically diverse and have provided many important antibiotics, a potentially richer resource for antibiotic discovery lies within as-yet-uncultured bacteria (Gillespie et al. 2002). In this study, culture-independent methods are used to access novel biosynthetic pathways of potential medical importance from diverse soil microorganisms. Direct cloning of genomic DNA isolated from a complex microbial assemblage, also referred to as a metagenomic approach, allows for isolation of recombinant clones which may include partial or complete antibiotic biosynthetic pathways. This approach bypasses the limitations associated with traditional cultivation methods, but also has its own unique set of biases (e.g., cloning, heterologous expression).

MATERIALS AND METHODS

(The following 2 methods were performed previously.)

Soil collection and DNA isolation. Soil samples were collected from soil cores (10 cm to 50 cm depth) from an agricultural field at the University of Wisconsin-Madison's Hancock Agricultural Research Station in Hancock, WI. The soil was determined to be a sandy loam by scientists at the research station. Bacterial cells were collected by soil homogenization and differential centrifugation. Cells were then embedded and lysed within an agarose plug, and high molecular weight metagenomic DNA was recovered by electroelution (Liles et al. 2004).

Fosmid library construction. The library was constructed according to the CopyControl™ fosmid library production kit according to manufacturer's protocols (www.epicentre.com, Madison, WI. Purified genomic DNA was randomly sheared and end repaired. DNA of optimal size was isolated and ligated into the CopyControl™ vector. DNA was then transformed into TransforMax EPI300 *E. coli* and plated onto LB + 12.5 µg/ml chloramphenicol media. In EPI300 *E. coli* cells, arabinose is used to control the copy number of fosmid DNA by targeting the pEpiFOS copy control vector. The *E. coli* clones arrayed onto membranes at Clemson University's Genomics Institute (www.genome.clemson.edu) were grown with and without arabinose induction of copy number. In the absence of inducers, the P_{Bad} promoter is turned off and the *trfA* gene of the EPI300 (*E. coli*) cells is repressed. The clones are controlled by the *E. coli* F replicon and grown at single copy to maintain stability. Individual clones were chosen to be grown in culture. CopyControl Induction Solution (Arabinose) was added to induce the P_{Bad}

promoter, turning on the *trfA* gene. *trfA* initiates the *oriV* to induce high copy number of clones. Macroarrays were generated with and without arabinose-induction of fosmid copy number.

Screening of fosmid macrarray. The individual clones were grown in 96-well format and subsequently pooled for preparation of a fosmid library template for downstream PCR reactions. Specific domains from PKS pathways from the pooled DNA was PCR amplified using the primer set 5LL (GGR TCN CCI ARY TGI GTI CCI GTI CCR TGI GC) and 4UU (MGI GAR GCI YTI CAR ATG GAY CCI CAR CAR MG)(David Sherman, University of Michigan, personal communication), as well as other Type I and Type II PKS-specific and non-ribosomal peptide synthetase (NRPS)-specific primer sets. A digoxigenin (DIG)-labeled KS probe from pooled fosmid DNA was generated for macroarray hybridization. The hybridization was performed using the DIG detection system (Roche; Indianapolis, IN), with detection mediated by CSPD chemiluminescent substrate. Results were obtained after visualizing the nylon membrane after overnight incubation.

PCR amplification and sequencing of KS domains. The 34 clones that showed chemiluminescence (in duplicate) during hybridization were used as DNA template for verification that these clones contained a KS domain using the 4UU/5LL primer set. Because the 4UU/5LL primers did not work well for sequencing purposes, a nested PCR method was used by modifying the 4UU/5LL primers with T7 (4UU) and T3 (5LL) recognition sequences added to the 5' end of the respective primer sequence. PCR products were gel extracted and submitted for BigDye sequencing using either the T7 or T3 primer, respectively.

KS domain phylogenetic analysis. KS genes from KS-positive clones and known KS gene sequences from bacterial and fungal sources were aligned using ClustalX software. A maximum parsimony phylogenetic tree was generated using the MEGA4 (Tempe, AZ) software, using *E. coli* FabB sequence as a root, with 1000 iterations of the maximum parsimony analysis performed for calculation of bootstrap support.

Southern hybridization. A Southern blot was performed on the 13 clones that were positive by hybridization but negative by PCR. Fosmid clone DNAs were restriction digested with BamHI and electrophoresed on a 1% agarose gel. Clones were transferred to a nylon membrane (Whatman 2007) using the Whatman TurboBlotter Kit (Kent ME14 2LE ,UK) and hybridized with a heterogenous KS domain probe generated from amplification of all of the KS-positive fosmid clones (n=21) using primer set 4UU/5LL.

Generation of shotgun subclone libraries. The objective of subcloning was to identify fosmid clones containing KS domains and to enable complete fosmid insert sequencing. Fosmid DNA was extracted from each *E.coli* clone. To remove genomic DNA, various methods were used:

1. Alkaline lysis with Plasmid Safe Exonuclease (Epicentre, Madison, WI) treatment
2. Large Construct Kit (Qiagen, Germantown, MD) with Plasmid Safe Exonuclease (Epicentre, Madison, WI)
3. Gel extraction and electroelution of the fosmid DNA
4. Use of BAC/PAC DNA kit (Omega, Norcross, GA)
5. An improved alkaline lysis method was used for 454 sequencing that included a large scale prep (1 L culture size), the neutralization step was subjected to two sequential centrifugation spins at 20,000 x g, and two plasmid safe exonuclease digests were performed.

Isolation and 454 sequencing of fosmid DNA. Fosmid DNA was extracted from the clones A2, A12, B8, and C1. In order to maintain a high concentration of DNA, plasmid-safe exonuclease digestion was performed to remove genomic DNA.

454 Sequencing. Four fosmid clones with strong positive signals on a Southern blot were submitted for 454 sequencing in a pooled format, and conducted at the University of South Carolina's Genomics Center (Sciences 2006). Through this method, samples were fractionated into small 300 to 800 base-pair fragments and adaptors were ligated to each fragment. The "emulsion PCR" technique allowed the adaptors and single stranded fragments to bind to individual and unique beads. The bead and attached fragment were encapsulated by a water-oil immersion. This creates a "microreactor" which contains one bead with one unique fragment. The fragments were then amplified to a copy number of several million per bead. The fragments remain bound to their beads and the amplified beads were loaded onto a plate for sequencing.

Initial bioinformatics analysis. Fosmid DNA sequences determined by 454 sequencing were analyzed for the presence of PKS-specific genes. After identifying ORFs using Chromas Pro software, the predicted protein sequence was submitted for BLASTp analysis. Any predicted ORF without significant homology to other gene products in GenBank were submitted for BLASTx and BLASTn analysis.

Restriction Fragment Length Polymorphism. Restriction fragment length polymorphism (RFLP) analysis was used to determine if the fosmid clones identified as containing a KS domain were unique and to determine the size of each cloned insert DNA. Multiple restriction enzymes were used to determine the best enzyme for

generating RFLPs: BamHI, ClaI, EcoRI, EcoRV, Hind III, Not I, PstI, SalI, SmaI, SphI, and XbaI.

Pulse Field Gel Electrophoresis. In order to get significant separation of the DNA bands after restriction enzyme digests, pulsed field gel electrophoresis was used. Restriction digested DNA was run on a 1% TAE gel at 5 volts/centimeter with a 1 to 15 second switch time for 16 hours.

Contig assembly. Sequences obtained from shotgun subcloning and 454 sequencing were assembled using Sequencher 4.1.4 using a clean data algorithm with a 70% sequence homology setting and a minimum 30 base pair overlap, and separately a dirty data algorithm was used that had a 80% sequence homology setting and a minimum 20 base pair overlap. Assemblies were analyzed after each round of assembly for elimination of low quality sequences. Assembled sequences were submitted for BLASTx and BLASTn analysis. Forward and reverse primers were constructed from contig sequences for primer walking.

PCR joining of contigs. Different combinations of primers were used with A12 DNA template to join existing contigs and extend the degree of contiguous sequences available from this A12 clone. Samples that yielded a PCR product were submitted for sequencing at Auburn University's Sequencing Facility.

Primer Walking. Primers were also designed to be specific to the available fosmid A12 contigs, and both fosmid A12 DNA template and primer DNA was supplied to Lucigen for primer walking to extend the contiguous fosmid A12 sequences.

RESULTS

Generation of a PKS probe

The fosmid library contained 9,984 clones in duplicate with an average insert size of 42kb. Multiple primer sets specific to PKS or NRPS pathways were used in this study to prepare a probe for macroarray hybridization. Table 2.1 includes the eight primers used for PCR. In each experiment, a pooled fosmid library DNA template was used with each primer set, and with several primer sets a PCR product was observed (Figure 2.1); however, a PCR product of the same size was also observed using *E. coli* genomic DNA as template (Figure 2.1). Every primer set that produced a product from library template (but not *E. coli* DNA template) was cloned and a representative number of clones (n=8) was sequenced to validate that the amplicon was the intended gene target. The only primer set that produced the desired PCR amplicon using a metagenomic library template and not with host genomic DNA template was primer set 4UU/5LL, which is specific to the ketosynthase (KS) domain of Type I PKS pathways. The PCR product obtained from the pooled fosmid DNA using the 4UU/5LL primer set was DIG-labeled and used for colony blot hybridization.

Identification of Type I PKS-containing clones

A total of 34 clones exhibited chemiluminescence (in duplicate), indicating the presence of a KS domain and therefore a possible PKS pathway (or at the very least a KS domain of a PKS pathway) within the cloned insert. Only 19 fosmid clones were detected in the absence of arabinose-induction. Each of these 19 clones was also detected from the arabinose-induced macroarrays as well, which in total revealed 34 KS-positive clones

(Fig. 2.2). The position of the duplicate clones relative to each other enabled determination of fosmid clone identity. The KS-positive fosmid clones were grown from glycerol stocks stored in the -80°C freezer, and stored as separate glycerol stocks. Fosmid DNA was isolated from each putative KS-positive clone for use in validating the presence of the KS domain on each clone. Of the 34 fosmid clones that were identified by hybridization, only 21 of these clones yielded a PCR product.

Phylogenetic Analysis of Fosmid KS domains.

Of the 34 fosmid clones that were identified by hybridization, only 21 of these clones yielded a PCR product using the 4UU/5LL KS domain-specific primer set, despite repeated attempts to amplify the KS domain from the other 13 clones. The DNA sequence derived from the 21 KS PCR products had a KS domain as a top hit (51-73% identity range) in the GenBank nr/nt database, but as the highest percent homology observed was 73% it is apparent that these KS domains are phylogenetically distinct from the previously characterized KS domains present in the GenBank nr/nt database. Maximum parsimony analysis of library-derived KS domains and diverse KS domains retrieved from the GenBank nr/nt database confirms that the KS domains identified from the soil metagenomic library represent diverse phylogenetic lineages but are related to various known type I PKS pathways (Fig. 2.3). As seen in Figure 2.3, the KS domains present on fosmid clones had similarity to KS domains from pathways responsible for PKS compounds like the myxobacterial compounds epothilone and stigmatellin (Beyer et al. 1999), and the cyanobacterial jamaicamide (Edwards et al. 2004), and the actinobacterial polyketide antibiotics erythromycin (Donadio et al. 1991), and pikromycin (Xue et al. 1998). It should be noted that although many of these library-

derived KS domains have similarity to known KS domains, the resultant chemical moieties produced by these pathways in their native host may be structurally distinct, as the other modules present in these PKS pathways may not be similar in organization to known PKS pathways. Further knowledge of the PKS pathways presumably present on each KS-positive fosmid clone will indicate the degree of similarity of these cloned pathways to the existing database of PKS pathways.

Restriction Fragment Length Polymorphism analysis of KS-positive fosmid clones

Restriction fragment length polymorphism (RFLP) analysis was necessary to determine whether the fosmids identified as containing a KS domain were unique clones and whether the size of the cloned insert was sufficient in size to potentially contain an entire biosynthetic pathway. Restriction digests provided the best digests of the fosmid clones (nine fosmid clones shown in Fig. 2.4, other fosmid clones not shown). These data indicated that no two fosmid clones had similar RFLP results, as expected if each fosmid clone represents a unique cloned DNA fragment from a different soil microorganism. Furthermore, each fosmid clone determined to have a KS domain had a large DNA insert size of approximately 42 kb, consistent with the average insert size of the fosmid library.

Evaluation of methods for the removal of chromosomal DNA contamination. Several methods were used to attempt to eliminate chromosomal DNA contamination and still achieve a high yield of fosmid DNA :

1. Alkaline lysis with plasmid safe exonuclease (Epicentre, Madison, WI) treatment generated a shotgun subclone library which unfortunately had about 50% frequency of subclones that contained *E. coli* genomic DNA.

2. Large Construct Kit (Qiagen, Germantown, MD) with plasmid safe exonuclease (Epicentre, Madison, WI) digest had very pure fosmid DNA but with a poor yield.
3. Gel extraction and electroelution of the fosmid DNA showed pure fosmid DNA with a poor yield.
4. Use of a BAC/PAC DNA kit (Omega, Norcross, GA) revealed significant chromosomal DNA present on an agarose gel.
5. An improved alkaline lysis for 454 sequencing was successful in reducing the extent of chromosomal contamination (although still present) compared to the original alkaline lysis protocol.

Southern hybridization of PCR-negative fosmid clones

Of the 34 fosmid clones identified by the original library macroarray hybridization, 13 of these clones never produced a KS domain PCR product, despite multiple attempts at PCR amplification using a variety of PCR conditions. To determine if each of these fosmid clones contained a KS domain, a Southern hybridization was performed using a KS domain probe generated by PCR amplification of the KS domains from the pooled collection of 21 PCR-positive fosmid clones. The agarose gel of the restriction digested fosmid clones (e.g., Fig. 2.4) was used to transfer DNA onto a nylon membrane for Southern blot hybridization. The PCR product generated from pooled fosmid DNA using 4UU/5LL was subsequently DIG-labeled and used to hybridize to the fosmid DNAs under medium stringency conditions. Eight of the 13 PCR-negative clones yielded KS-positive Southern blot results (e.g., Fig. 2.5).

Of all of the fosmid-derived sequences generated by 454 sequencing, there was one contig resulting from this data that had significant homology to a PKS pathway. The fosmid contig has 52% identity (70% similarity) over a portion of a 577 bp sequence to MtaD from *Stigmatella aurantiaca*, a fruiting body-forming myxobacterium (Fig. 2.6).

MtaD is predicted to be necessary for chain extension for a hybrid PKS and NRPS pathway for myxothiazol biosynthesis (Silakowski et al. 1999). Myxothiazol is an inhibitor of cytochrome *bc*₁ complex (complex III) and acts by inhibiting electron transport (Thierbach et al. 1981).

Unfortunately, the 454 sequencing did not produce a complete fosmid insert sequence. It was hoped that the end-sequences generated from each fosmid clone (by sequencing with vector-specific primers) would allow contig assembly of fosmid end-sequences with sequences derived from 454 sequencing, thereby identifying the exact fosmid from which each insert sequence was derived. Therefore, in order to determine which fosmid clone this PKS sequence had originated from, PCR primers were designed (see Fig. 2.6) and used to PCR amplify this PKS-related sequence from each of the fosmid clones submitted for 454 sequencing, resulting in the determination that these sequences were derived from fosmid clone A12 (Fig. 2.7). Only the primers highlighted in blue in Fig. 2.6 produced a PCR product in this reaction. This fosmid A12 PCR product was submitted for sequencing, to confirm that the sequence shown in Fig. 2.6 is present on fosmid A12.

Sequence and bioinformatic analysis of fosmid A12 insert sequences

Contigs obtained from assembly of A12 sequences were compared to the complete genome of *Solibacter usitatus* Ellin 6076. Regions of homology are shown in Figure 2.8. The sequence containing the target sequence from A12 showed greatest homology to *Stigmatella aurantiaca*. Further BLASTx analysis revealed alignments and gene products of the contigs. Figure 2.9 panels A-F represents the pairwise alignments derived from BLASTx analysis for fosmid A12 contig sequences.

Percent G + C plots and percent identity plots were performed with each of the fosmid A12 contigs and respective regions of the *Solibacter usitatus* genome. Results of the % G+C plots and the % identity with regions of the *S. usitatus* genome are shown in Figure 2.10 panels A-F revealed a consistent % G + C content across all contig sequences of approximately 60%, consistent with the average % G+C content of Acidobacteria sp.

DISCUSSION

Multiple primer sets targeting both PKS and NRPS pathways were evaluated for use in screening metagenomic libraries. The criterion used for selecting a primer set was that it produced a PCR product with a pooled library DNA template, but not with *E. coli* genomic DNA template, and that sequence generated from PCR amplicons were verified to have some degree of homology to known PKS or NRPS pathways. Several primer sets fulfilled this criterion, yet only the Type I PKS KS domain-specific primer set (4UU/5LL) produced a library amplicon that when cloned and sequenced, yielded the expected KS domain sequences. This KS domain-specific primer set was used for all subsequent work in this study.

The metagenomic library macroarrays were prepared by growing the *E. coli* fosmid clones with and without arabinose induction of fosmid copy number. Without the induction of arabinose, plasmid copy number would have remained at a low number, allowing fewer targets for the probe to hybridize to.

The observation that a greater number of fosmid clones were identified with copy number induction (n=34) compared to without copy number induction (n=19) is attributable to the increase in the number of fosmid DNA targets upon copy induction on each macroarray spot. It is possible that some fosmid clones may encode gene products that are toxic to the *E. coli* host; therefore it was decided to do this experiment with and without copy number induction. However, each of the 19 clones detected without copy number induction were also detected when arabinose was added to induce copy number, suggesting that at least among this set of clones none of the fosmid DNA inserts proved toxic to *E. coli* at higher copy number.

A phylogenetic analysis of the KS domains present on the 21 fosmids that yielded a PCR product confirms the similarity to known KS domains (51-73% identity) as well as the diverse phylogenetic origins of these cloned genes. This diversity is further confirmed by the banding patterns seen on RFLP analysis. One of the most interesting points of the phylogenetic tree is the significant relation to cyanobacterial species. Cyanobacteria species carry out photosynthesis. Because the first 5cm of soil was removed before DNA extraction, it is unusual that photosynthetic Cyanobacteria would be found in the deeper layers of soil. Further sequence analysis of these fosmid clones will provide a greater degree of phylogenetic resolution as to the origin of these cloned PKS pathways and the genetic organization of the modular PKSs.

A significant percentage of the 34 fosmid clones identified by library macroarray hybridization never yielded a KS domain-specific PCR product, leading to the initial conclusion that these may be false positive identifications, or that the incorrect fosmid clone had been selected from the 384-well formatted library in the -80°C freezer. It was also possible that some of these fosmid clones were true positives and that a small mismatch in the PCR primers could prevent PCR amplification yet fosmid sequences may still be detected by Southern blot hybridization with a ~700 bp heterogeneous KS domain probe. To test the hypothesis that some of these reproducibly PCR-negative fosmid clones did in fact contain a KS domain, a Southern blot was performed using a KS domain probe prepared from the pooled KS-positive fosmid DNAs, and hybridized against each of the PCR-negative fosmid DNAs along with positive controls. The Southern blots revealed that 8 of the 13 clones found to be negative by PCR were still KS domain-positive by Southern blot hybridization. Among these eight fosmid clones that

are PCR-negative and Southern blot-positive, many revealed multiple KS-domain hybridizing bands. It is hypothesized that these latter fosmid clones do contain a PKS pathway and that the genes present within these fosmid clones will be found to have an even lower percent homology to the available PKS sequences in GenBank compared to the KS domains already identified from the soil metagenomic library.

BLAST analysis of fragments obtained from fosmid A12 revealed homology to a gene found in *Stigmatella aurantiaca*, encoding a hybrid PKS and NRPS pathway myxothiazol biosynthesis (Silakowski et al. 1999). Myxothiazol is an inhibitor of cytochrome *bc*₁ complex (complex III) and acts by inhibiting electron transport (Thierbach et al. 1981). Interestingly, BLASTx analysis of the target sequence from fosmid A12 revealed top hits to members of Mxyobacteria and Cyanobacteria species. Subsequent investigation following contig assembly of A12 sequences revealed that most contigs were homologous to *Solibacter usitatus*. *S. usitatus* is a member of the phylum Acidobacteria. Only one known Acidobacterial species isolated from acidic drainage has been shown to demonstrate PKS pathways. No PKS pathways from soil Acidobacteria have been isolated. This phenomenon reflects the scarcity of Acidobacteria genomes present in the GenBank database.

SWAAP software was used to obtain G + C content and percent similarity of A12 contig sequences to the *S. usitatus* genome. Figure 2.10 shows G + C percentages in the 60% range, consistent with known Acidobacteria G + C content. Similarity graphs of contig sequences also revealed strong resemblance to *S. usitatus*.

CHAPTER 3
CHARACTERIZATION OF METAGENOMIC CLONES EXPRESSING
ANTIMICROBIAL ACTIVITY

INTRODUCTION

In addition to sequence-based methods for identifying recombinant clones, function-based screening has also been utilized to detect secondary metabolites with antibiotic activity. Clones identified on the basis of antibiotic activity have been selected for the capacity for heterologous expression of antimicrobial activity against specific bacterial strains. Therefore, these active clones may be promising candidates for the expression of bioactive compounds with antimicrobial activity. As opposed to the fosmid clones that contain PKS pathways, these recombinant clones identified from a bioassay may have extremely different biosynthetic pathways, potentially with little or no homology to known biosynthetic pathways in the GenBank databases.

This method has a significant advantage in that it does not require that the clones of interest be identified by sequence analysis, and that structurally diverse compounds may be identified from a single screen for antimicrobial activity. However, the limitation of this approach is that genes products involved in the biosynthesis of antibiotic compounds in a native host may be present but not expressed in an *E. coli* heterologous host (Handelsman 2004). Since there has been evidence of heterologous expression of metagenomic DNA from diverse microbial phyla, and the bioassay used will select for those recombinant clones that do express a bioactive compound with antimicrobial activity, the biases inherent in this approach will limit the number of clones that may be detected in a bioassay but not prevent some subset of clones from being detected in a bioassay.

MATERIALS AND METHODS

Sampling and DNA isolation. Soil samples were collected from University of Wisconsin-Madison's Hancock Agricultural Research Station. Bacterial cells were collected by soil homogenization and differential centrifugation. Cells were embedded and lysed within an agarose plug, purified using a formamide denaturation step, and high molecular weight metagenomic DNA was used for restriction digestion and cloning (Liles et al. 2004).

BAC library construction. Metagenomic DNA was partially restriction digested with Sau3AI and electrophoresed using pulsed field gel electrophoresis to select DNA fragments greater than 50 kb in size. The size-selected DNA was electroeluted from the agarose gel and ligated into a BamHI-cut pEZBAC vector (Lucigen, Middleton, WI). Ligated DNA was transformed into *E. coli* and transformants were selected on LB agar containing 12.5 µg/ml of chloramphenicol. Colonies were robotically picked into 384-well plate format. The library contained 9,216 clones with an average insert size of 68 kb.

Screening of BAC library. Instead of removing chloramphenicol from *E. coli* supernatants, tester strains that were chloramphenicol resistant were chosen including *Alcaligenes faecalis*, *Flavobacterium meningosepticum*, *Pediococcus cerevisiae*, *Pseudomonas aeruginosa* and *Streptococcus agalactiae*. The supernatant of each BAC clone was grown at 37°C for 72 hours in Luria-Bertani broth with 12.5 µg/ml chloramphenicol. Each clone was spotted onto an Omni plate with Luria-Bertani + 12.5 µg/ml chloramphenicol media and grown for 72 hours. The clones were then subjected to choliform lysis, and tester strains were overlaid onto lysed cells. A positive result was indicated by the formation of halos or an inhibition of pigment production.

Liquid bioassays were also established to quantify growth inhibition of tester strains resulting from *E. coli* clone supernatants. Each *E. coli* clone was grown and chloroform lysed within 96-well deep well plates. The supernatants were subjected to centrifugation for 10 min at 3,600 x g and 150 µl of supernatant from individual clones was transferred into 96-well plate format and 50 µl of tester strain in the log phase of growth was added. Ampicillin was also added to each well at a concentration of 50 µg/ml to inhibit the growth of any *E. coli* cells remaining in the supernatant (each tester strain was also resistant to ampicillin).

Transposon Mutagenesis. Twelve of the 44 clones were mutagenized with a mini-Tn5-Kan^R transposon cassette (Epicentre, Madison, WI). Mutants were arrayed into 96-well format for evaluation of loss-of-function mutants.

Evaluation of growth media. Antibiotic activity of clones against tester strains was tested in several different growth media: Luria-Bertani broth, 2x YT broth, Tryptic Soy Broth, M9 minimal media with 20% glucose, and Brain Heart Infusion broth, and Hsu-Shotts broth containing 12.5 µg/ml chloramphenicol.

Evaluation of supernatant activity. Because secondary metabolites are likely produced in stationary phase, extended growth times were tested for enhanced activity against tester strains. Clones were grown for 72 hours in Hsu-Shotts broth with 12.5 µg/ml chloramphenicol and the OD₆₀₀ was taken to assess the inhibition of tester strains at 24, 48, and 72 hour time points.

RESULTS

Screening of BAC Libraries

E. coli BAC clones were tested using several methods against a panel of chloramphenicol-resistant bacterial and yeast tester strains. Table 3.1 is a representation of the results of soft agar overlays and liquid bioassays conducted against *Alcaligenes faecalis* and *Pseudomonas aeruginosa*. Assays were performed by obtaining individual BAC clone culture supernatant, overlaying tester strain, and reading the bacterial culture optical density at 600nm.

After overlays and liquid bioassays, 160 clones were identified and tested in liquid bioassay in duplicate to confirm the presence of an antimicrobial activity. From the 160 initial clones, 44 clones showed consistent activity in duplicate and were selected for retransformation and among these, the best candidates were chosen for transposon mutagenesis.

Evaluation of Clone Activity against *Alcaligenes faecalis*

It was determined that growth of the *E. coli* recombinant clones for 72 hours showed the best growth inhibition of tester strains for some recombinant clones. Supernatants isolated from BAC Clone P15G24 were evaluated over a time course of 72 hours. Assays in 96-well plate format were incubated at 37°C and OD₆₀₀ was taken every 24 hours to evaluate the growth of tester strains, with and without supernatants from clone P15G24 and each respective transposon mutant. An increase in inhibitory activity for P15G24 was observed compared to some P15G24 transposon mutants (Fig. 3.1).

Clone P15G24 has shown strong and consistent activity against both *A. faecalis* and *P. aeruginosa*. Interestingly, higher antibiotic activity was observed when this

recombinant clone was grown in Hsu-Shotts broth with compared to LB, 2x YT, Tryptic Soy Broth, M9 minimal media with 20% glucose, or Brain Heart Infusion media.

DISCUSSION

In order to enhance the expression of antimicrobial activity from recombinant clones, *E. coli* cultures were incubated for 72 hours. Prolonged incubation time results in the depletion of the growth medium leading to a complex cascade of regulatory signals which shuts down expression of growth genes, and turns on survival genes that may encode secondary metabolites (Nystrom 2004). Secondary metabolites are synthesized only by some microorganisms and are often produced in connection with cellular differentiation. These metabolites can include secreted products that inhibit the growth of competing bacteria. From at least one of the BAC clones expressing antimicrobial activity (i.e., P15G24), an increase in antimicrobial activity within the supernatant of this clone was observed at 72 hours, supporting the idea that prolonged incubation of the *E. coli* culture may increase the rate of discovery of recombinant clones with bioactivity. This also suggests that the growth inhibitory compound(s) are being produced and/or excreted during stationary phase, as predicted.

Table 3.1 represents the activity of individual clones against *Alcaligenes faecalis* and *Pseudomonas aeruginosa*. Clones were tested by soft agar overlay and by liquid bioassay in order to identify individual clones with antimicrobial activity. Soft agar overlays with chloroform lysis were carried out to ensure that clones with biosynthetic products that are not secreted from the *E. coli* cell may still be identified. Performing liquid assays in 96-well plates reduced the potential for cross-reactivity of clones as well as allowed for quantification of any inhibitory effect of clone supernatant on the tester strains. Because of cellular aggregation and poor growth associated with some tester

strains, only *Pseudomonas aeruginosa* and *Alcaligenes faecalis* were chosen for use in further tests.

FUTURE DIRECTIONS

Each of 16 different *E. coli* clones that have been observed to have antimicrobial activity will be grown in different media (e.g., Hsu-Shotts, LB, others) and supernatants tested for growth inhibition of *A. faecalis* and *P.aeruginosa*. The negative control of *E. coli* expressing the empty vector will be included in each experiment. Each of the cultures will also be treated with and without chloroform lysis to determine if this increases antimicrobial activity.

Clones with significant antimicrobial activity will be studied to determine the nature of the chemical compound(s) produced from their respective clone. As each recombinant clone could produce a unique chemistry, it will be important to winnow down the number of clones to study. The molecular weight of the active compound will be estimated by passing active supernatants through membranes with different size exclusions. For example, an active compound in the supernatant of P15G24 seems to be less than 1 kDa in size, whereas other clone supernatants have a much larger molecular weight compound. Smaller molecular weight compounds will have preference for future development. Each active supernatant will be subjected to boiling for 10 min as well as proteinase K digestion. Supernatants with and without heat-inactivation or enzymatic treatment will be tested against a susceptible tester strain to determine if the bioactive compound may be proteinaceous. Preference will be given to non-proteinaceous compounds.

Other strains of the tester bacterial species will be obtained to determine if strains that are resistant to known antibiotics have the ability to grow in the presence of the growth inhibitory compound(s). This could provide evidence of the biochemical nature of the

unknown antimicrobial compound. Preference would be given to those compounds that continue to inhibit the growth of other antibiotic-resistant bacterial strains.

Each non-proteinaceous compound present within *E. coli* clone supernatants will be evaluated for its ability to be extracted by ethyl acetate (and possibly other) organic solvents. This could make purification of the antimicrobial compounds much easier in further steps. The ethyl acetate fraction would be dried and then resuspended in 100% DMSO. Dilutions of DMSO resuspensions will be added to tester strain media at concentrations of DMSO that are sub-inhibitory to the tester strain, and evaluated for antibiotic activity. Negative controls will include ethyl acetate extractions of the *E. coli* culture with empty vector in each bioassay.

The chemical structure of an antibiotic compound may be determined by LC/MS/MS and other biochemical techniques dependent upon the availability of funds to support this research effort.

Transposon mutagenesis has already been conducted on 12 of the BAC clones with antimicrobial activity. The BAC clone(s) to be studied further at a genetic level will be dependent upon biochemical tests in Objectives 1 and 2. A collection of loss-of-function transposon mutants already exists for some BAC clones, and loss-of-function mutants would be identified from other clones based on results of antimicrobial bioassays as performed above. Inverse PCR would be conducted to identify the gene(s) necessary for antimicrobial biosynthesis in each recombinant clone, followed by sequencing of the PCR product. Multiple loss-of-function transposon mutants would be evaluated for each bioactive BAC clone.

CONCLUSION

The great need for novel antimicrobial compounds is clear. Because of the reduced interest by the pharmaceutical industry, the responsibility to discover and describe the compounds has fallen on academia. Methods depending on microbial culturing are falling short, therefore shifting the focus to culture-independent detection. PKSs are promising candidates given their modular arrangement and repeating domains. Further characterization of these KS compounds will reveal novel compounds as potential antibiotic candidates.

LITERATURE CITED

- Allen, E. E. and J. F. Banfield (2005). "Community genomics in microbial ecology and evolution." Nat Rev Microbiol **3**(6): 489-98.
- Austin, M. B. and J. P. Noel (2003). "The chalcone synthase superfamily of type III polyketide synthases." Nat Prod Rep **20**(1): 79-110.
- Ayuso-Sacido, A. and O. Genilloud (2005). "New PCR primers for the screening of NRPS and PKS-I systems in actinomycetes: detection and distribution of these biosynthetic gene sequences in major taxonomic groups." Microb Ecol **49**(1): 10-24.
- Beja, O., L. Aravind, et al. (2000). "Bacterial rhodopsin: evidence for a new type of phototrophy in the sea." Science **289**(5486): 1902-6.
- Beja, O., M. T. Suzuki, et al. (2002). "Unsuspected diversity among marine aerobic anoxygenic phototrophs." Nature **415**(6872): 630-3.
- Beja, O., M. T. Suzuki, et al. (2002). "Unsuspected diversity among marine aerobic anoxygenic phototrophs." Nature **415**: 335-345.
- 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.
- Bevitt, D. J., J. Cortes, et al. (1992). "6-Deoxyerythronolide-B synthase 2 from *Saccharopolyspora erythraea*. Cloning of the structural gene, sequence analysis

- and inferred domain structure of the multifunctional enzyme." Eur J Biochem **204**(1): 39-49.
- Beyer, S., B. Kunze, et al. (1999). "Metabolic diversity in myxobacteria: identification of the myxalamid and the stigmatellin biosynthetic gene cluster of *Stigmatella aurantiaca* Sg a15 and a combined polyketide-(poly)peptide gene cluster from the epothilone producing strain *Sorangium cellulosum* So ce90." Biochim Biophys Acta **1445**(2): 185-95.
- Birch, A. J., P. A. Massy-Westropp, et al. (1955). "Studies in relation to biosynthesis III. 2-Hydroxy-6-methylbenzoic acid in *Penicillium griseofulvum* Dierckx." Australian Journal of Chemistry **8**: 539-544.
- Bull, A. T., A. C. Ward, et al. (2000). "Search and discovery strategies for biotechnology: the paradigm shift." Microbiol Mol Biol Rev **64**(3): 573-606.
- Chater, K. F. and C. J. Bruton (1985). "Resistance, regulatory and production genes for the antibiotic methylenomycin are clustered." EMBO J **4**(7): 1893-7.
- Cottrell, M. T., J. A. Moore, et al. (1999). "Chitinases from uncultured marine microorganisms." Appl Environ Microbiol **65**(6): 2553-7.
- 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.
- Curtis, T. P. and W. T. Sloan (2005). "Microbiology. Exploring microbial diversity--a vast below." Science **309**(5739): 1331-3.
- Davies, J. (1994). "Inactivation of antibiotics and the dissemination of resistance genes." Science **264**(5157): 375-82.

- Davies, J. (2006). "Are antibiotics naturally antibiotics?" J Ind Microbiol Biotechnol **33**(7): 496-9.
- Demain, A. L. and J. Davies (1999). Manual of industrial microbiology and biotechnology. Washington, D.C., ASM Press.
- Demain, A. L. and L. Zhang (2005). Natural Products and Drug Discovery. Totowa, Humana Press, Inc.
- Distler, J., C. Braun, et al. (1987). "Gene cluster for streptomycin biosynthesis in *Streptomyces griseus*: analysis of a central region including the major resistance gene." Mol Gen Genet **208**(1-2): 204-10.
- Donadio, S., M. J. Staver, et al. (1991). "Modular organization of genes required for complex polyketide biosynthesis." Science **252**(5006): 675-9.
- Edwards, D. J., B. L. Marquez, et al. (2004). "Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*." Chem Biol **11**(6): 817-33.
- Eisen, J. A. (1998). "Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis." Genome Res **8**(3): 163-7.
- 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.
- Funa, N., Y. Ohnishi, et al. (2002). "Properties and substrate specificity of RppA, a chalcone synthase-related polyketide synthase in *Streptomyces griseus*." J Biol Chem **277**(7): 4628-35.

- Gill, S. R., M. Pop, et al. (2006). "Metagenomic analysis of the human distal gut microbiome." Science **312**(5778): 1355-9.
- 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.
- Ginolhac, A., C. Jarrin, et al. (2004). "Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones." Appl Environ Microbiol **70**(9): 5522-7.
- Hallam, S. J., N. Putnam, et al. (2004). "Reverse methanogenesis: testing the hypothesis with environmental genomics." Science **305**(5689): 1457-62.
- Handelsman, J. (2004). "Metagenomics: application of genomics to uncultured microorganisms." Microbiol Mol Biol Rev **68**(4): 669-85.
- Handelsman, J., M. R. Liles, et al. (2002). "Cloning the metagenome: Culture-independent access to the diversity and functions of the uncultivated microbial world." Methods in Microbiology **33**: 241-255.
- 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.
- 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.

- Henne, A., R. A. Schmitz, et al. (2000). "Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on Escherichia coli." Appl Environ Microbiol **66**(7): 3113-6.
- Hugenholtz, P., B. M. Goebel, et al. (1998). "Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity." J Bacteriol **180**(18): 4765-74.
- Hutchinson, C. R., Ed. (2005). Manipulating Microbial Metabolites for Drug Discovery and Production. Natural Products Drug Discovery and Therapeutic Medicine. Totowa, Humana Press.
- Hutchinson, C. R. and I. Fujii (1995). "Polyketide synthase gene manipulation: a structure-function approach in engineering novel antibiotics." Annu Rev Microbiol **49**: 201-38.
- Jiao, Y. L., L. H. Wang, et al. (2008). "Isolation of new polyketide synthase gene fragments and a partial gene cluster from East China Sea and function analysis of a new acyltransferase." Appl Biochem Biotechnol **149**(1): 67-78.
- Kaerberlein, T., K. Lewis, et al. (2002). "Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment." Science **296**(5570): 1127-9.
- Khatun, S., A. C. Waugh, et al. (2002). "Distribution of polyketide synthase genes in bacterial populations." J Antibiot (Tokyo) **55**(1): 107-8.
- Lancini, G. and R. Lorenzetti (1993). Antibiotics and Bioactive Microbial Metabolites. New York, Plenum Press.

- Liles, M. R., L. L. Williamson, et al. (2004). "Isolation of high molecular weight genomic DNA from soil bacteria for genomic library construction." Molecular Methods in Environmental Microbiology: 839-852.
- Malpartida, F. and D. A. Hopwood (1984). "Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host." Nature **309**(5967): 462-4.
- McDaniel, R., S. Ebert-Khosla, et al. (1993). "Engineered biosynthesis of novel polyketides." Science **262**(5139): 1546-50.
- McDaniel, R., M. Welch, et al. (2005). "Genetic approaches to polyketide antibiotics. 1." Chem Rev **105**(2): 543-58.
- Metsa-Ketela, M., L. Halo, et al. (2002). "Molecular evolution of aromatic polyketides and comparative sequence analysis of polyketide ketosynthase and 16S ribosomal DNA genes from various streptomyces species." Appl Environ Microbiol **68**(9): 4472-9.
- Molnar, I., T. Schupp, et al. (2000). "The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from Sorangium cellulosum So ce90." Chem Biol **7**(2): 97-109.
- Moore, B. S. and J. N. Hopke (2001). "Discovery of a new bacterial polyketide biosynthetic pathway." ChemBiochem **2**(1): 35-8.
- Motamedi, H. and C. R. Hutchinson (1987). "Cloning and heterologous expression of a gene cluster for the biosynthesis of tetracenomycin C, the anthracycline antitumor antibiotic of Streptomyces glaucescens." Proc Natl Acad Sci U S A **84**(13): 4445-9.

- 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.
- Nystrom, T. (2004). "Stationary-phase physiology." Annu Rev Microbiol **58**: 161-81.
- O'Hagan, D. (1991). The Polyketide Metabolites. New York, Ellis Horwood.
- Parsley, L. and M. R. Liles (2006). Discovery of polyketide synthase pathways from an arrayed soil metagenomic library. ASM Meeting. Orlando, FL.
- Pfeifer, B. A., S. J. Admiraal, et al. (2001). "Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli." Science **291**(5509): 1790-2.
- Pfeifer, B. A. and C. Khosla (2001). "Biosynthesis of polyketides in heterologous hosts." Microbiol Mol Biol Rev **65**(1): 106-18.
- Piel, J. (2002). "A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles." Proc Natl Acad Sci U S A **99**(22): 14002-7.
- Roberts, G. A., J. Staunton, et al. (1993). "Heterologous expression in Escherichia coli of an intact multienzyme component of the erythromycin-producing polyketide synthase." Eur J Biochem **214**(1): 305-11.
- 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.
- Sandler, S. J., P. Hugenholtz, et al. (1999). "Diversity of radA genes from cultured and uncultured archaea: comparative analysis of putative RadA proteins and their use as a phylogenetic marker." J Bacteriol **181**(3): 907-15.

- Schirmer, A., R. Gadkari, et al. (2005). "Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta*." *Appl Environ Microbiol* **71**(8): 4840-9.
- Schleper, C., E. F. DeLong, et al. (1998). "Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Cenarchaeum symbiosum*." *J Bacteriol* **180**(19): 5003-9.
- Schleper, C., R. V. Swanson, et al. (1997). "Characterization of a DNA polymerase from the uncultivated psychrophilic archaeon *Cenarchaeum symbiosum*." *J Bacteriol* **179**(24): 7803-11.
- Schroder, J. (1999). "Comprehensive Natural Products in Chemistry." 749-771.
- Sciences, R. A. (2006). Genome Sequencer FLX System. Indianapolis, Indiana.
- Seow, K. T., G. Meurer, et al. (1997). "A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms." *J Bacteriol* **179**(23): 7360-8.
- Shen, B. (2003). "Polyketide biosynthesis beyond the type I, II, and III polyketide synthase paradigms." *Current Opinion in Chemical Biology* **7**: 285-295.
- Silakowski, B., H. U. Schairer, et al. (1999). "New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1." *J Biol Chem* **274**(52): 37391-9.
- 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.

- Staunton, J. and K. J. Weissman (2001). "Polyketide biosynthesis: a millennium review." Nat Prod Rep **18**(4): 380-416.
- 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.
- Summers, R. G., E. Wendt-Pienkowski, et al. (1992). "Nucleotide sequence of the tcmII-tcmIV region of the tetracenomycin C biosynthetic gene cluster of *Streptomyces glaucescens* and evidence that the tcmN gene encodes a multifunctional cyclase-dehydratase-O-methyl transferase." J Bacteriol **174**(6): 1810-20.
- Summers, R. G., E. Wendt-Pienkowski, et al. (1993). "The tcmVI region of the tetracenomycin C biosynthetic gene cluster of *Streptomyces glaucescens* encodes the tetracenomycin F1 monooxygenase, tetracenomycin F2 cyclase, and, most likely, a second cyclase." J Bacteriol **175**(23): 7571-80.
- Thierbach, G. and H. Reichenbach (1981). "Myxothiazol, a new antibiotic interfering with respiration." Antimicrob Agents Chemother **19**(4): 504-7.
- Tyson, G. W. and J. F. Banfield (2005). "Cultivating the uncultivated: a community genomics perspective." Trends Microbiol **13**(9): 411-5.
- Vergin, K. L., E. Urbach, et al. (1998). "Screening of a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order Planctomycetales." Appl Environ Microbiol **64**(8): 3075-8.
- Wakil, S. J. (1989). "Fatty acid synthase, a proficient multifunctional enzyme." Biochemistry **28**(11): 4523-30.

- Waksman, S. A. (1961). "The role of antibiotics in nature." Perspectives in Biology and Medicine **IV**: 271-286.
- Wanibuchi, K., P. Zhang, et al. (2007). "An acridone-producing novel multifunctional type III polyketide synthase from *Huperzia serrata*." FEBS J **274**(4): 1073-82.
- Wawrik, B., L. Kerkhof, et al. (2005). "Identification of unique type II polyketide synthase genes in soil." Appl Environ Microbiol **71**(5): 2232-8.
- Whatman. (2007). "Turboblotter Protocol." Retrieved 12/2/2008, from <http://www.whatman.com/TurboBlotter.aspx>.
- Woese, C. R., O. Kandler, et al. (1990). "Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya." Proc Natl Acad Sci U S A **87**(12): 4576-9.
- Xue, Y., L. Zhao, et al. (1998). "A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: architecture of metabolic diversity." Proc Natl Acad Sci U S A **95**(21): 12111-6.
- Yim, G., H. H. Wang, et al. (2006). "The truth about antibiotics." Int J Med Microbiol **296**(2-3): 163-70.
- Zaehner, H. and F. P. Fiedler, Eds. (1995). Fifty years of antimicrobials: past perspectives and future trends. The Need for New Antibiotics: Possible Ways Forward. Fifty-Third Symposium of the Society for General Microbiology. Cambridge, UK, Cambridge University Press.
- Zengler, K., G. Toledo, et al. (2002). "Cultivating the uncultured." Proc Natl Acad Sci U S A **99**(24): 15681-6.

Table 2.1 PCR primers for design of KS probe. (Seow et al. 1997; Metsa-Ketela et al. 2002; Piel 2002; Ginolhac et al. 2004; Ayuso-Sacido et al. 2005; Wawrik et al. 2005)

| | | |
|---|---------|----------------------------------|
| 1 | A3 | GCSTACSYSATSTACACSTCSGG |
| | A7R | SASGTCVCCSGTSCGGTAS |
| 2 | K1 | TSAAGTCSAACATCGGBCA |
| | M6R | CGCAGGTTSCSGTACCAGTA |
| 3 | KSLF | CCSCAGSAGCGCSTSYTSCTSGA |
| | KSLR | GTSCCSGTSCCGTGSGYSTCSA |
| 4 | KSDPQQF | MGNGARGCNNWNSMNATGGAYCCNCARCANMG |
| | KSHGTGR | GGRTCNCCNARNSWNGTNCCNGTNCCRTG |
| 5 | 5LL | GGRTCNCCIARYTGIGTICCGTICCRTGIGC |
| | 4UU | MGIGARGCIYTICARATGGAYCCICARCARMG |
| 6 | F | TSGCSTGCTTGGAYGCSATC |
| | R | TGGAANCCGCCGAABCCGCT |
| 7 | 540F | GGITGCACSTCIGGIMTSGAC |
| | 1100R | CCGATSGCICCSAGIGAGTG |
| 8 | kSALPHA | TTCGGCGGXTTCCAGTCXGCCATG |
| | ACP | CCXATGCTCAGCXACCGCGACGACCT |

Table 3.1. Activity of BAC clones tested against *Alcaligenes faecalis* and *Pseudomonas aeruginosa*

| Clone | Activity Against <i>Alcaligenes faecalis</i> (times positive activity/times tested) | Activity against <i>Pseudomonas aeruginosa</i> (yes or no) |
|--------|--|---|
| P5K24 | 3/4 | Yes |
| P8B10 | 1/4 | No |
| P8P8 | 2/4 | Yes |
| P10L17 | 4/5 | Yes |
| P10M15 | 2/2 | Yes |
| P11K11 | 4/5 | Yes |
| P15C24 | No activity | Yes |
| P15G24 | 2/5 | Yes |
| P15L24 | 4/4 | Yes |
| P15P12 | 5/5 | Yes |
| P16O8 | 2/4 | Yes |
| P17C10 | 1/4 | No |
| P17H13 | No activity | Yes |
| P17L5 | 3/5 | Yes |
| P18H24 | No activity | Yes |
| P19B26 | 2/5 | No |
| P19J24 | No activity | Yes |
| P21H5 | 2/4 | No |
| P23D10 | No activity | Yes |
| P23P1 | No activity | Yes |
| P24A23 | 1/2 | Yes |
| P24H20 | 3/4 | Yes |
| P24P22 | 1/4 | No |

Figure 1.1. Modular arrangement of type I PKS pathways (Staunton et al. 2001)

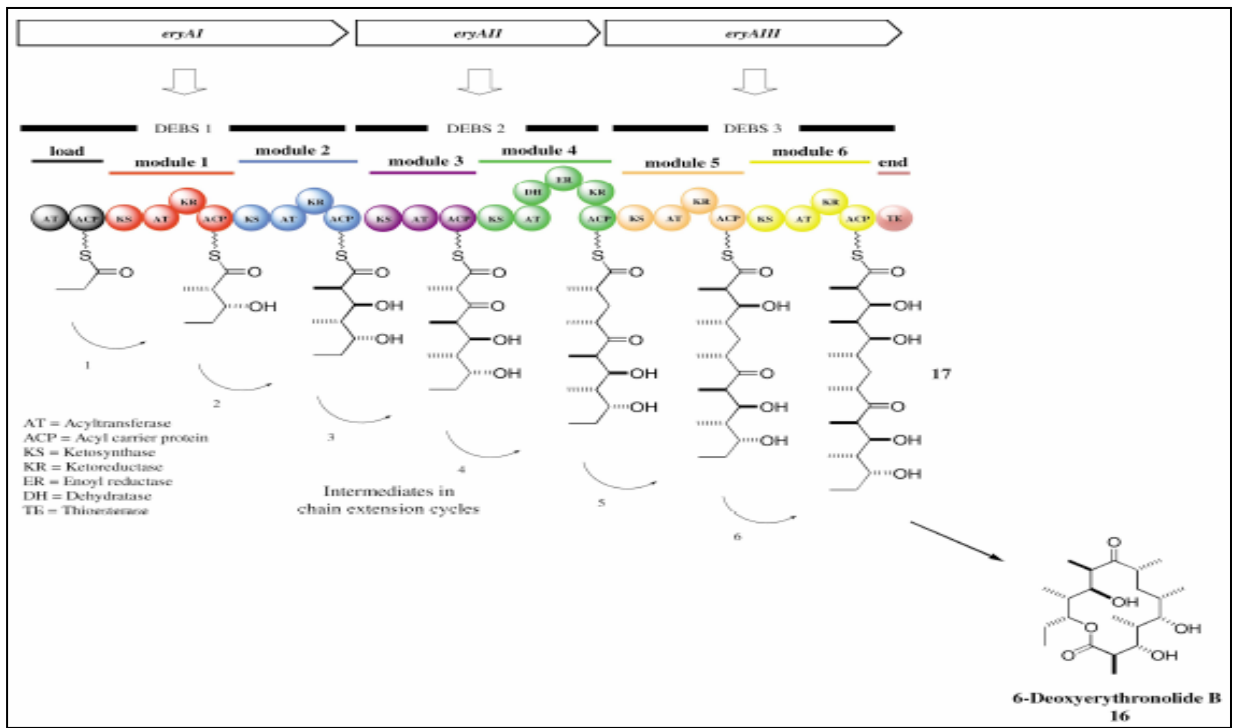


Figure 2.1 Gel electrophoresis with multiple KS primer sets using pooled fosmid, diluted fosmid DNA, positive KS clones, and *E. coli* as template. The first eight lanes are primers 1-8, respectively, using undiluted fosmid template. The second eight lanes are primers sets 1-8 using 1/10 diluted fosmid template. The third eight lanes are primer sets 1-8 with pooled KS fosmid clones as template. The fourth set of eight lanes are primer sets 1-8 with *E. coli* as template.

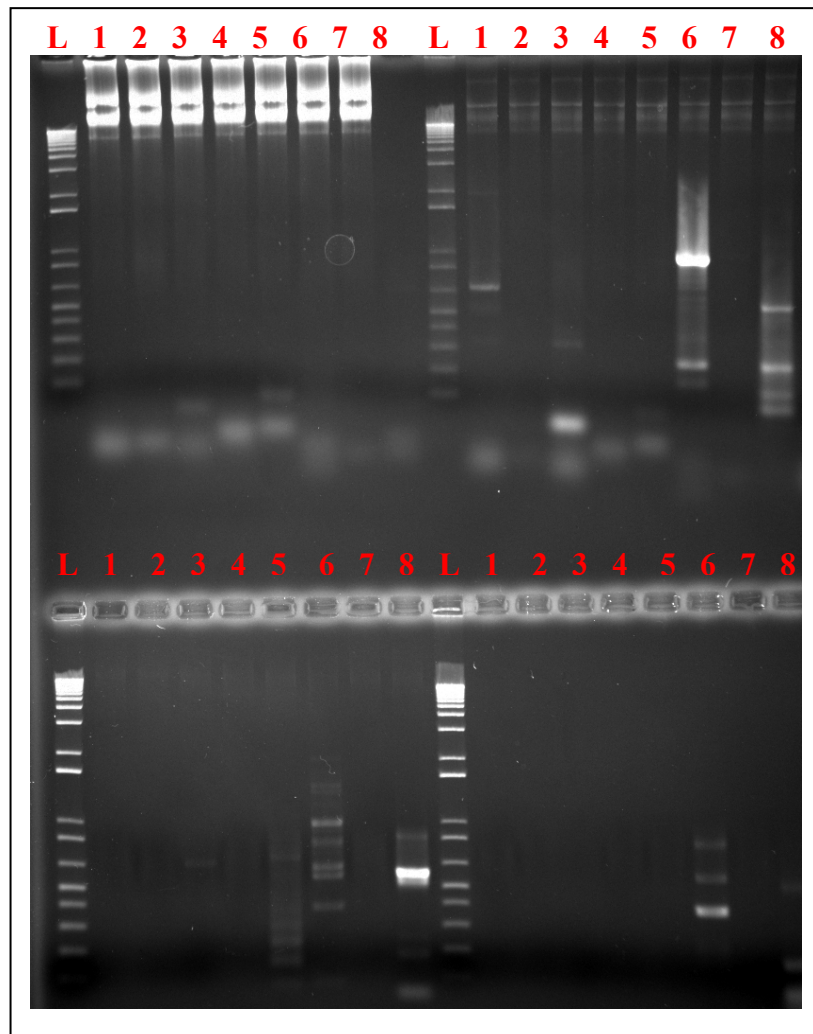


Figure 2.2. Detection of KS-domain positive fosmid clones. A digoxigenin (DIG)-labeled KS probe from pooled fosmid DNA template was used for colony blot DNA hybridization. A subset of the KS-positive fosmid clones is shown.

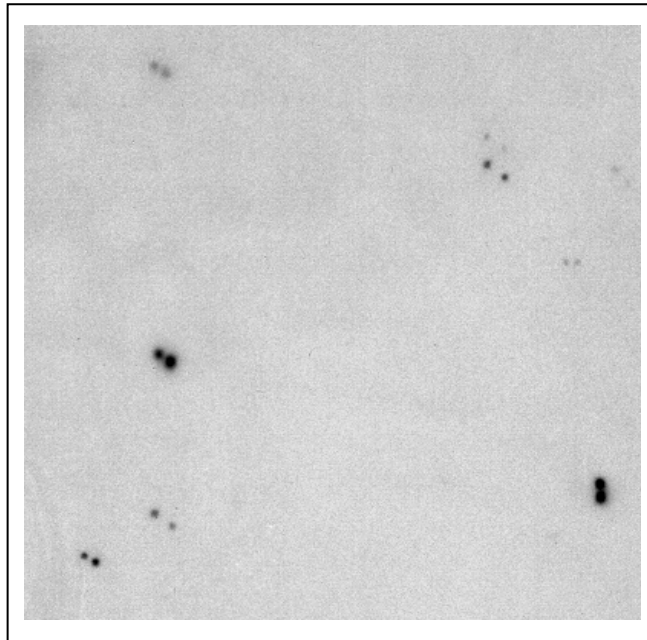


Figure 2.3 Maximum parsimony tree of KS domains recovered from fosmid clones. KS domain sequences recovered from the soil metagenomic library are underlined, and in bold. Numbers at tree nodes represent bootstrap values, and the tree is rooted using the *E. coli* FabB sequence. Scale represents sequence changes/bp.

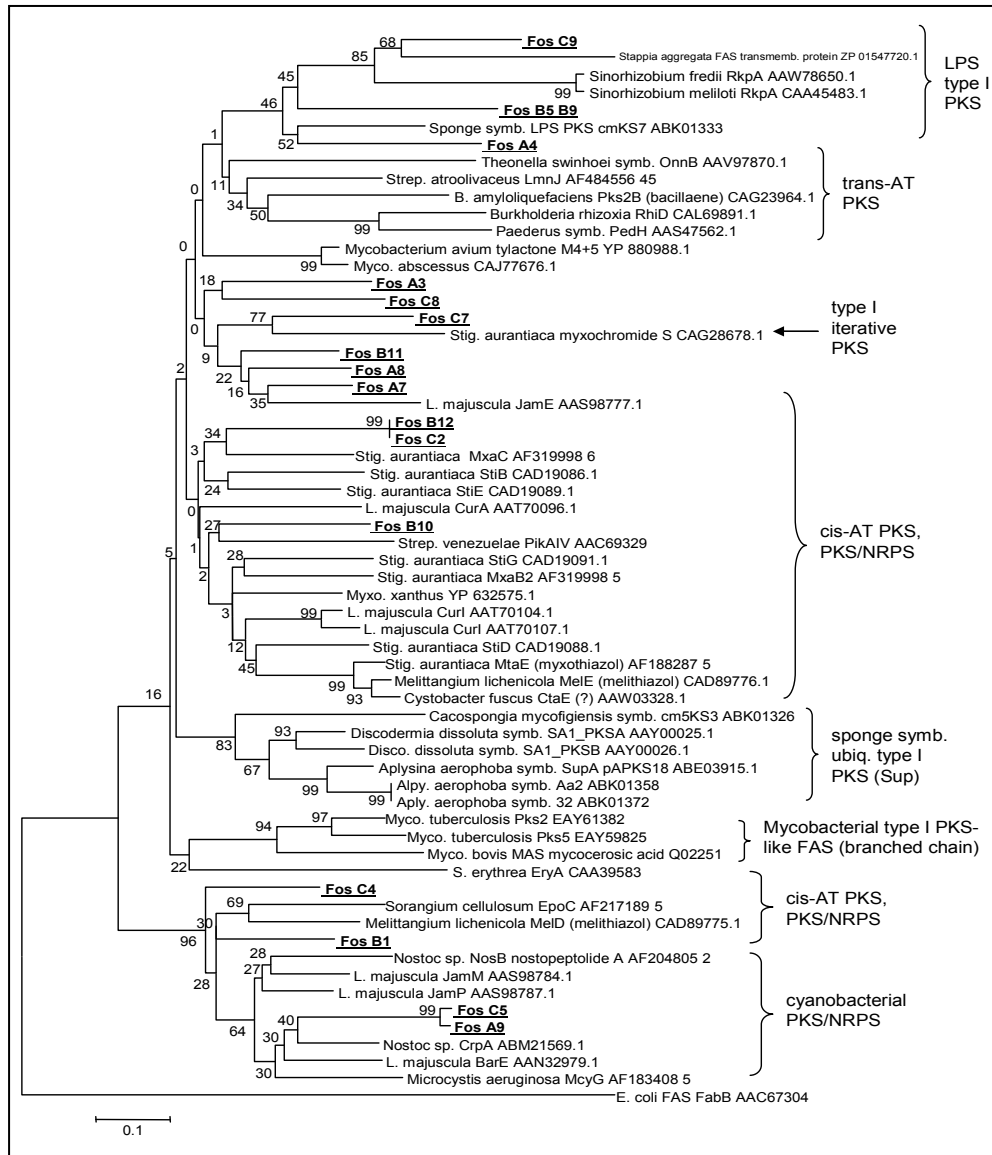


Figure 2.4. Restriction Fragment Length Polymorphism Analysis of fosmid clones.

Fosmid clones were restriction digested with BamHI to determine the size and uniqueness of each clone. The lanes are as follows: lane 1: ladder, lane 2: Clone A1, lane 3: Clone A2, lane 4: Clone A5, lane 5: Clone A12, lane 6: Clone B3, lane 7: Clone B5 (control), lane 8: Clone B8, lane 9: Clone C1, lane 10: Clone C2 (control).

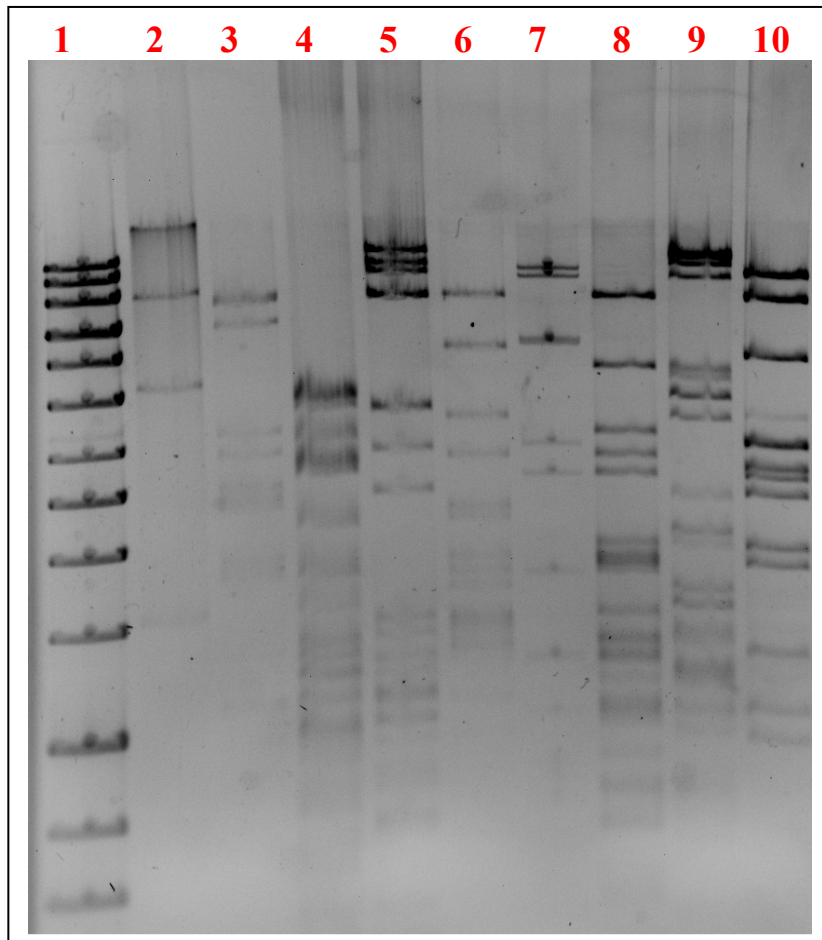


Figure 2.5. Southern Blot analysis of Fosmid clones negative by PCR. Fosmid DNAs were restriction digested with BamHI and hybridized with a heterogeneous KS domain probe. Lane 1, clone C1, Lane 2, clone C2.

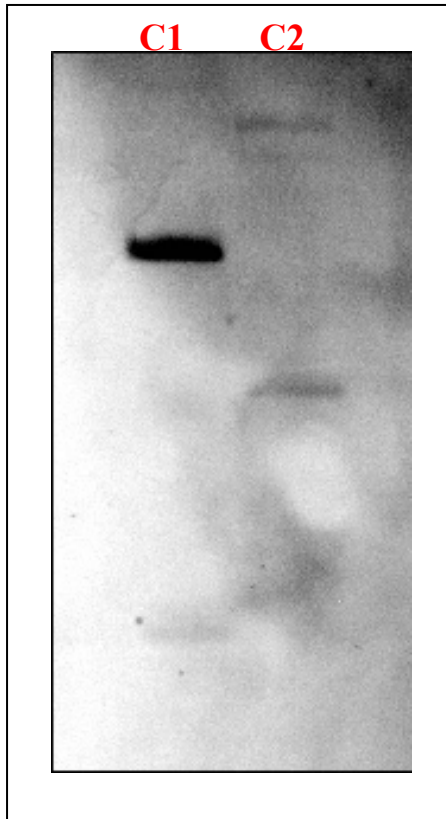


Figure 2.6. Contig derived from fosmid 454 sequencing with homology to a PKS sequence. Primers designed specific to this sequence are highlighted. The top hit from a BLASTx comparison to the GenBank nr/nt database is also shown.

```

tCGACGGCCGCGCCCTTGCCACGCCGCCGCTACGgTAGCCAGTAACTCGACCCTCGC
TTCACCGCTCAGCGATGTCGTGTTGCCGACAACCGGCGATTGTCCGGTTGCAAATTGC
GCGTCGACTtGTCTTGGACCCTTGCCGCAATAAATTGTTGAGCAGTAGGItcGAGGCG
GTCGACCGCTGCTTCCAAATGCGAGTCCGTGTCACGCCAGTACTTCCGCCGCTGCCT
GCGAAAAGCCAGTGAACCGGCGAGCGATCTCCAATCGGATTCGAGCGTCAGCAGCCGC
AGTTTGTGAGCAGCTGATCTCGATACCGCGCTACGATCGCGACGCGTTCGCTGAACT
CGACGCGGCCGACGTTTCGTGTATGGCAAACGTCCGCCAGCTGCCACGAGTCGTCCCG
GGAAAGGTGGCTCGCATATTGGGCGACCATTTGCCGGAGCGCACGCTCGTCGGCGCGC
CGACAGGTCAGAATATGCGCCGTTGGTGCCTCCTCTGAAACCTGGGCGTGTGGCGC
GGCGACTGCTCCAGCACCACATGCGCATTGGTCCCTCCAAACCCAAATGCGCTCa
>gb|AAAF19812.1|AF188287.4 MtaD [Stigmatella aurantiaca] Length=3291
Score = 95.9 bits (237), Expect = 2e-18
Identities = 45/85 (52%), Positives = 60/85 (70%), Gaps = 0/85 (0%) Frame = -3

Query 575 SAFGFGGTNAHVVLEQSPRQHAQVSEDAPTAHILTLSDERLALRQMVAQYASHLSADDS 396
S+FG GGTNAHVVLE++P QV+E H+L+LSAR E AL+ + +YA HL++ +
Sbjct 1867 SSFGIGGTNAHVVLEEPEPGRQVAERERPLHVLSLSARSEVALKALAGRYAQHLASSGA 1926

Query 395 WQLADVCHTTNVGRVFEFSERVAIVA 321
LADVCHT N GR +F R+A+VA
Sbjct 1927 VNLADVCHTANAGRAQFKHRLALVA 1951

```

Figure 2.7. Identification of fosmid clone A12 containing a predicted Myxobacterial MtaD (Hybrid PKS/NRPS pathway) sequence. Multiple primer sets and concentrations of DNA were used to determine that the clone containing the sequence is fosmid A12.

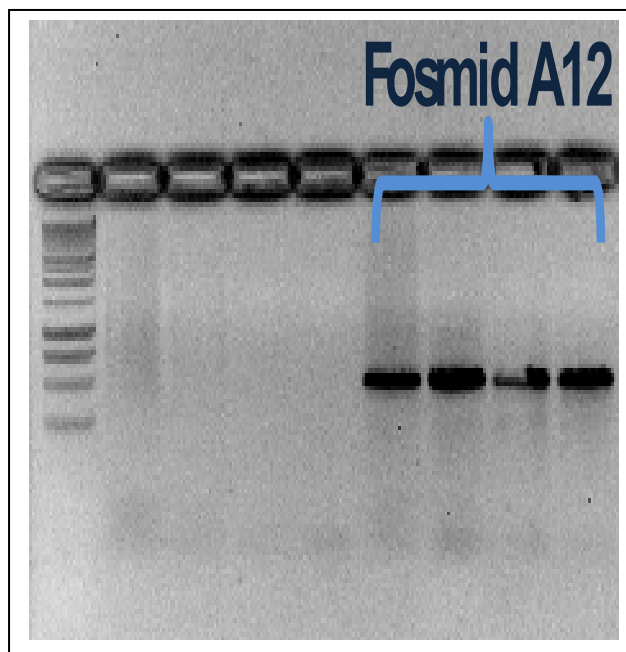


Figure 2.8 Alignment of *Solibacter usitatus* genome with A12 contig sequences. Colored boxes represent regions of homology. Multiple regions of the *S. usitatus* genome have significant identity (i.e., E value < 0.001) to fosmid A12 contig sequences.

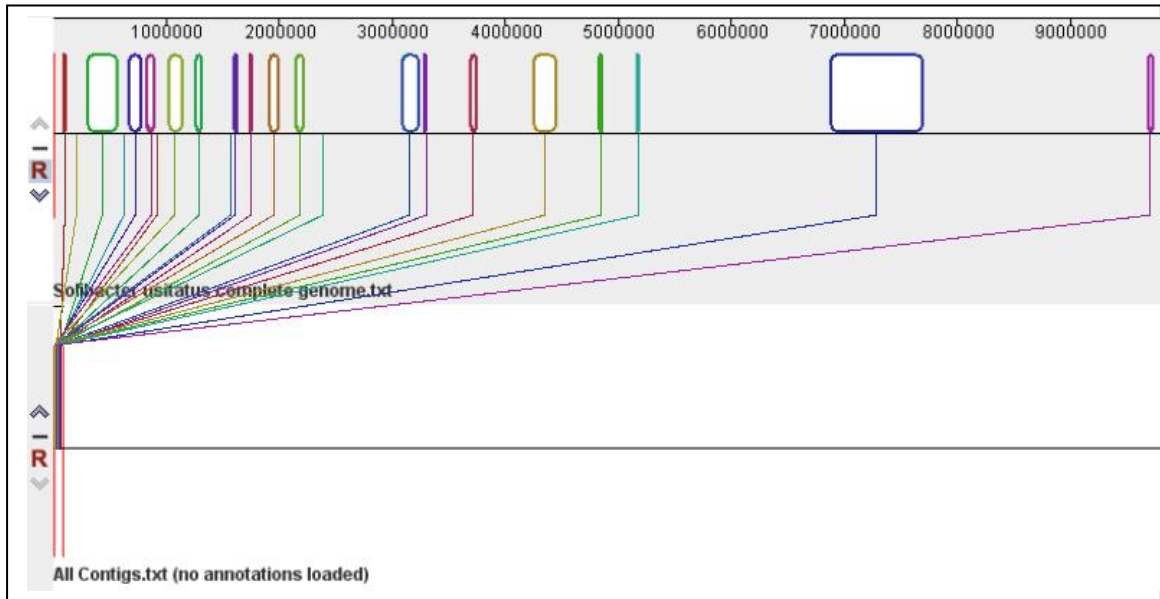


Figure 2.9, Panels A-F. BLASTx analysis of fosmid A12 contig sequences, and comparison with predicted gene products from the *Solibacter usitatus* genome.

A. BLASTx analysis of contig 176992, showing the top hit as the *MtaD* gene of *Stigmatella aurantiaca*

```

>gb|AAF19812.1|AF188287_4 MtaD [Stigmatella aurantiaca]
length=3091

Score = 120 bits (381), Expect = 1e-25
Identities = 58/108 (53%), Positives = 77/108 (71%), Gaps = 1/108 (0%)
Frame = -2

Query  661  SEFFINDRLRSWDQACTV-RRAGVSAEFCPCCTNAEVVLEQSPRQHAQVSEDAPTAEHLILS  405
      E +I N L+ wPAG I RAGVSTFG GTNAPVVIK LF QV - F LILS
Sbjct  1844  SEFYVNIQ SDGPRAGHPRRAGVSTFGIGGTAI VVIRKAPKPRGQVAIRREKFLVLSIS  1903

Query  404  ARDEFALRQMWVAQYASMLGADDSNQLADVCHTINVCRVFSEERVAIVA  301
      AR E RL+ + +YA KL++ + LADVCHT N CR -P R+A+VA
Sbjct  1904  ARSEVALKALACRYAQLASSCAVNLADVCHTANACRAQPKERLAIVA  1951
    
```

B. BLASTx analysis of contig 177416, showing the top hit as the Ton-B dependent receptor of *Solibacter usitatus*

```

>ref|YP_823366.1| G TonB-dependent receptor [Solibacter usitatus E11in6076]
>gb|ABJ83081.1| G TonB-dependent receptor [Solibacter usitatus E11in6076]
Length=1111

GENE ID: 4429732 Acid 2091 TonB dependent receptor
[Solibacter usitatus E11in6076] (10 of lower PubMed links)

Score = 37.4 bits (137), Expect(3) = 7e-13
Identities = 54/104 (52%), Positives = 54/104 (52%), Gaps = 5/104 (4%)
Frame = 12

Query  407  +ATAKQGVVILNFDITQGRS LQIQVQALSRKSGTARYNGSRYGYFSDD4QMAKQI ANKVI  586
      +A+ IF+V I+ F C - A -SC+SS+ G+ Y + R- NS++I A
Sbjct  207  DALQIFKVVQISNFTADTRCTSCAVLNATVRSOSKSTFDGNAYEFLRNSLNRNRTALPCA  266

Query  537  F YANQOIGGALGGPILKDKAHPPISEK RNRFPYIITWY  703
      F G IGGP IK K FFI I I M IIV
Sbjct  267  EKQGLNKANYGASIGGPIKAKW RPNRYDVISQEQAVTISTWY  310
    
```


E. BLASTx analysis of contig 178144b, showing the top hit as hypothetical protein from *Solibacter usitatus*

```

Query[YP_027704.1] G hypothetical protein YP_027704 [Solibacter usitatus E11n6076]
gb|ABJ87419.1| G protein of unknown function YP214 [Solibacter usitatus E11n6076]
Length: 347

GENE ID: 4425384 Acid: 6485 hypothetical protein
[Solibacter usitatus E11n6076] (10 or fewer PubMed links)

Score = 92.0 bits (100%), Expect = 2e-13
Identities = 36/122 (29%), Positives = 40/122 (33%), Gaps = 1/122 (0%)
Frame = +1

Query 13  DRTIMTDRTRKQVYDRITADTRTAIILIKVTAASVSTIANTGLAKSVATVYASQVET 100
             DR 111 E R G L L A TRF I I I I T F S I A A I A C I L V I L E V L E F
Sbjct 670  DENKVLMLPFRNQQLVFRKQASTFRQIQLIDYAKKIALSLRQVGLDYVYKLLV-RKPKAL 730

Query 181  LQVMAFASVTCGFIASKTHDPRDPRDTRDTRDQVQAYVSRKMAIYVQVATATDQAFEA 242
             LQVMA F A S V T C G F I A S K T H D P R D P R D T R D Q V Q A Y V S R K M A I Y V Q V A T A T D Q A F E A
Sbjct 735  LQVMAFASVTCGFIASKTHDPRDPRDTRDTRDQVQAYVSRKMAIYVQVATATDQAFEA 794

Query 314  PR 316
             PR
Sbjct 796  PR 798
  
```

F. BLASTx analysis of contig 178176, showing the top hit as radical SAM domain-containing protein of *Solibacter usitatus*

```

gb|ABJ83399.1| G Radical SAM domain protein [Solibacter usitatus E11n6076]
|tergth-714

GENE ID: 4429393 Acid: 2410 radical SAM domain containing protein
[Solibacter usitatus E11n6076] (10 or fewer PubMed links)

Score = 186 bits (147%), Expect(3) = 4e-52
Identities = 82/89 (92%), Positives = 87/89 (97%), Gaps = 0/89 (0%)
Frame = -1

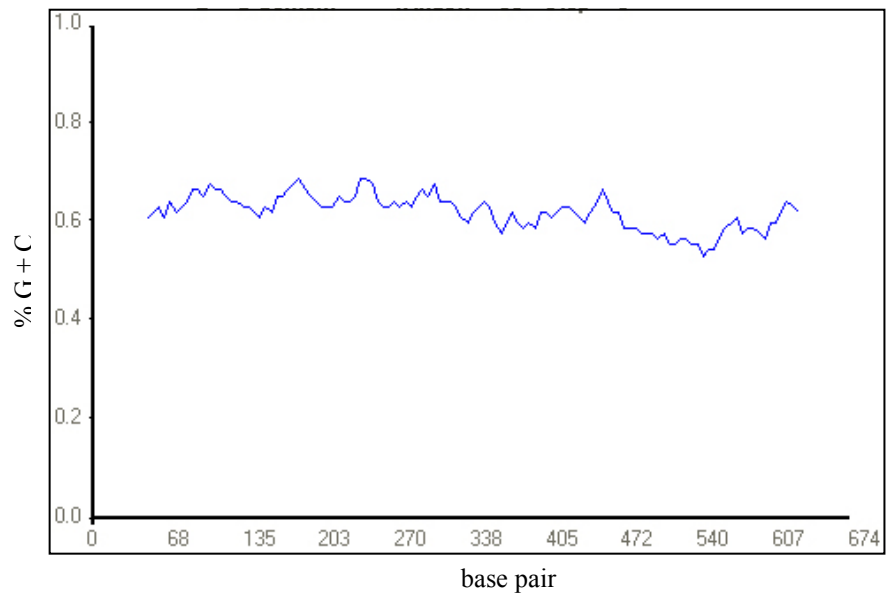
Query 299  CFIHGHIEDVMAIDTAPFKLEEVFPGRDIAENDEKLEKHCSSSTIKHCRGAVLTVDLGN 34
             CFIHGHIEDVMAIDTAPFKLE+VFPGRDI+AIENDEKLEKHCSSSTIKHCRG+VLTVDLGN
Sbjct 111  CFIHGHIEDVMAIDTAPFKLEEDVFPGRDIRAENDEKLEKHCSSSTIVTHCRGAVLTVDLGN 170

Query 93  RNNMCDICFMDANQVGFVHELSDWEDI 7
             RNNMCDICFMDANQVGFVHELSDWEDI+T
Sbjct 191  RNNMCDICFMDANQVGFVHELSDWEDIQT 199
  
```

Figure 2.10 Panels A-F. Percent similarity of fosmid A12 contig sequences to *Solibacter usitatus* genome and G + C content of A12 contig sequences

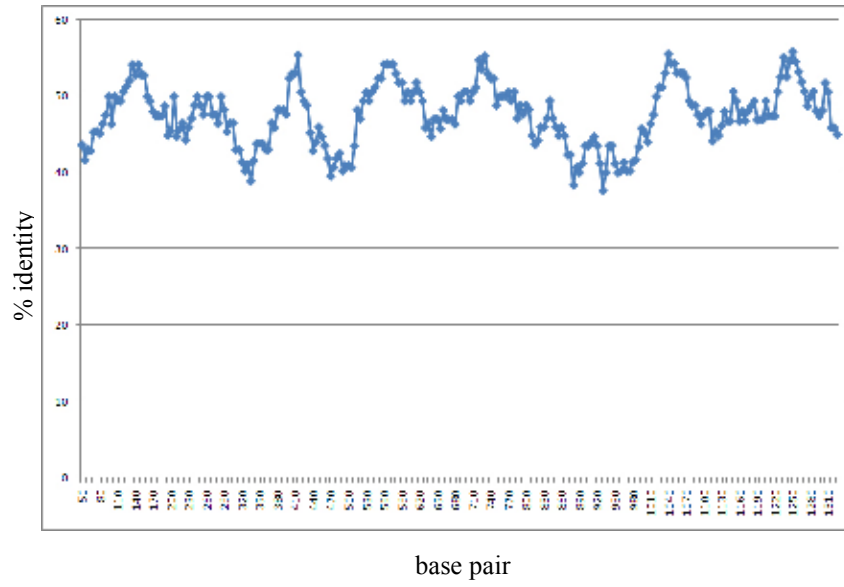
A. Contig 176992 (containing the A12 MtaD homologous sequence) G + C content.

G + C Content

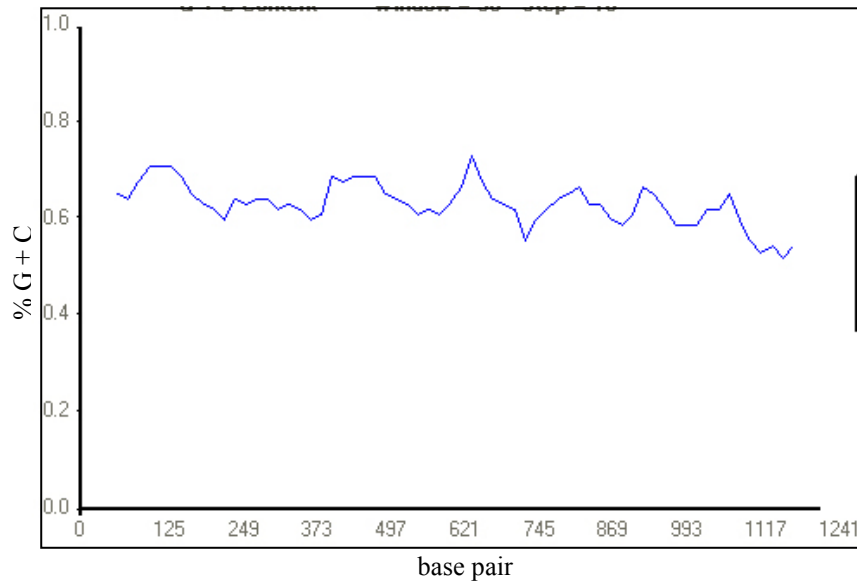


B. Contig 177362 identity to *Solibacter usitatus* genome and G + C content.

Identity Plot

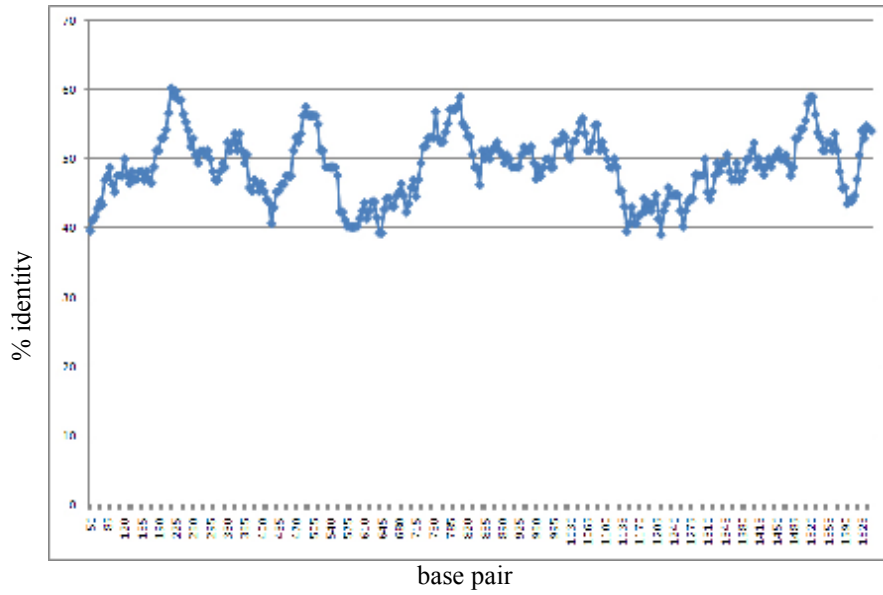


G + C Content of Contig 177362

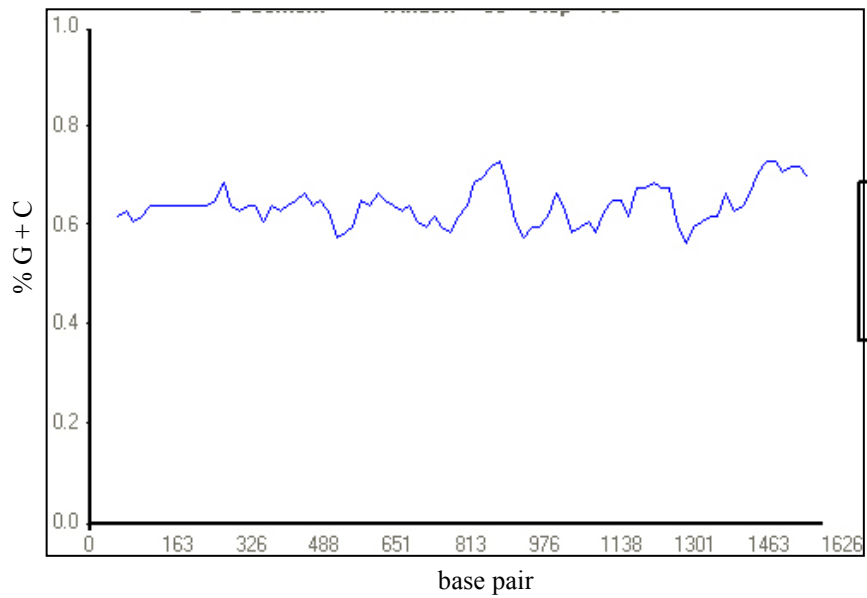


C. Contig 177416 identity to *Solibacter usitatus* genome and G + C content.

Identity Plot

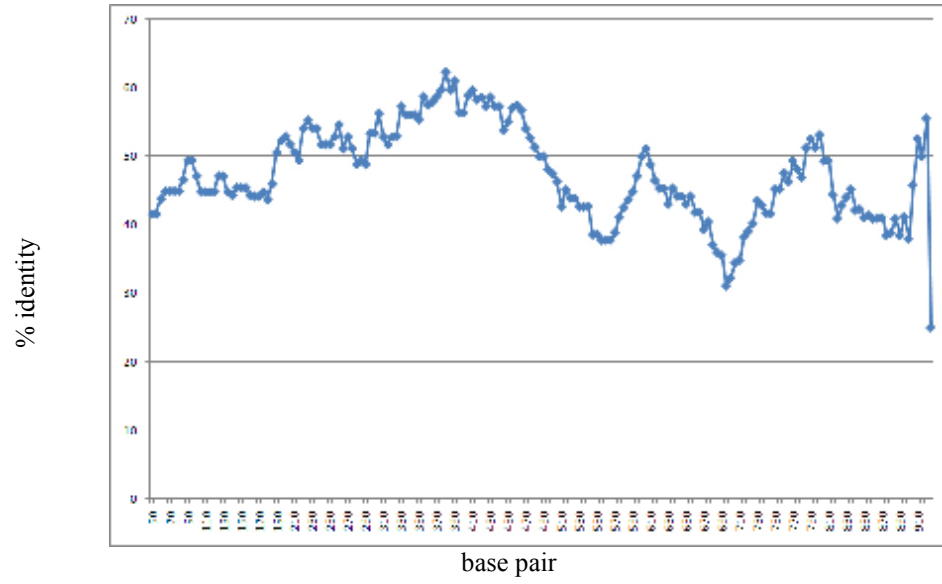


G+C Content of Contig 177416

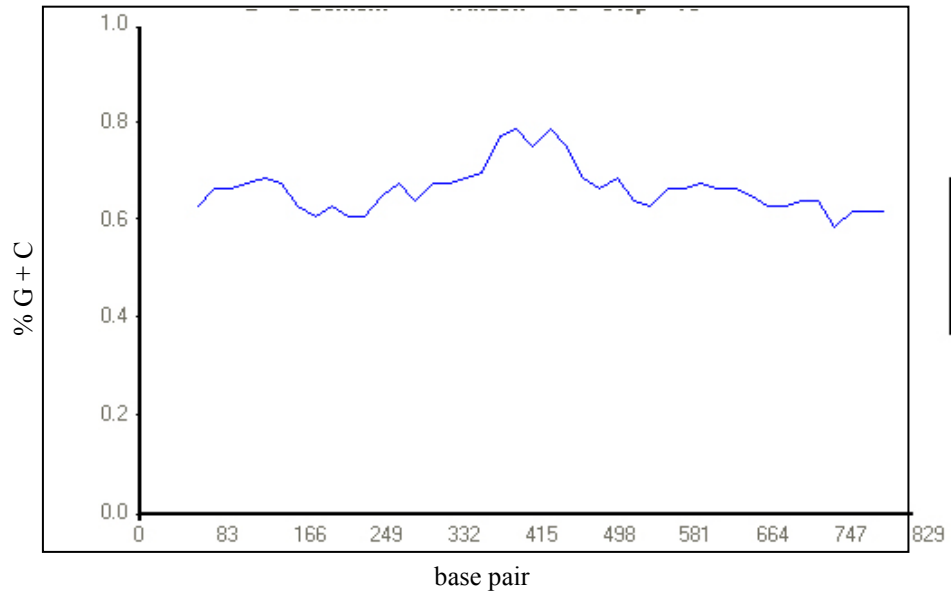


D. Contig 177799 identity to *Solibacter usitatus* genome and G + C content.

Identity Plot

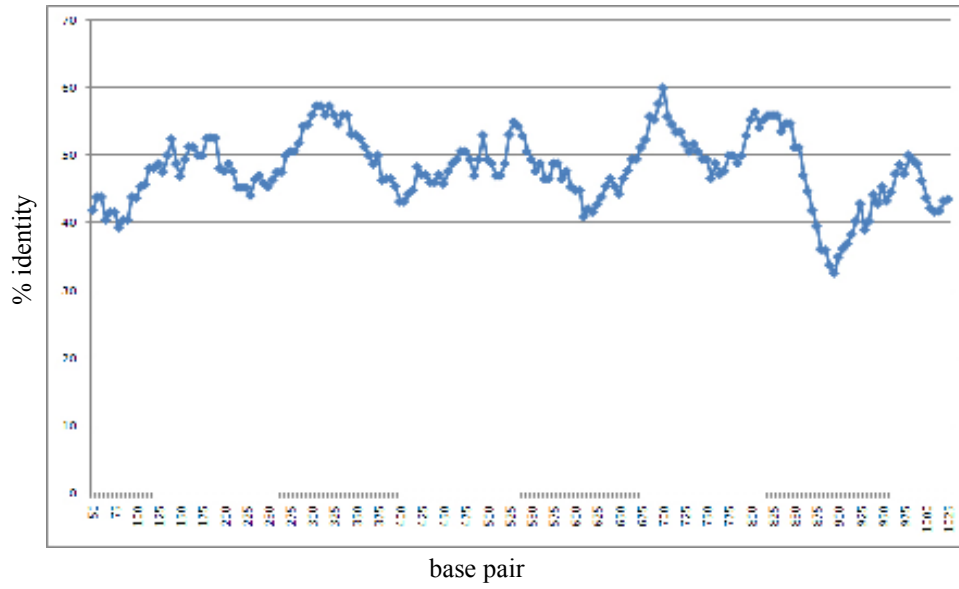


G + C content of Contig 177799

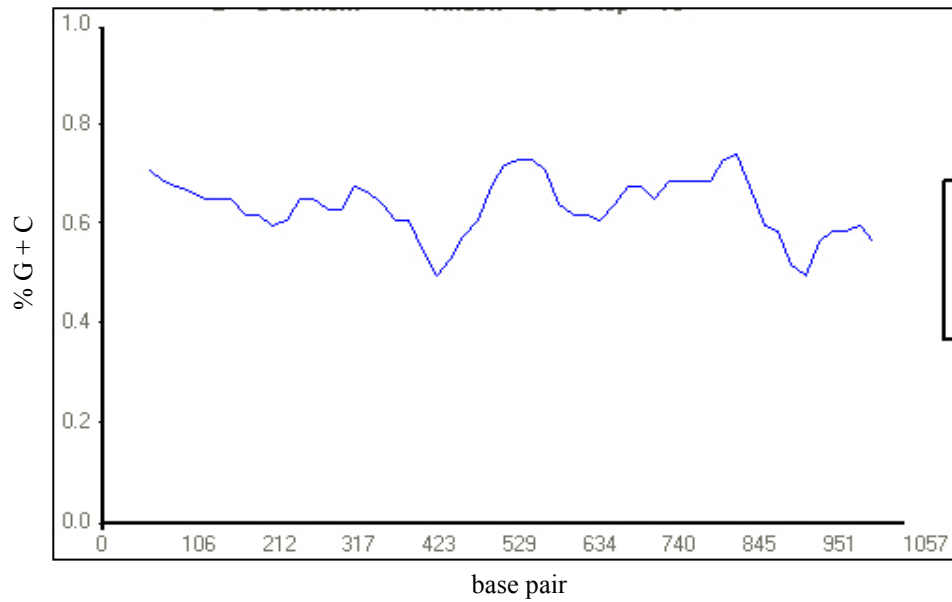


E. Contig 178123 identity to *Solibacter usitatus* genome and G + C content.

Identity Plot

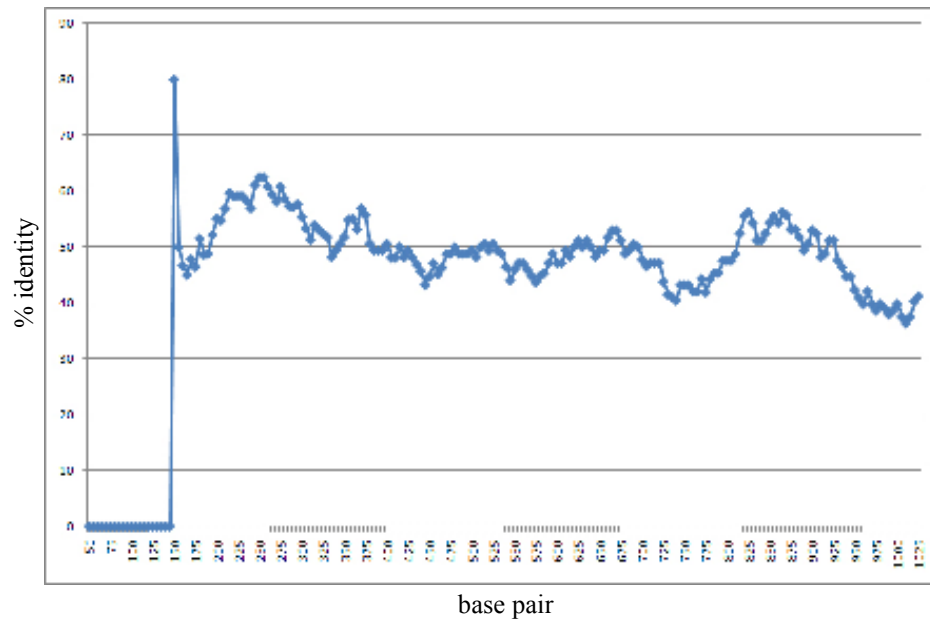


G + C content of Contig 178123



F. Contig 178131 identity to *Solibacter usitatus* genome and G + C content.

Identity Plot



G + C content of Contig 178131

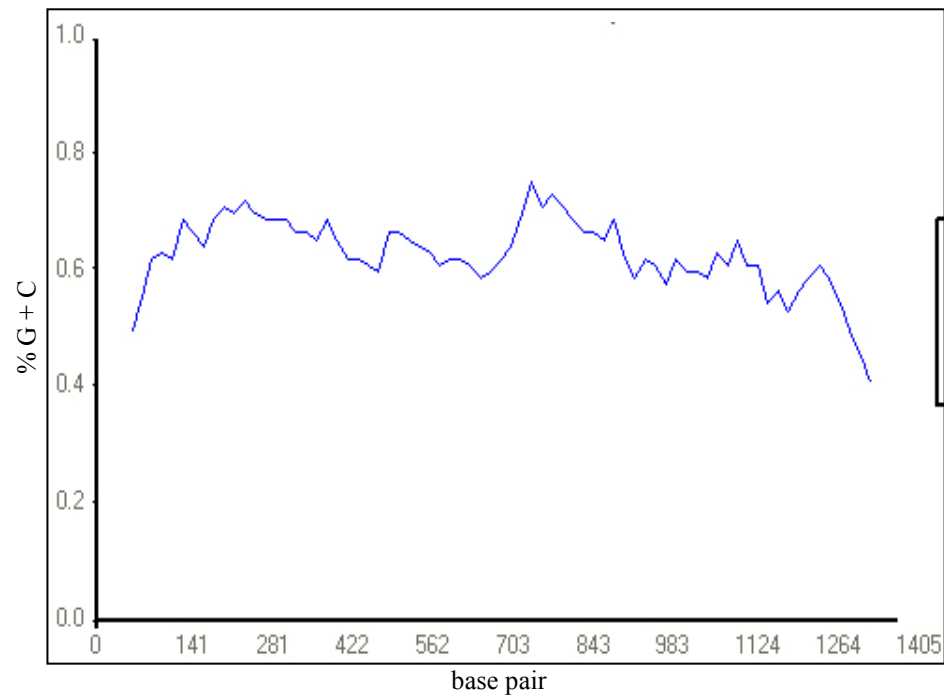


Figure 3.1. Growth inhibition of *A. faecalis*. Observed in the presence of supernatant from BAC clone P15 when the *E. coli* BAC clone culture is grown for 72 hrs in Hsu-Shotts broth, but not with certain P15 transposon mutants. Data represent the OD₆₀₀ of each *A. faecalis* culture, relative to the OD₆₀₀ of *A. faecalis* grown in the presence of supernatant of the *E. coli* negative control (containing empty vector). Cultures were analyzed in triplicate, and error bars reflect standard deviation.

