QUATTLEBAUM, AMY L., M.S. Characterization of Biosynthetic and Catabolic Pathways of *Bacillus subtilis* Strain 168. (2009)
Directed by Dr. Jason J. Reddick. 86 pp.

Bacillus subtilis has long been a model bacterium for understanding biological mechanisms, such as fatty acid catabolism and polyketide biosynthesis. Our interest in the latter was centered on the polyketide synthase (PKS) mechanism responsible for β-branching polyketides. The unique structural moiety is attributed to a HMG-CoA synthase homolog, such as the *pksG* gene in *B. subtilis*.

The first goal was a metagenomic survey of local soils, using the conserved *pksG* homolog sequence as a genetic marker. After optimizing techniques for the extraction and purification of environmental DNA, the β-branching polyketide population was not detected in any local soil samples. While working with a *pksG* homolog, an apparent sequence anomaly prompted us to verify the taxonomic classification of *B. subtilis* research strains ATCC 39374 and 39320. Comparison of DNA sequences (*pksG* homologs, hypervariable regions of 16S rRNA and rDNA) and species-specific genes showed the two ATCC strains are more closely related to *B. amyloliquefaciens*.

A group of genes named the *mmg* operon, able to catalyze fatty acid catabolism, are active only during *B. subtilis* sporulation. Our hypothesis was that, by creating a conditional genetic knockout mutation, the bacterium would be capable of *in-vitro* growth in propionate media. Attempts to subclone portions of *mmg* DNA were unsuccessful. However, the neighboring YqiQ protein was successfully isolated and purified.

CHARACTERIZATION OF BIOSYNTHETIC AND

CATABOLIC PATHWAYS OF *BACILLUS*

SUBTILIS STRAIN 168

by

Amy L. Quattlebaum

A Thesis Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Greensboro
2009

Approved by

_____
Committee Chair

To Toby Hedrick

For your unending love and support

APPROVAL PAGE

This thesis has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair   _____

Committee Members   _____

                                _____

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION TO POLYKETIDE BIOSYNTHESIS

I.A. Polyketides

As one of the most common classes of natural products, polyketides are made by nearly all types of organisms. Produced by bacteria and fungi alike, they are a structurally diverse family of natural products known. The bioactivity shown by many of these secondary metabolites has made them the focus of much scientific inquiry; pharmacological applications range from antibiotic, antiparasite, antitumor, immunosuppressive, and other chemotherapeutic effects.[1] Four examples of these naturally occurring polyketides are shown below in Figure 1.1.

First is rapamycin, an immunosuppressant compound that halts tumor growth by cell-cycle inhibition; this antibiotic is administered to patients receiving organ transplants.[2, 3] Next is soraphen, a fungal, acetyl CoA carboxylase inhibitor that is used as a crop protectant for its ability to control plant pathogens. [4] From *Streptomyces glaucescens*, the decaketide tetracenomycin C antibiotic is a cytotoxic antibiotic. [5] It has broad activity against the actinomycete molecular class. Erythromycin A is a broad-spectrum antibiotic that exhibits therapeutic effects against Gram-positive bacteria via a ribosome growth inhibition mechanism.[6] Initially isolated in 1952 from *Saccharopolyspora erythraea*, it has since become the model for both polyketide antibiotics and their biosynthesis.[6]
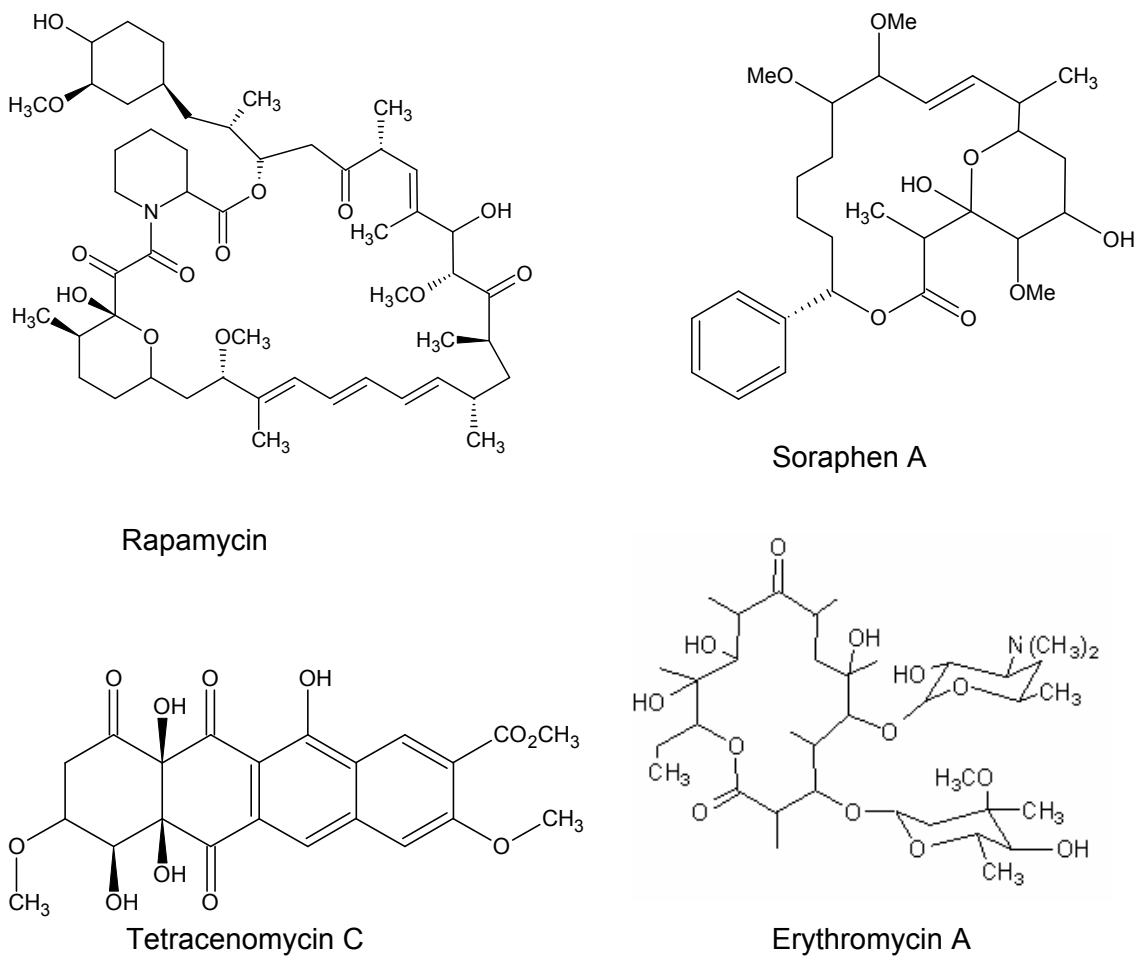
Figure 1.1. Polyketide natural products. [3-6]

Structural diversity among polyketides is elucidated by understanding their biosyntheses. Assembly occurs within polyketide synthases (PKS)– large groupings of multifunctional enzymes, similar to peptide-forming nonribosomal peptide synthetase (NRPS) systems and fatty acid synthases (FAS).[7]

There are three PKS classifications, denoted Type I, II, and III; each has different enzyme arrangements and overall mechanisms.[8] Type I synthases contain multiple modular arrangements of specific enzyme domains, and use each module once for the

production of a single β-keto acyl moiety.[8] Two subclasses exist within Type I PKSs: for bacteria the above system prevails, while iterative fungi use modules more than once for biosynthetic processes.[1]

Three enzyme domains are required for the "minimal" Type I PKS module; these are ketoacyl synthase (KS), acyl transferase (AT), and an acyl carrier protein (ACP) that contains at least one thiolation domain.[9] Polyketide chains are elongated by an acyl addition that is catalyzed by these groups. Three additional common domains that may be present for reduction of the chain are ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER). Lastly, thioesterase (TE) catalyzes either hydrolysis or macrolactonization to release the polyketide product from the acyl carrier protein (ACP) that binds it to a PKS.[8]

Type II PKSs contain synthases arranged among different polypeptides and are known for producing compounds such as tetracenomycin C and aromatic polyketides. This arrangement is similar to fungal synthases, with each synthase used repeatedly during chain elongation.[1] The Type III composition contains single multifunctional modules that are used once, yet these PKSs lack the 4'-phosphopantetheine moeity.[8] The terminal sulfhydryl group of this structure is responsible for essential Coenzyme A utilization in polyketide biosynthesis; it is also responsible for modifying synthase ACPs and enabling them to process growing polyketide chains.[7, 10]

I.B. Polyketide Biosynthesis

I.B.1. Erythromycin

Discovered in soil samples in the Philippines, Erythromycin A (EryA) was first isolated from the Gram-positive bacteria *S. eythraea*; EryA and its derivatives have shown neurotrophic, immunosuppressant, antineoplastic, antiparasitic, and anti-inflammatory bioactivity, along with other therapeutic applications.[11, 12] Its biosynthetic mechanism has become a model for other polyketides for many reasons; one is its utilization of the large, multifunctional 6-deoxyerythronolide B synthase (DEBS) system, and another is the possibility of novel drug design by structural modification. DEBS contains three large proteins, each with multiple enzymatic modules.[13] Stereospecific representations of erythromycin A and its precursor, 6-deoxyerythronolide B (6-dEB), are shown in Figure 1.2.
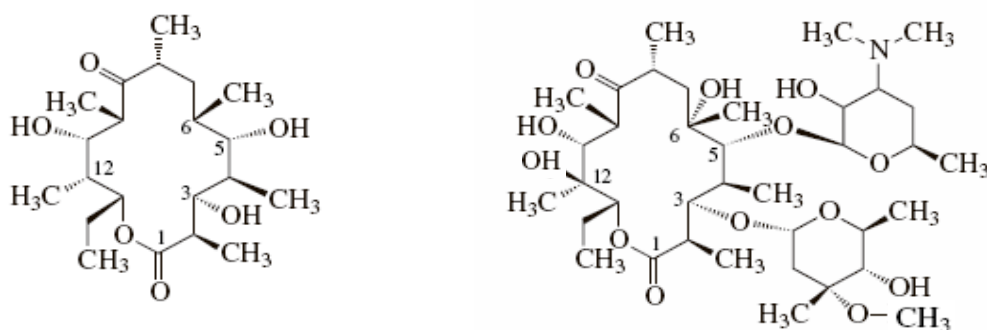
Figure 1.2. 6-dEB precursor and erythromycin A.[12]

As with the majority of Type I PKS gene clusters, a propionyl-CoA starter unit is first loaded onto an acyl carrier protein. Condensation reactions then occur consecutively with six [*S*]-methylmalonyl-CoA derivatives, one per module. Each resulting β-ketoester

group may then be reduced by an enzyme of the PKS; modular organization dictates the resulting polyketide chain structure. The initial 6-dEB macrolide ring is formed via the PKS machinery, after which one P450 enzyme hydroxylates the precursor, and two deoxysugars are attached. Two reactions are needed to complete erythromycin A biosynthesis, and may occur in any order. These are an *EryG*-catalyzed methylation step and hydroxylation by *EryK*, a second P450 molecule.[12]

Propionyl-CoA is typically made available by either carbohydrate or fatty acid catabolism through the tricarboxylic acid cycle; while the loading module also recognizes acetyl-CoA, it is not the preferred substrate for erythromycin biosynthesis.[12, 14] The modules that make up the three erythromycin DEBS proteins (Figure 1.3) catalyze polyketide chain growth and perform cyclization of the end product.

Macrolide biosynthesis: Once the propionyl group is bound to the acyl carrier protein "loading" module, the first [*S*]-methylmalonyl-CoA is attached to a second ACP in module 1 by acyl transferase (AT). The malonyl group is decarboxylated and participates in a Claisen condensation reaction with the propionyl-ACP, catalyzed by β-ketosynthase (KS).[12] This results in a β-ketone group, which is then reduced to an alcohol by ketoreductase (KR) in an NADPH-dependent reaction. Modules two and three repeat the chain elongation with two new malonyl derivatives. In module four the β-carbon is reduced further – a dehydratase (DH) eliminates the alcohol to produce an unsaturated alkene, and, in another NADPH-dependant reaction, enoyl reductase (ER) catalyzes the complete reduction to a fully saturated carbon. Modules 5 and 6 repeat the

chemistry of the 2nd module. 6-dEB biosynthesis is completed by thioesterase (TE) catalysis of ring cyclization and release from the acyl carrier protein.[12]
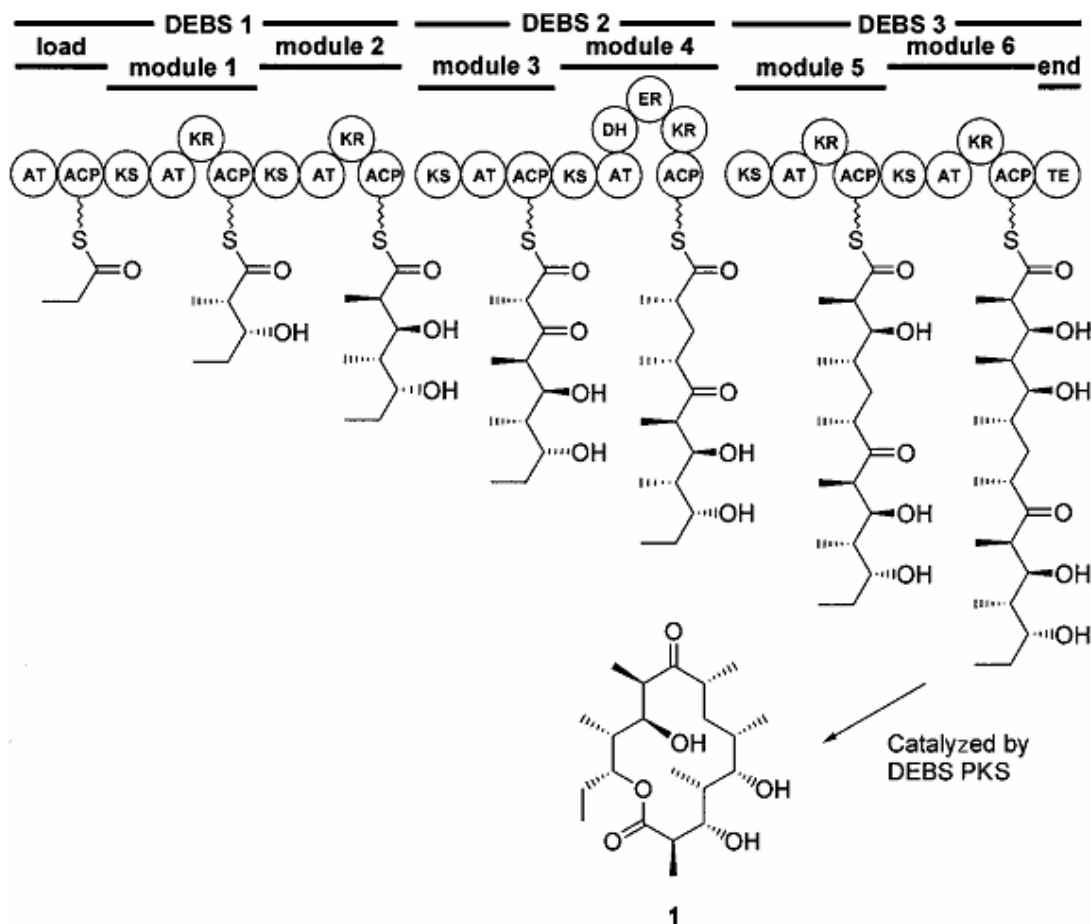
Figure 1.3. Modular biosynthesis of 6-dEB.[14] This is the polyketide precursor to erythromycin.

Erythromycin biosynthesis: The completed synthesis of erythromycin A requires addition of the deoxysugars *D*-desosamine and *L*-mycarose. This occurs by one of two possible mechanisms, both of which result in the production of erythromycin antibiotics EryB, EryC, and EryD.[12] The synthetic scheme are shown in Figure 1.4.
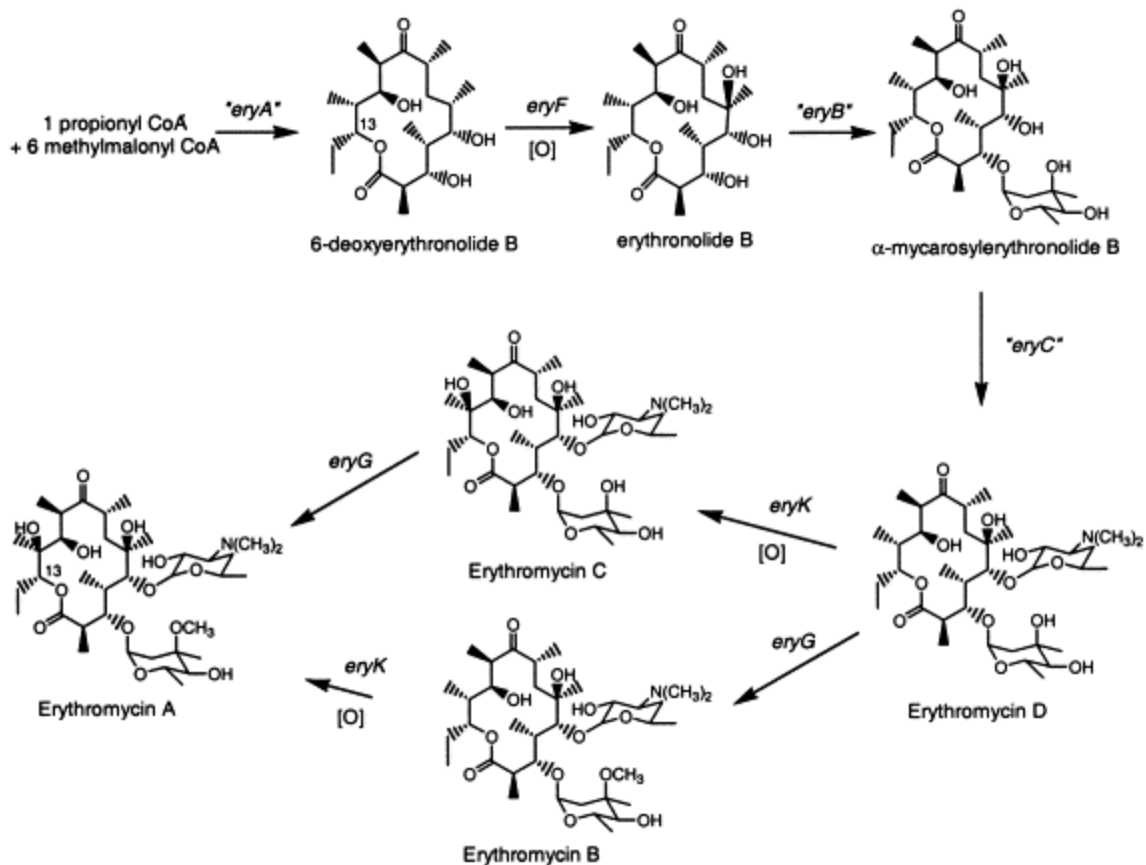
Figure 1.4. Biosynthesis of erythromycin antibiotics.[11, 12]

First, 6-dEB is hydroxylated to erythronolide B by the cytochrome P450 homolog *eryF*, a stereospecific monooxygenase system.[11] The two sugars are then added to this lactone; synthesis and addition of these is believed to be the rate-limiting step in erythromycin production.[12] EryBV, a glycosyltransferase encoded by the *eryB* gene, first catalyzes addition of *L*-mycarose at C-3 to form 3α-mycarosylerythronolide B; EryC desosaminyltransferase further glycosylates at C-5 by addition of *D*-desosamine to produce erythromycin D.[13] At this stage of the synthesis, there are two possible scenarios for the structural modification of erythromycin D to erythromycin A.

Erythromycins B and C are produced by one of two pathways. The two genes involved are *eryG* and *eryK*; these encode for an O-methyltransferase enzyme and a cytochrome P450 homolog, respectively. Methylation by *eryG* at C-3 of the mycarose sugar results in erythromycin B, while *eryK* hydroxylation of this carbon gives erythromycin C.[13] Depending on the order in which these reactions occur, erythromycin A biosynthesis is next completed when either the erythromycin B lactone is hydroxylated or erythromycin C is methylated. While both processes occur, the latter pathway is prevalent; with erythromycin B as its substrate, the catalytic activity of the *eryK* P450 enzyme is 1200-1900 times higher.[13]

This type of polyketide construction has been discovered across multiple species, including Streptomycetes, Nocardioforms, Myxobacteria, Pseudomonadaceae, and other bacteria.[15] In addition to a PKS gene cluster, some organisms also contain non-ribosomal peptide synthase (NRPS) modules; the combination of these two allows for the incorporation of amino acids into an elongating polyketide chain.[16, 17] Following modular biosynthesis, additional modification to structures may also occur, as with the erythromycin deoxysugars.

Analysis of the specific mechanisms of erythromycin biosynthesis has led to an understanding of many polyketide production systems; it is widely applicable, as many Type I PKS gene clusters show homologous processes. In certain cases where natural polyketide products exhibit structural anomalies, there are additional mechanistic explanations to be discovered.

I.C. β-Branching Polyketides

One structural moiety found in naturally occurring molecules that is not explained by the conventional understanding of polyketide biosynthesis is the β-branch. A growing polyketide chain is elongated two carbons at a time by addition of malonyl derivatives. PKS enzymatic domain arrangement dictates one of four possible functional groups at the β-carbon. Table 1 shows the structures normally seen at the keto position of a polyketide, and the modular assembly responsible for each; none of these encode for the β-methyl group.
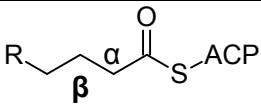
MALONYL DERIVATIVE	BETA FUNCTIONAL GROUP
KS, ACP, AT	
KS, ACP, AT, **KR**	
KS, ACP, AT, **DH**, KR	
KS, ACP, AT, KR, DH, **ER**	
???	

Table 1. Functional groups expected from PKS modular domains.

The β-branch is found in several known polyketides, including the Jamaicamides and curacin A of marine cyanobacteria, the antibiotics TA and mupirocin of *Pseudomonas fluorescens*, which inhibits bacterial cell wall synthesis, all shown in Figure 1.5.[18-22] The keto position of the initial malonyl derivative unit is designated by a black dot, and each β-branch highlighted by a dashed line.

Jamaicamides

Mupirocin

Curacin

Antibiotic TA
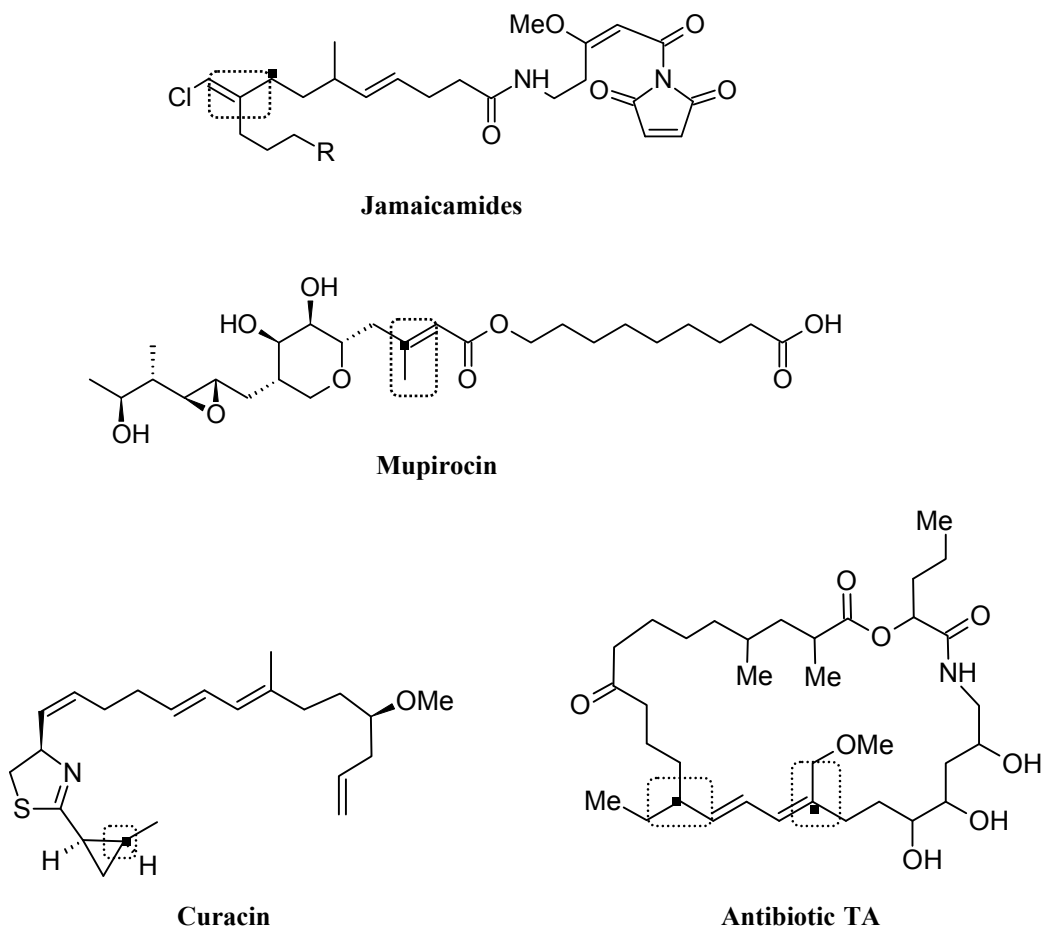
Figure 1.5. Examples of β-branch containing polyketides.[18-22]

Shown first in Figure 1.5 is the general structure of the jamaicamides, isolated from the marine bacterium *Lyngbya majuscula* by the William H. Gerwick research group in 2004.[19] They exhibit sodium channel-blocking activity and cytotoxicity towards both human and mouse cell cancer lines, an example of the pharmacological applications of novel polyketide discovery.[19] Mupirocin has bioactivity towards Gram-positive bacteria and is one of the few antibiotics in use for treatment of methicillin-resistant *Staphylococcus aureus*.[22]

Curacin A, found in the marine cyanobacteria in Curacao, is an antitubulin compound with cancer cell toxicity that affects cell microtubules at the colchicine drug binding site.[18] Its biosynthesis utilizes both the PKS and NRPS gene clusters described above, as does the last compound in Figure 1.5. Myxovirescin, or antibiotic TA, is produced by the myxobacterium *Myxococcus xanthus* when it is under stress from nutrient shortages; one of its polyketide synthases shows strong sequence homology with those of erythromycin and soraphen A.[15]

All sources of the β-branch structural moiety share the presence of large, type I PKS gene clusters. Each of these groups contains similar conserved regions, including a freestanding thiolation domain or ACP, and a 3-hydroxy-3-methylglutaryl (HMG)-CoA synthase. Many of the enzyme-encoding ORFs are homologous, even among seemingly unrelated species. These clusters are unique in that they do not follow the colinearity rule typical of PKS organization.[7, 17] Many also lack acyltransferase (AT) domains within the PKS condensation module; instead a single, separate protein acts as a *trans*-AT to facilitate the loading of monomers onto the synthase ACPs.[22] This class of syntheses are thus called "AT-less" PKSs.[23]

I.D. Difficidin

Four of the *Bacillus subtilis* antibiotics that contain the β-branch structural moiety are difficidin, oxydifficidin, bacillaene, and dihydrobacillaene.[23-25] Difficidin and oxydifficidin are broad-spectrum antibiotics that were first isolated in 1987 from ATCC strains 39274 and 39320 of *B. subtilis*, the most commonly studied Gram-positive bacterium.[26-28] It was assumed for ten years that the biosynthesis of difficidin, shown in

Figure 1.6, arose from a large PKS cluster named *pksX* that was discovered upon sequencing of *B. subtilis* and is represented in Figure 1.7.[16, 29-31]
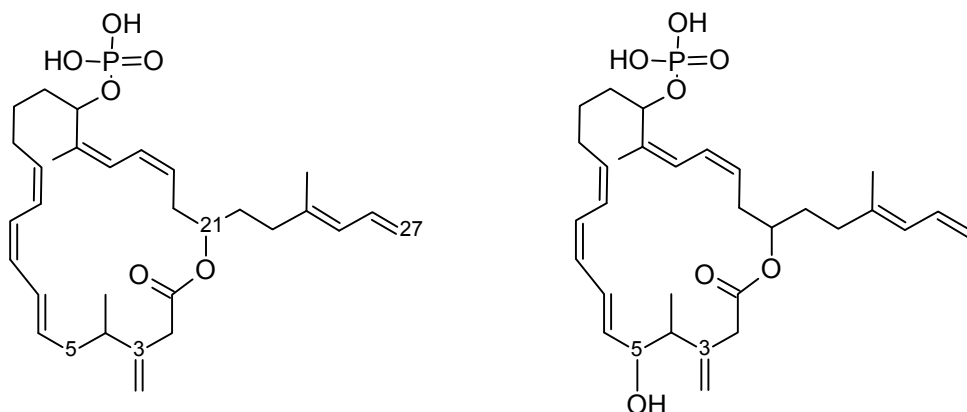
Figure 1.6. Difficidin and oxydifficidin both contain a β-branch on C-3.[32]
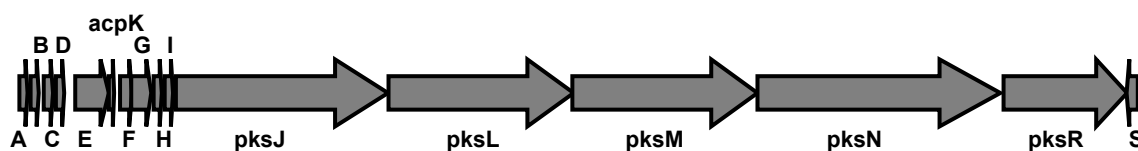
Figure 1.7. The *pksX* operon of *Bacillus subtilis*.[23]

In 1997, *B. subtilis* 168 was one of the first organisms to be fully sequenced, and while it contained *pksX*, the strain is incapable of producing difficidin. An *sfp* gene that encodes for 4'-phosphopantetheine transfer to PKS ACPs is essential for CoA utilization in polyketide synthases, and thus for enzymatic production of the polyketide antibiotic. However, the gene is silenced in B. subtilis 168 (sfp°), and so the two "wild-type" strains, 39374 and 39320, have been considered the models for this bacterium's polyketide production.[33-36]

I.E. Bacillaene

I.E.1. β-branch Biosynthesis

The 16 *pks* genes were discovered in 1995 and make up ~2% of the *B. subtilis* genome.[17, 23] Bacillaene was first isolated from *B. subtilis* in 1995, but not structurally defined until 2007 when it was discovered to be the product of the *pksX* gene cluster.[7, 24] In 2006, it was revealed that a related species, *Bacillus amyloliquefaciens* FZB 42, also contained three polyketide synthase clusters which produced bacillaene, macrolactin, and difficidin/oxydifficidin,.[7] These PKSs are referred to as *pks1(bae)*, *pks2* (macrolactin biosynthesis), and *pks3(dif)*, respectively, and all are "AT-less" type I synthases.[7]

Chen et al. defined the organization of all three and showed the *bae* gene cluster to be a *pksX* ortholog.[7] Both are large PKS/NRPS hybrid complexes, uncommon but highly conserved among multiple *Bacillus* species.[7, 8, 17, 24, 25] Bacillaene is an unstable compound, and while further characterization has proven difficult, the amount of space dedicated to this polyketide machinery indicates possibly important biological functions.[17]

Characterization of *bae*'s ortholog, the conserved *pksX* gene cluster, has made clear the manner of β-methyl inclusion in polyketides.[7, 23, 30, 37] Each ORF within the *pksX* cluster is homologous to enzymes utilized by Type I PKSs (Table 2.)[23] The proteins AcpK, PksC, and PksF-I, along with the tandem thiolation domains of PksL, make up a "gene cassette" that is now noted in all PKS or PKS/NRPS hybrid systems that synthesize a β-branched product.[7, 24, 30] Once the functions of these genes were known, the specific reactions responsible for incorporating the methyl unit was clear.[23]

13

Gene Name	Gene Function Based on Homology
pksA	Transcriptional regulator
pksB	Zinc-dependant hydrolase
pksC-D	AT (acyl transferases)
pksE	AT with flavin mononucleotide-dependant oxidase
acpK	ACP transferase
pksF	Ketosynthase
pksG	Hydroxymethylglutaryl (HMG)-CoA synthase
pksH-I	Enoyl CoA dehydrases
pksJ-R	"AT-less" type I PKSs
pksS	Cytochrome P450 (hydroxylase)

Table 2. Genes of the *pksX* operon and their supposed functions.[7]

The first and last enzymes of a Type I PKS modular domain would typically be ketosynthase and an ACP, per the colinearity rule.[7, 17] The six-protein cluster utilizes a Type II PKS mechanism instead by way of a non-C-terminal KS enzyme. Though lacking an amino acid residue required for Claisen condensation, it is still capable of decarboxylation. Figure 1.8 illustrates the mechanistic scheme for β-branch insertion determined by the Calderone and Walsh research group.[23]
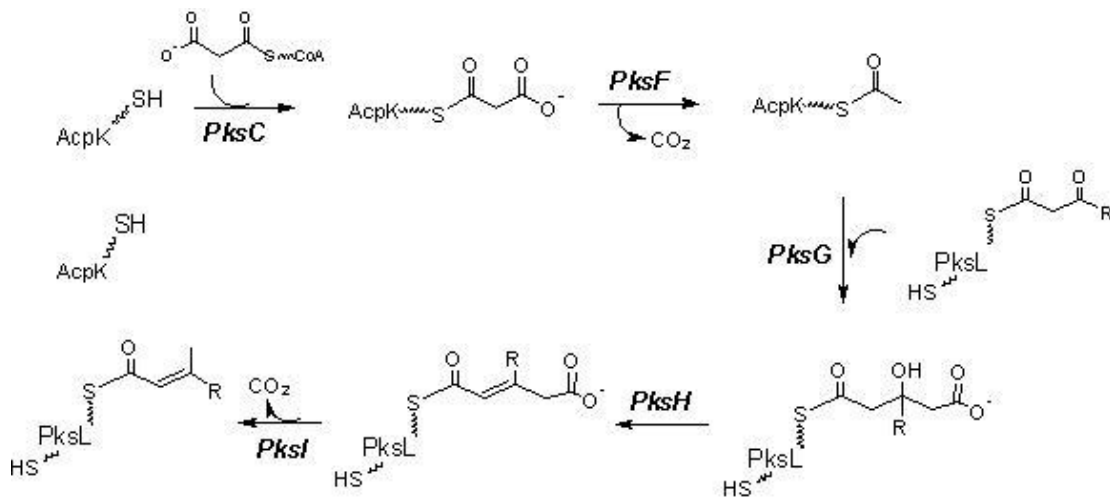
Figure 1.8. Mechanism of β-branch formation.[7, 23, 24]

PksC is an acyltransferase specific for malonyl-CoA, and promotes its transfer to

the sulfhydryl group of AcpK. The PksF ketosynthase then catalyzes the decarboxylation

of the malonyl-ACP to acetyl-ACP. Next the reaction initiated by the HMG-CoA

synthase homolog PksG incorporates the β-methyl moiety into bacillaene through a

critical carbon-carbon bond formation that introduces a single carbon unit for chain

elongation.[23, 24] Condensation occurs between the acetate unit of acetyl-CoA and a β-

ketothioester polyketide intermediary located in PksL, attached to one of its two tandem

thiolation domains.[23] PksH and PksI, both enoyl-CoA hydratases, then catalyze

dehydration of the hydroxyl group and decarboxylation, respectively, forming a β-

methylated structure.[23]

I.E.2. Bacillaene Biosynthesis

The AcpK/PksFGHI gene "cassette" mimics an isoprene biosynthesis; once the product (isoprenyl-S-PksL-2ACP) is constructed, bacillaene is synthesized by the remainder of the *pksX* operon.[18, 19, 23] For both *pksX* and *bae*, the gen clusters involved and the respective portions of the growing bacillaene structure are shown in Figure 1.9.
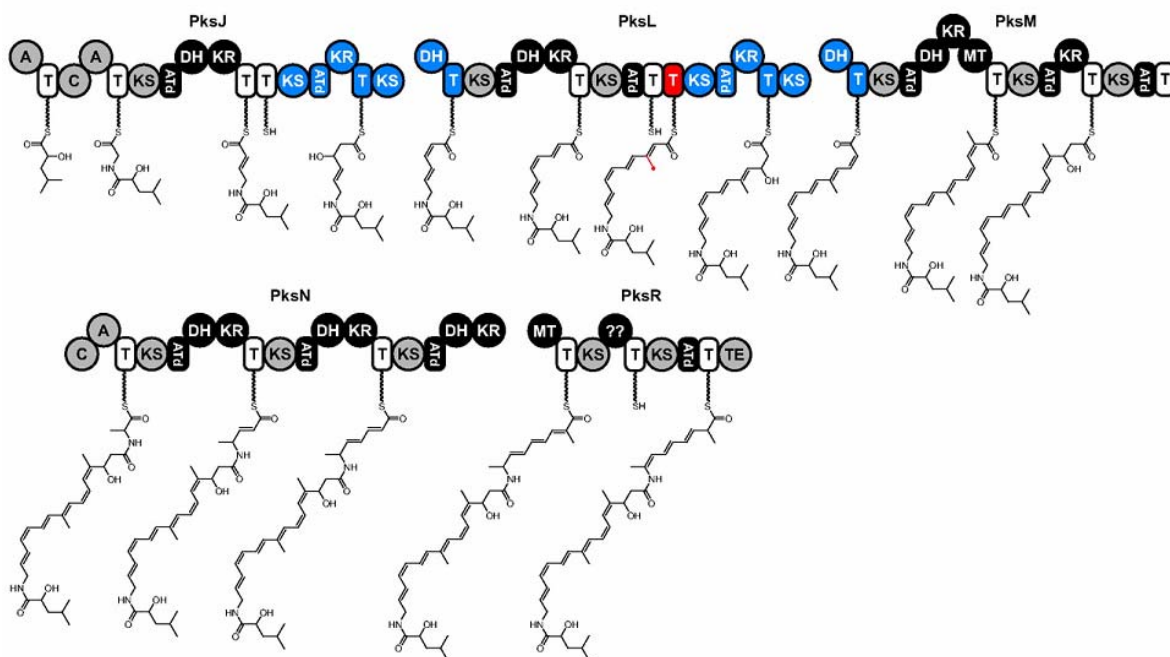
Figure 1.9. Model of bacillaene biosynthesis.[24]

A linear and highly conjugated structure, the antibiotic contains both a hexaene, triene, and the α-hydroxy-isocaproic acid incorporated by PksJ.[24] Multiple stereoisomers of bacillaene are observed by LC-MS; with a reduced double bond between C22 and C23, dihydrobacillaene is the most prevalent (Figure 1.10.)[37]

Dihydrobacillaene
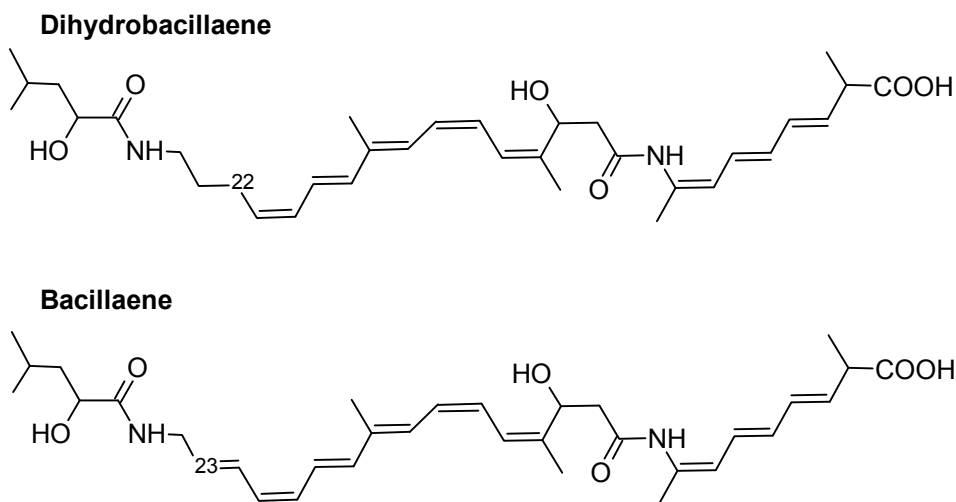
Bacillaene

Figure 1.10. Bacillaene and dihydrobacillaene.[37]

Biosynthesis begins with the PKS/NRPS hybrid, *pksJ*, that lacks four characteristics of a typical synthase: acyltransferase domains to introduce monomers for chain elongation, a β-ketoacyl synthase domain at the N-terminus, a C-terminal ACP, and a conserved aspartate residue in the ketoreductase domain.[7] This absent amino acid, a trait of all PKS/NRPS split module orthologs, is thought to influence the stereochemistry of KR reaction products and so the configuration of any subsequent alterations.[7, 24, 30]

Figure 1.11 is the enzymatic domain structure of *pksJ* and corresponding portions of the growing bacillaene polyketide chain. In the 1st module, an α-hydroxyl-isocaproic starter compound is loaded, and glycine added by the first NRPS domain in module 2.[24, 37] A malonyl unit is added to the chain by ketoacyl synthase in the 3rd module, with the resulting ketone reduced to an alcohol by KR. The alcohol is then eliminated completely by dehydratase, resulting in a *cis*-alkene. Chain elongation continues in the 4th module,
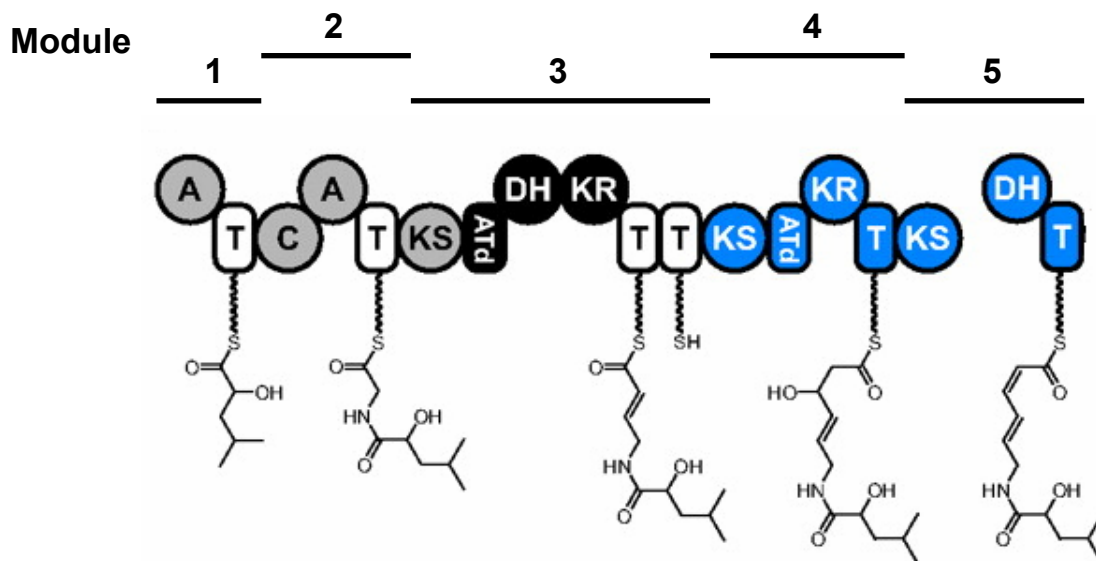
Figure 1.11. Domains of *pksJ* and the beginning of bacillaene biosynthesis. Enzymatic domains include: A (adenylation), T (thiolation), C (condensation), KS (ketosynthase), ATd (acyltransferase docking), DH (dehydratse), and KR (ketoreductase). [24, 37]

while the 5th module is shared with *pksL*. This is a split module in which both proteins

contain an incomplete enzymatic domain; by interfacing at these domains, the two

proteins share the work of a single module.[24]

CHAPTER II

PKSX OPERON OF *BACILLUS SUBTILIS*

II.A. Environmental DNA

In recent years there has been a renewed focus on natural products as a vastly untapped resource for novel antibiotics. Although an estimated ≤1% of microbes worldwide have been studied in depth, they have yielded greater than 17,000 biologically active molecules.[38] While the remaining 99% is unculturable by known methods, it must certainly contain a large number of undiscovered products with biological activity; any number of these could be a source of novel pharmacological and/or other clinical applications. Despite difficulties related to culturing these microbes, the isolation of bacterial nucleic acids, or metagenomic DNA (eDNA), directly from environmental samples is possible and relatively simple.

Biological sciences often rely on bacterial models to elucidate the specifics of biological processes. Polyketide metabolites, observable throughout nature, possess both structural and bioactive diversity. Discovery of these individual compounds began decades ago; it is now known that varying strains of a single bacterial species, such as *Bacillus subtilis*, actinomycetes, and streptomyces, are capable of producing more than two dozen antibiotics.[25] Although diverse, highly conserved sequence conservation exists among specific metabolites. Consequently, metagenomic methods which survey

the environment for the sequences responsible for polyketides can yield a large amount of data.

Despite this opportunity, reconstituting the biosynthesis of polyketides from eDNA has not been successful. However, technological and/or theoretical advances may yet open the way for a more complete survey of this environmental microbe assembly. DNA sequencing has become a routine and readily available research method, with completed genomes accessible for scores of organisms. Isolated yet unidentified DNA sequence regions may now be assessed through sequence homology, giving the researcher immediate feedback and a narrowed gene target for further development.

Numerous methods have been described and utilized to extract and purify eDNA; this is a reflection of the variety of environmental samples available, as well as their chemical complexity. Nucleic acid isolation may be done in several ways, yet further applications using this material can be difficult because of factors such as impurities. Other potential limitations arise from shearing by harsh lysis methods, DNA adsorption to soil surfaces, and the coextraction of pollutants. The most prevalent soil contaminants are humic substances and tightly bound soil colloids, both making DNA detection and measurement problematic and unreliable.[39, 40]

Due to these obstacles, successful eDNA isolation is not sufficient for metagenomic studies – further purification of the sample prior to usage in the laboratory is vital. As with extraction procedures, there are a number of described methods. Efficiency and yield of this step helps determine the success of future processing (i.e. by

polymerase chain reaction (PCR), restriction enzyme digestion, and subcloning efficiency.)[40]

II.B. Goals

We hypothesized that the biosynthetic genes responsible for the β-branch subclass found in *Bacillus* and other known bacteria were sufficiently widespread to be useful as highly specific genetic beacons for new polyketide natural products. By using routine laboratory methods, we would be able to dectect *pksG*-system homologs in local environmental samples. These soils should contain nucleic acids from a large host of organisms, and we expected that these were produced by unknown organisms as well as the known *Bacilli* species.

The first objective of this research was to determine optimal techniques for the extraction and purification of bacterial DNA from local soils. Once achieved, the second goal was to obtain environmental genetic data for the β-branching subclass of polyketides using existing analytical techniques. This would be done by specifically using the PCR method to survey for the highly conserved *pksG* homologs described above. Sequencing of these may yield insight into the biosynthetic mechanisms afforded by the conserved regions.

II.C. Results

Bacterial nucleic acid extraction and purification were done by various methods as described herein, with the resulting products used as the template DNA for a PCR. Once isolated from soil, DNA may be processed further only after all contaminants are removed. PCRs are thus a rapid way to evaluate the purity of metagenomic extracts; the

amplification of multiple target sequences should also provide a means to estimate their prevalence in various collected soils.

Two PCRs were performed for each eDNA isolate. The first was amplification of a 590 base pair (bp) ribosomal RNA sequence, and the second targeted a DNA sequence that codes for a conserved amino acid region of *pksG* homologs shown and highlighted in Figure 2.1. Once amplification was substantiated by agarose gel electrophoresis, the PCR products were purified with the QIAquick® PCR Purification Kit from QIAGEN Sciences (Valencia, CA). The target sequences were then subcloned into the *Eschericia coli* pBAD vector via topoisomerase cloning methodology (Invitrogen Corporation). This vector confers antibiotic resistance, and so all agar plates and cultures used thereafter (except where noted) contained 100 µg/mL ampicillin.

```
PksG      ----------------------MVSAGIEAMNVFGGTAYLDVMELAKYRHLDTARFENL      37

JamH      ----------------------MQSVGIEALSVYGGAAQLDVRKLVQARQLDISRFDNL      37
CurD      MPSAIALLLAASIQLYYWARSKAMQQVGIEALSVYGGAAQLELRKLAQARQLDISRFDNL      60
TaC       ----------------------MGPVGIEAMNAYCGIARLDVLQLATHRGLDTSRFANL      37
PedP      ---------------------MPVGIEAMNLYGGSAFVDVMELAVYRKHDLSRFDNL       36
MupH      ----------------------MTRQVGIEAMNVFGGAASLDVTMLARHRKLDSQRFGNL      38


PksG      LMKEKAVALPYEDPVTFGVNAAKPIIDALSEAEKDRIELLITCSESGIDFGKSLSTYIHE      97
JamH      MMKEKAVSLPYEDPVSYAVNAAKPIIDRLSNADKQRIEMVITCSESGIDFGKSMSTYIQE      97
CurD      MMKEKAVSLPYEDPVSYAVNAAKPIIDRLSDADKQRIEMVITCSESGIDFGKSMSTYIQE      120
TaC       LMEEKTVPLPYEDPVTYGVNAARPILDQLTAAERDSIELLVACTESSFDFGKAMSTYLHQ      97
PedP      LMKQKTVALPYEDSVSFAINAAQPLIDTLSETEKERIELLITCSESGIDFGKSISTYIHH      96
MupH      LMHEKSVALHSEDPVTFAVNAAKPLIDALSPQERDQIELLITCTESGIDFGKSVSTYVHD      98


PksG      YLGLNRNCRLFEVKQACYSGTAGFQMAVNFILSQTSPGAKALVIASDISRFLIAEGGD-A      156
JamH      YLGLSRNCRMFELKQACYSGTAGLQMAINLILSQTFPGAKALVIATDISRFLVAEGGE-A      156
CurD      YLGLSRNCRMFELKQACYSGTAGLQMAINLILSQTFPGAKALVIATDISRFLVAEGGE-A      179
TaC       HLGLSRNCRLIELKSACYSGVAGLQMAVNFILSGCSPGAKALVVASDLSRFSIAEGGD-A      156
PedP      YLGLNRNCRLFELKQACYSGTAGLQMAVNFIVSKVSPGAKALVIASDLSRFLVTRDEEGE      156
MupH      YLGLKKNCRVFELKQACYSGTAGFHSAVQFVLSQTSPNAKALVVAVDLCRFLMIGEGG-A      157


PksG      LSEDWSYAEPSAGAGAVAVLVGENPEVFQIDPGANGYYGYEVMDTCRPIPDSEAGDSDLS      216
JamH      LSEDWSFAEPSSGAGAVALLVSDTPYIFQVDVGCNGYYGYEVMDTCRPNPDSEAGDADLS      216
CurD      INYDWSFAEPSSGAGAVALLVSDTPHIFQIDVGCNGYYGYEVMDTCRPNPDSEAGDADLS      239
TaC       STEDWSFAEPSSGAGAVAMLVSDTPRVFRVDVGANGYYGYEVMDTCRPVADSEAGDADLS      216
PedP      LTMDWDFFEPSGGAGAIAMLVSEKPDIFELDIGANGYYTYEVMDTCRPIADSEAGNADLS      216
MupH      AAQDWAFSEPSGGAGAVAMLISDKPHVFSLDIGASGYYSFEVMDTCRPEPDSEAGDADLS      217


PksG      LMSYLDCCEQTFLEYQKRVPGANYQDTFQYLAYHTPFGGMVKGAHRTMMRKVAKVKTSGI      276
JamH      LLSYLDCCENAYRHYQNRVEGVDYQESFDYLSFHTPFGGMVKGAHRSMMRRLKRAKPADI      276
CurD      LLSYLDCCENAYRHYQNRVEGVDYRESFDYLSFHTPFGGMVKGAHRNMMRRLKRAKPAEI      299
TaC       LLSYLDCCENAFREYTRRVPAANYAESFGYLAFHTPFGGMVKGAHRTMMRKFSGKNRGDI      276
PedP      LLSYLDCCDQSFKEYKKRVFKVDYCDTFQYLCFHTPFGGMVKGAHRNMMRKYSDLSREAI      276
MupH      LMSYLDCCEQTFNEYRLRVDDVDYARTFDYLAFHTPFGGMVKGAHRLMMRKFAPADPPAI      277


PksG      ETDFLTRVKPGLNYCQRVGNIMGAALFLALASTIDQGRFDTPKRIGCFSYGSGCCSEFYS      336
JamH      EADFQRRVMPGLVYCQQVGNIMGATLFLSLASTIDNGDFSTPRRIGMFSYGSGCCSEFYS      336
CurD      EADFQRRVMPGLVYCQQVGNIMGATLFLSLASTIDNGDFSTPRRIGMFSYGSGCCSEFYS      359
TaC       EADFQRRVAPGLTYCQRVGNIMGATMALSLLGTIDHGDFATAKRIGCFSYGSGCCSSEFFS      336
PedP      EADFQQRVQPGLTYCQRVGNIMGGTIMLSLASTIDNAIMRNTCRVGCFSYGSGCCSEFYS      336
MupH      EADFAQRVAPGLAYCQRVGNIMGATMFLSLAGIISTARFDEPRRIGCFSYGSGCCSEFFS      337


PksG      GITTPQGQERQRTFGIEKHLDRRYQLSMEEYELLFKGSGMVRFGTRNVKLDFEMIPGIMQ      396
JamH      GVVTPEGKQIQAQQGISSQLADRYSLSMEEYEQLLYHSSAVAFGTRNVTLDYQLFPGVWK      396
CurD      GVVTPEGKQIQAQQGISAQLADRYSLSMEEYEQLLYHSSAVAFGTRNVTLDYQLFPGVWK      419
TaC       GVVTEEGQQRQRALGLGEALGRRQQLSMPDYDALLKGNGLVRFGTRNAELDFGVVGSIRP      396
PedP      GVVSPSGQAKLRDWALGKHLDSRHRLTMEQYEIVLQRSELVKFGTRNVVLSNDLIPGAWK      396
MupH      GTVTPQSRDHLSAQAIAQQLDSRYSLSLAEYESILLGNDAVRFGTRDHTVACPVTDSVIA      397


PksG      ---------STQE-KPRLFLEEISEFHRKYRWIS                           420
JamH      ---------QIAG-KGRLVLKAIKEFHREYEWV-                           419
CurD      ---------KIAG-KGRLVLKAIKEFHRKYEWV-                           442
TaC       ---------GGWG-RPLLFLSAIRDFHRDYQWIS                           420
PedP      TITRNVRENKDIDIGKRIFLSEIKEFHRHYEWIE                           430
MupH      ----------AGGLSGKLLLSAINGYHREYSFQP                           421
```

Figure 2.1. Targeted conserved amino acid sequence of *pksG* homologs.

Once target sequences were ligated, the plasmid was transformed into "One Shot®

TOP10 Chemically Competent E. coli Cells." Transformants were grown overnight in

liquid cultures and screened by PCR for the insert. If a product of approximately 600 bp

was amplified, cloning was considered successful. As a final step, DNA sequencing was

performed to verify the target insert; this was done by either SeqWright, Inc. (Houston,

TX) or the Institute for Health, Science, and Society (UNC Greensboro).

II.C.1. Target Bacterial Regions

The incidence of *pksG* homologs in any bacterial population is likely minor, thus

a positive control reaction was needed to validate each step of the experiment. A small

16S ribosomal RNA (rRNA) region is found in all forms of life and, among related

species, its sequence contains both highly conserved and hypervariable regions. In order

to assess the presence, concentration and purity of eDNA prior to *pksG* screening, a

conserved 16S region was targeted for PCR.

Bacterial (BAC) 16S primers (Table 2) were designed to amplify this 590 bp 16S

rRNA region. PCRs, using *B. subtilis* ATCC strains 39374/39320 genomic DNA as a

template, were performed to confirm primer viability; each produced a band at ~600 bp,

verifying that the target sequence was a reliable positive control. Listed at the bottom of

Table 2 are universal 16S rRNA primers, designated 16S For and Rev, obtained from

Integrated DNA Technologies (Coralville, IA).

Primer	Sequence
pksG degen UP	5' – CTN CCN TAY GA**R** GAY CCC GT – 3'
pksG degen DN	5' – NCK RTG NGC NCC YTT NAC CAT – 3'
pksG deg2 UP	5' – ACA CGG TTT GAC AAT CTA ATG ATG VRN SAR AA – 3'
pksG deg2 DN	5' – CAN TAC CTR TGA ACA GCT GGC TAG GGT – 3'
BAC 16S Up	5' – AAG TCG AGC GGA CAG ATG – 3'
BAC 16S Dn	5' – CCA GTT TCC AAT GAC CCT CCC G – 3'
Universal 16S For	5' – AGA GTT TGA TCC TGG CTC AG – 3'
Universal 16S Rev	5' – ACG GCT ACC TTG TTA CGA CTT – 3'

Table 3: Primers used for metagenomic survey PCRs. Abbreviations used: N (any nucleotide), Y (any pyrimidine), and R (any purine).

These additional rRNA primers were designed to augment a 570 bp region, and meant for a secondary control reaction, but resulted in no amplification from eDNA or strains 39374/39320.

Degenerate *pksG* primers were originally designed by Edwards et al., and used here with one change in primer sequence.[19] The corrected forward primer, *pksG* degen UP, reflects coding DNA for the highlighted "R" in place of a reported "Y." Degenerate primers target specific amino acids, and for PCR these must be translated into complementary DNA sequences for amplification. Where multiple triple codons may encode a single amino acid, PCR primers account for this by including general pyrimidine or purine nucleotides. The highly conserved LPYEDPV sequence was the forward primer target; these are the ~46th – 52nd residues of *pksG* homologs. The

downstream primer, *pksG* degen DN, was similarly designed to target DNA coding for the highly conserved MVKGAHR amino acid sequence that makes up the ~57th – 62nd residues of *pksG* homologs.

II.C.2. eDNA Isolation

Three direct extraction techniques were tested on soils collected from various areas: wooded and cultivated locations adjacent to the Sullivan Science building of UNC Greensboro (Greensboro, NC); a forest, field, and dried creek bed in Greensboro, NC park; sand-like sediment from a stream in Falls Park (Greenville, SC); and finally, industrial potting soil.   Two commercially available DNA extraction kits were utilized: the EPICENTRE® Soilmaster™ DNA Extraction Kit (Madison, WI) and the PowerMax™ Soil DNA Isolation Kit from MO BIO Laboratories, Inc. (Carlsbad, CA). A third method was a sodium dodecyl sulfate (SDS)-based extraction described by Jizhong Zhou et al. in February 1996.[40]

EPICENTRE® Soilmaster™ DNA Extraction: EPICENTRE® required 100 mg of soil for a 300 µL yield of eDNA isolate. Soil was processed per manufacturer's instructions, and initial attempts to amplify 16S and *pksG* were unsuccessful. Alterations to the procedure included: omission of vortexing steps during extraction to prevent DNA shearing, incubating on ice instead of at room temperature during precipitation, and extended centrifugation. Despite these adjustments, ***no PCR reaction led to amplification following eDNA isolation with the EPICENTRE® kit.***

SDS-Based Extraction: Another isolation attempt used a multi-step, SDS-based approach: An initial 5–10 g soil sample yielded greater than 10 mL of isolate solution.

The final yield was an amber-colored 500 µL solution. This coloration was likely due to the presence of inhibitory humic substances, and so no PCR was attempted on the isolate. Instead, two 100 µL aliquots were purified with commercially available kits, each yielding 50 µL. These were the QIAquick® PCR Purification Kit, and the Wizard® PCR Preps DNA Purification System from Promega Corporation (Madison, WI). An agarose gel of the purified samples showed bands above 10 kbp; this indicated DNA presence, but all PCR attempts were unsuccessful. ***SDS-extracted eDNA was dilute and contaminated; even after purification, it was unsuitable for PCR amplification.*** This prompted use of a third bacterial DNA extraction method.

MO BIO PowerMax^{TM} Soil DNA Isolation: PowerMax^{TM} required an initial 5–10 g soil sample and yielded 5 mL eDNA isolate solution. Manufacturer protocol recommended the concentration of 1 mL eluent by ethanol (EtOH) precipitation before further processing. This was done and the eDNA resuspended in 10 mM Tris (pH 8.0) to a final volume of 100 µL. On an agarose gel, the bright band above 10 kbp indicated high levels of DNA. Initial PCRs were performed with *Taq* Master Mix polymerase; although one gel showed faint bands for 16S and *pksG* at ~600 bp, all attempts at subcloning failed, as did efforts to repeat amplification.

As *Taq* is particularly sensitive to humic contaminants, PCR was repeated using a High Fidelity Phusion polymerase.[40] The *pksG* reaction showed a band near 600 bp; it was purified, subcloned, and transformation into TOP10 cells resulted in the growth of three colonies. Each was screened for the *pksG* insert; only one PCR resulted in a band but, at more than 1 kbp, was not the appropriate target. Sequencing attempts for this

product were unsuccessful, and it was determined that the PowerMax^{TM} enabled

sufficient amplification but not further processing. ***However, as PowerMax^{TM} extraction***

of eDNA resulted in successful pksG and 16S PCRs, it was used thereafter, unless

otherwise noted.

II.C.3. PCR Optimization

Despite amplification, agarose gel bands of 16S and *pksG* were barely visible and

so 1 mL of eDNA was concentrated twice. The initial 100 µL product was resuspended

in dH₂0, and then in 100 µL 10 mM Tris (pH 8.0) after a second EtOH precipitation.

Resulting DNA had a concentration of 25 µg/mL; by using 4-5 µL (100-125 ng), 16S

PCR resulted in much brighter gel bands. Figure 2.2 shows the effects of DNA
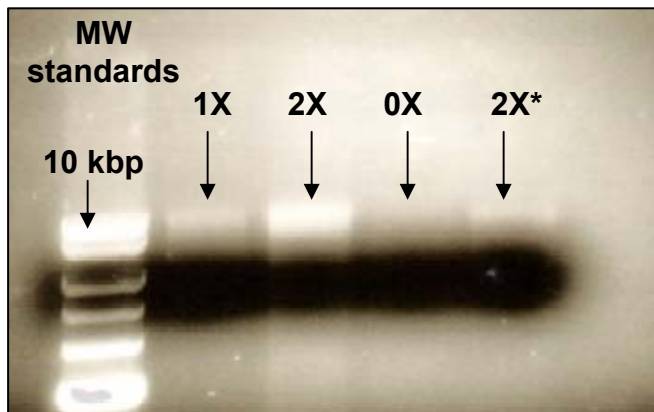
**Figure 2.2: Effects of concentrating e DNA and its stability over time. A single EtOH precipitation
(1X) increases [DNA]; when repeated (2X), doubly so. One month at -20ºC (2X*) led to ~50% DNA
breakdown/loss.**

concentration on extracted nucleic acid samples, as well as the degradation of eDNA

during storage. Although immediately transferred to -20ºC following all experimental

procedures, approximately 50% less DNA was noted in a sample that had been stored just thirty days as compared to more recent isolate.

The 16S positive control reaction was, at this stage, considered a reliable and repeatable indicator of PCR success; using the same conditions, an initial attempt to amplify *pksG* failed. Using a High Fidelity *Taq* polymerase, PCR conditions were optimized for *pksG* amplification. While monitoring the effects of primer and magnesium concentration, PCR products began to show up as blurred gel lanes. Conditions were adjusted to minimize streaking, resulting in a reaction recipe that included: 100-125 µg eDNA, 0.8-1.0 µM each *pksG* primer, 2.0 µM MgSO$_4$, and anannealing temperature between 34.5ºC and 36.0ºC. With these parameters, gel streaks were concentrated around the 550-700 bp region, but were still too delocalized for cloning.

To improve precision, a second set of degenerate primers, *pksG* deg2 UP and *pksG* deg2 DN, were designed. As they target the same 600 bp region, these were meant to be used as a set or in conjunction with the original *pksG* primers. The degenerate primers, still targeting the above conserved amino acid sequences, contained different nucleotides than the first set, in an attempt to maximize the possibility of different bases encoding for the desired residues. As the conserved amino acid region was the same, it was hypothesized that increasing the degeneracy of the primers would improve the chances of PCR success. No amplification was achieved with any primer combination, despite optimization attempts.

II.C.4. eDNA Purification

MO BIO PowerClean^{TM} DNA Clean-Up: One possible explanation for both the

non-specific *pksG* bands and low concentration of 16S products, was the presence of

coextracted inhibitors in our eDNA samples. The PowerClean^{TM} DNA Clean-Up Kit

from MO BIO was used to purify isolates; via spin column filtration with three wash

buffers, organic and non-organic matter was precipitated out of 150 µL extract to a final

50 µL yield. It was noted that, if used to purify the isolate immediately after extraction,

no DNA band was detectable on an agarose gel. The kit was found to be effective only

after a 2x EtOH precipitation of the crude extract; resulting eDNA had a concentration of

20 µg/mL and PCR amplification was markedly improved. *Following PowerMax^{TM}*

extraction, 2 x EtOH precipitations, and PowerClean^{TM} purification, PCRs resulted in

a bright band at ~600bp for 16S, and a faint but distinct band at ~600 bp for **pksG.**

Each purified PCR product was successfully subcloned into pBAD with

subsequent transformation; 16S transformants were screened, and insert presence verified

by DNA sequencing. Transformation of *pksG* resulted in 34 colonies. All were

screened, and PCRs showed three different products – one was potentially accurate for

pksG at ~600 bp, thirteen were between 200-300bp, and the remaining 20 were ~1300bp.

Plasmid concentration varied from 15-90 µg/mL, which was insufficient for DNA

sequencing; using 10 mL inoculated cultures increased the range to 75-315 µg/mL.

Sequencing attempts were unsuccessful for all three products.

Gel Purification: Around this time an article by Professor Sean Brady at

Rockefeller University appeared in *Nature Protocols,* and described an electroelution

method for purifying crude eDNA extracts.[41] Based on this, 170 µL eDNA isolate, from

soil obtained at UNC Greensboro, was loaded into a 1% agarose gel that contained no

ethidium bromide. To prevent the oily isolate from floating into solution, the gel was run

at 100V for 1 h until samples were absorbed into the agarose; the gel was then submerged

in buffer and run for an additional 45 minutes. One lane was excised and stained with

ethidium bromide until a band was visible at ~1-2 cm from the top. The portion of

unstained gel containing DNA was excised and put in dialysis tubing with 10 mL 0.5x

Tris/boric acid/EDTA (TBE) buffer. Submerged in buffer, a 100V current was run

perpendicularly through the tubing for 2.5 hours. Following centrifugation to remove gel

residue, 10 mL of solution remained. Two EtOH precipitations of a 1 mL aliquot

resulted in DNA with a 15 µg/mL concentration.

The eDNA obtained from each purification technique was similar in

concentration, yet electroelution provided 100 µL eluent as opposed to 50 µL from

PowerClean[TM]. To compare inhibitor removal abilities of the two methods, PCRs were

performed using reaction conditions as described by J.R. Spear et al.[42] The 16S rRNA

products seen in Figure 2.3 verified that both methods were successful at purifying
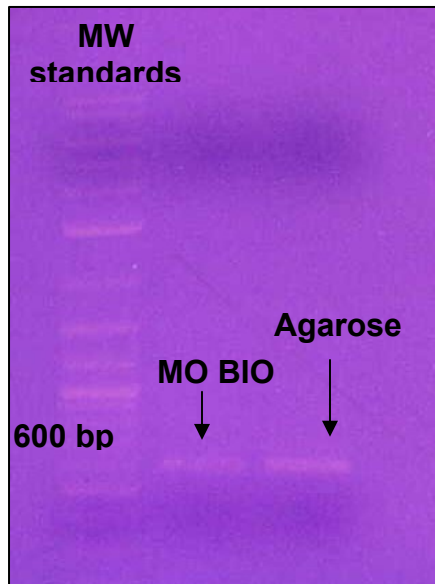
Figure 2.3: eDNA sample purity and concentration following gel purification. Both were similar following either method, and bands at ~600 bp verify 16S rRNA amplification.

extracted DNA, and with comparable results. It should be noted that 10 g of the sand/soil collected from Falls Park in Greenville, S.C. was prepared by PowerMax[TM] extraction, gel purification, and 2 x EtOH precipitations. Resulting eDNA had a 40 µg/mL concentration, twice that of any soils collected locally; however, no *pksG* amplification was observed using that sample.

Using gel-purified eDNA, PCR optimization of *pksG* was repeated. A reaction mixture of 75 µg DNA, 5µg BSA, 0.8-1.0 µM each primer, and 2.0 µM Mg^{2+}, resulted in 250 bp and 400 bp products. Sequencing showed that *pksG* degen UP served as both the forward and reverse primer in these PCR reactions. Homology screening using BLAST (Figure 2.4) revealed that the translated 400+ bp region showed 31% identity, 51% positive match to an *E. coli* putative carbohydrate kinase; a translated protein sequence

from the 250 bp product was had 21% identity, 42% positive match to a bacterial α-rhamnosidase.

```
eDNA/pksG 400 bp sequence   1    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
carbohydrate kinase         1    M F L N V R R C A M C D K E Y I I T I G S A N I D I A G Y S Y A S L N Y A D S N P G K I K F T P G G

eDNA/pksG 400 bp sequence   1    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
carbohydrate kinase         51   V G R N I A H N L A L L G K N S W L M S A V G D D F Y G Q S L L A Q T T Q S G V N V D K C L I V N G

eDNA/pksG 400 bp sequence   1    - - - - - - - - - - - - - - - - - - - - - - L P Y D D P V P D E L L R R K S - - - - - - - - - - - - -
carbohydrate kinase         101  E N T S S Y L S L L D K T G E M L V A I N D M T I S E H I S A E F L S Q H I D F I Q G A K V I V A D

eDNA/pksG 400 bp sequence   16   - - - - - - - - - I R L L L R I C S L P L L R G - - - - - R D I R L R R H L P P I Q Q H Y P V A V D Q
carbohydrate kinase         151  C N I S E S T L V W L L N N A G K V P V F V D P V S A W K C V K I R E H L A Q I H T L K P N R L E A

eDNA/pksG 400 bp sequence   54   D S G A G E A L A D R H N R R - - - T A L H R R A T V S P V L P V R G D R L H T R D H H R L R R Y L
carbohydrate kinase         201  E T L S G I S L S D R N D A E K V A V W F H T H G L N R L V L S M G C D G V Y Y S E L Y G A S G W S

eDNA/pksG 400 bp sequence   101  P P V Q R C D E G P D Y Q D R S V G D C L V V G L P A S W A L L R - - - - - - - C A A P R V L L L H
carbohydrate kinase         251  P P V K T Q V V N V T G A G D A M M A G L A S C W V D N M P F I D S I R F A Q G C S S M A L A S E Y

eDNA/pksG 400 bp sequence   144  R N C L Q P R D H H K X - - - - - - - -
carbohydrate kinase         301  T N N P E L S V A N V K S L V E K Q N V


eDNA/pksG 100 bp sequence   1    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
alpha rhamnosidase          1    X S L I L L G A L S L A S S T F A Q T W I W Y P G D Y E I W L G N Q X N N R R T E R G A F F P P F W

eDNA/pksG 100 bp sequence   1    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
alpha rhamnosidase          51   K T D S H Y V V V E F S K V L N L S E P E E V F I A A E G T Y N V K L D G K L Q F G X P E T L L L P

eDNA/pksG 100 bp sequence   1    - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - S G - - - -
alpha rhamnosidase          101  A G K H S L N I K V W N Q A T P P T I Y V K G K T V N S D S S W R V T Y E D K E W I D E S C K A S D

eDNA/pksG 100 bp sequence   2    - - - - - - - - S C F R S R S A P R R P T A A R L X X R P R P G P P G T A S Q Q V Q G G - - - - -
alpha rhamnosidase          151  T S A T I Y X D A C C W N F D G A T Q R P S Q F S L X - - R E P Q Q P V A K T E Q P E C C I L Y D F

eDNA/pksG 100 bp sequence   39   - - - - - - - - - - - - - T C S C S R S X F R T R - - - - - - - - - - C S - - - - - - - R C - - - -
alpha rhamnosidase          199  G K E T F G F I T L K N L S G K C K I D L Y Y G E S P E E A K D K A Y C E T L D K L L L E P C Q I T

eDNA/pksG 100 bp sequence   55   - - - - - - - - - - H H T E - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
alpha rhamnosidase          249  D L A I R S T S P L H H S D N E Y T L E N S K A F R Y V Y I T H E P E V Q I G E V S X Q Y E Y L P E

eDNA/pksG 100 bp sequence   59   - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
alpha rhamnosidase          299  E Y R G N F R C N D E E L N C I W E V G A Y T X H L T T R E F F I D G I K R D R W V W S G D A I Q S
```

Figure 2.4. Homology results for the two eDNA-pksG sequences.

II.D. Conclusions

Access to nucleic acids from bacteria living in soil is hindered by the complex and diverse components of all soil types. Simple screening techniques such as PCR are possible, but only after careful preparation of collected samples. Preparing eDNA must

33

be done in a way that benefits research goals, specifically by adapting the procedure to access target molecules. Based on the success of 16S rRNA control reactions, we have concluded that SDS-based and PowerMax^{TM} extraction of crude eDNA are effective methods for soils collected in our region of North Carolina. The SDS yield is twice that of the kit, and so more appropriate for larger-scale trials. To access Gram-positive bacteria, crude extract purification by agarose gel electrophoresis and a 100-fold eDNA concentration increase are sufficient preparation for a PCR survey.

In comparison to the control reaction, all attempts to amplify *pksG* HMG-CoA homologs were unsuccessful. We concluded that the relatively small β-branching polyketide population was not well-represented in our purified eDNA samples for one of two reasons; these metabolites are either absent from the local soil population or exist in such small amounts that detection is not possible with the described procedure.

II.E. Future Metagenomic Work

In order to pursue the β-branch polyketides further, a cosmid library of the environmental nucleic acids present in local soils could be constructed. By obtaining a large amount of DNA and cloning it into an expression host, it would be feasible to identify clone-specific molecules with antibacterial activity. An alternative host system with different expression capabilities may also benefit, to ensure that there are no heterologous expression difficulties.

The above techniques may also be used to screen for other classes of molecules. Future targets may include β-lactams, lantibiotics, and aminoglycosides. Additionally, other types of environmental samples are available for β-branch polyketide screening.

There are many other types of soil in North Carolina alone, varying by region and bacterial content. Also, microbial communities are to be found in other sediments and water; all of these are possible areas for future work.

II.F. Experimental

II.F.1. eDNA Extraction

EPICENTRE® Soilmaster™ DNA Extraction: All soils were collected by the author using sterile technique, and stored at 4ºC for up to one month. Approximately 100 mg of four soils from Greensboro, NC were processed as follows: the EPICENTRE® Soilmaster™ DNA Extraction Kit used a hot, SDS-based detergent and proteinase K at 65ºC for chemical and enzymatic cell lysis, respectively.   Proteins were then precipitated out of the solution, followed by spin column chromatography to remove inhibitors. The extracted DNA was resuspended in 300 µL of 10 mM Tris-HCl (pH 7.5), 1 mM EDTA.

Using this as template DNA, initial PCRs were unsuccessful. Alterations to the procedure included: omission of vortexing steps during extraction to prevent DNA shearing, incubating on ice instead of at room temperature during DNA precipitation, and extended centrifugation. Despite these adjustments, no PCR reaction resulted in a product.

SDS-based extraction: For SDS-extraction, 13.5 mL extraction buffer (100 mM Tris-HCl, pH 8.0, 100 mM Na-EDTA, pH 8.0, 100 mM Na·phosphate, pH 8.0, 1.5 M NaCl, 1% hexadecylmethylammonium bromide) and 20 µL proteinase K (50 mg/mL) were mixed with 5.0 g soil by horizontal shaking for 30 minutes (37ºC, 225 rpm).   DNA was extracted by adding 1.5 mL 20% SDS and incubated for 2 h (65ºC), during which the

tube was inverted every 20 min to mix. Supernatant was removed to a sterile tube following centrifugation (10 min, 6,000 x g). The remaining soil pellet was extracted twice more by adding 4.5 mL buffer and 0.5 mL 20% SDS, vortexing for 10 s, incubating at 65ºC for 10 min, and centrifuging as before. All three supernatants were combined to 12.5 mL in a clean tube; 12.5 mL of 24:1 vol/vol chloroform-isoamyl alcohol was added and the aqueous phase removed after a 10 min centrifugation. DNA was precipitated by adding 0.6 vol isopropanol and incubating 1 h at room temperature. After centrifuging (20 min, 16,000 x g), supernatant was removed, the brown pellet washed once with cold, 70% ethanol, and resuspended in 500 µL dH$_2$O.

As the crude extract was amber in color, 100 µL was purified using a vacuum manifold and the Wizard® PCR Preps System. The 50 µL eluted with sterile H$_2$O had a very faint tan-colored appearance, and the procedure was repeated. Another 100 µL of extract was purified with the QIAquick® PCR Purification Kit; although clear and colorless, the 50 µL eluent was processed again. Loaded directly into a 1% agarose gel, each sample showed DNA presence by a bright band located above 10 kbp. No PCR amplification was achieved with 16S rRNA or *pksG* primers.

MO BIO PowerMax^{TM} Soil DNA Isolation: Next we obtained the PowerMax^{TM} Soil DNA Isolation Kit. Soil is first added to a "bead-beating" tube to disperse particles and protect nucleic acids. After SDS-based cell lysis and 2 x EtOH precipitations, eDNA is bound to a salt solution for spin column chromatography. With the 100 µL isolate, a very faint band was seen above 10 kbp. To concentrate this, 1 mL crude extract was mixed with 0.04 mL NaCl (5 M), 2.1 mL cold 100% ethanol, and centrifuged (30 min,

13.2 rpm).   Supernatant was decanted, the pellet washed twice with cold, 70% ethanol, and resuspended in 100 µL sterile, 10 mM Tris.

II.F.2. PCR Optimization

All polymerase chain reactions were performed in a thermocycler, and cooled to 4.0 ºC after completion. For agarose gel electrophoresis of each PCR product, 16 µL was mixed with 4 µL loading dye, and the 20 µL solutions loaded into a 1% agarose gel containing 9 µL ethidium bromide. Gels were run at 120V to completion, bands viewed by UV light and their size determined by standard molecular weight markers.

Taq **Master Mix:** With the concentrated extract as template DNA, 16S rRNA and *pksG* were amplified using *Taq*® Master Mix from Invitrogen™. The reaction conditions were: a 2 minute "hot start" incubation at 94.0ºC, then 30 cycles of 30 s at 94.0ºC, 30 s at 55ºC for 16S and 45.0ºC for *pksG*, and 1.5 min at 65.0ºC. A faint band was noted around 600 bp for each PCR product; however, both PCRs were repeated twice with no observed product.

Phusion Polymerase: The same eDNA template was used for PCR with a high fidelity Phusion polymerase from New England BioLabs®, Inc. Reaction conditions were adjusted to: one 30 s hold at 98.0ºC, then 30 cycles of 98.0ºC for 10 s, 60.0ºC for 30 s, and 30 s at 72.0ºC, followed by a 5 min elongation hold at 72.0ºC. The 16S product showed indistinct streaking, and there was no band for *pksG*.

To improve these results, 1 mL crude eDNA extract was concentrated as above and resuspended in 100 µL dH2O; to this, 2 µL 5M NaCl and 104 µL cold, 100% ethanol were added before centrifugation (30 min, 13.2 rpm). The pellet was washed as before,

and resuspended in 100 μL sterile, 10 mM Tris. A repeat PCR of both target regions resulted in a band at ~600 bp for 16S, but none for *pksG*.

Taq High Fidelity Polyemerase: A High Fidelity *Taq*[®] polymerase was obtained from Invitrogen[TM] to optimize *pksG* PCR conditions; reaction conditions were identical to those using *Taq* Master Mix. The following parameters were varied: Mg^{2+}, primer, DNA, dNTP, bovine serum albumin (BSA), and DMSO concentrations, as well as annealing temperatures. The 16S primers were used for each trial, and conditions resulting in the brightest band (i.e. highest amount of amplification) were considered optimal.

During trials, PCR products started to show extensive streaking in gel lanes; once this was minimized, samples exhibited a broad "band" concentrated at the 600-700 bp region and were processed further. The PCRs contained: 100-125 ng eDNA, 3 μL of each 10 μM primer, 40 μL of a 50 mM Mg^{2+} solution, and 200 μM dNTP's. Addition of BSA or DMSO did not enhance amplification. For reactions that resulted in these broad bands, the *pksG* PCR product was purified with the QIAquick[®] PCR Purification Kit.

II.F.3. Initial TOPO Cloning Attempt

The pBAD TOPO[®] TA Expression Kit from Invitrogen[TM] was used in an attempt to clone and overexpress the purified *pksG* PCR product. As the vector confers antibiotic resistance to the bacterium, all agar plates and starter cultures were prepared with Luria-Bertani media (0.5% yeast, 1.0% Bacto tryptone, 1.0% NaCl) and contained 100 μg/mL ampicillin.

4 µL of the *pksG* insert was mixed with 1 µL TOPO® cloning vector, a standard salt solution, and incubated at room temperature for 15 min. Of this mixture, 2.0 µL was added to competent "TOP10®" cells, incubated on ice for 30 min, at 42.0ºC for 2 min, and 250 µL LB broth was added. Cells were shaken horizontally for 1 h (37.0ºC, 225 rpm), then 50 µL was spread over an agar plate. Following 24 h incubation at 37.0ºC, the plate had three colonies (G$_1$, G$_2$, and G$_3$) and was stored at 4ºC.

All three colonies were grown on fresh plates, and used to inoculate 5 mL starter cultures that were shaken overnight (225 rpm, 37.0ºC). Following purification the next day, 1 µL of each was used as DNA template for PCR. Only G$_3$ showed a product, and at ~1400 bp was a possible dimer of *pksG* that occurred during amplification. New G$_3$ cultures were purified and combined to a concentration of 260 µg/mL; DNA sequencing revealed no match to the *pksG* target insert.

II.F.4. eDNA Purification

To remove any coextracted inhibitors, 150 µL of PowerMax™ eDNA isolate was purified with the PowerClean™ DNA Clean-Up Kit from MO BIO Laboratories. With 4–5 µL of the 50 µL eluent as template DNA, PCRs for 16S and *pksG* resulted in bands around 600 bp. Following PCR clean-up and TOPO® cloning, 34 transformants were screened for the *pksG* insert. Of these, one was ~600 bp, 10 were between 200-300 bp, and the remaining around 1300 bp. Eight of the 1300 bp samples were digested using *Msp*I and *Hind*III restriction endonucleases. After 3.5 h incubation at 37ºC, samples were heat-deactivated for 20 min at 65.0ºC, and loaded into a 2% agarose gel. No bands

were seen for any reaction. DNA sequencing on each of the three TOPO® products was unsuccessful.

An eDNA purification procedure described by Sean F. Brady was performed in an attempt to eliminate coextracted inhibitors.[41] A 1% agarose gel was prepared with 0.5x TBE buffer (45 mM Tris, 45 mM boric acid, 1 mM EDTA, pH 8.0) and no ethidium bromide. The initial attempt to load crude extract failed, as the oily solution floated up to the buffer surface. Per a Brady recommendation, a gel was placed in the apparatus with TBE buffer up to the gel surface but not covering the loading wells.[41] 170 µL crude extract was heated for 35 min at 50ºC, loaded, and run in the gel for 30 min at 100V. Once eDNA was absorbed, the gel was submersed in TBE buffer and run for an additional 45 min.

A lengthwise piece of the gel containing one lane of eDNA, was excised and placed in 100 mL TBE buffer with 10 µL ethidium bromide. After mixing for 2 h, a bright band was noted by UV light at ~ 1-2 cm from the well; per this location, the corresponding portion of unstained gel was excised and placed in dialysis tubing with 10 mL TBE. (After staining the remaining gel pieces, it was noted that a significant amount of DNA remained in the agarose.) With the dialysis tubing placed perpendicularly to a 100 V current for 2.5 h, eDNA was electroeluted out of the agarose; resulting buffer was centrifuged for 10 min to remove gel residue, leaving 10 mL DNA solution. Using 1 mL of this, a 2x concentration was performed and the 100 µL product contained 15 µg/mL eDNA.

II.F.5. Final *pksG* Amplification Attempt

A literature search revealed PCR parameters to amplify 16S rRNA samples similar to those described above.[42] A control PCR using BAC 16S primers and *B. subtilis* 39374 led to a bright, distinct band near 600 bp. Using eDNA from both PowerClean™ and gel purification, two PCRs were mixed with *Taq* Master Mix to contain: 200 ng eDNA, 0.5 μL each BAC 16S primer, 0.25 μM Mg^{2+}, and 5 μg BSA. Reaction conditions were: a 2 min, 94.0ºC hot start, followed by 30 cycles of 30 s at 94.0ºC, 1 min at 55.0ºC, 1.5 min at 72.0ºC, and a final 12 min incubation at 72.0ºC. Both amplified successfully, with bands at ~600 bp of similar intensity.

The concentration of *pksG* primers and Mg^{2+} was varied in multiple PCRs to complement adjusted reaction conditions, and the annealing temperature decreased to 45.0ºC. With 3.0–4.0 mM $MgCl_2$ and 0.8–1.0 μM of each primer, two bands were seen for each reaction; a bright, distinct band at ~400 bp, with another near 250 bp that was noticeably fainter. TOPO® cloning products of each were sent for DNA sequencing, and the results are shown in Figure 2.4.

II.G. Cloning, Overexpression and Purification of the PksJ Protein

II.G.1. Background

The product of *pksX* and the *bae* ortholog is dihydrobacillaene, rather than its dehydrogenated counterpart bacillaene.[37] Per modular enzyme synthesis (Table 1), an enoyl reductase should be included in the 3rd module of *pksJ* for this structural moiety. Of particular interest is the unusual method by which *pksJ* processes a β-keto moiety to an α,β-enoyl. Rather than subsequent reduction and dehydration to produce the *cis*-

alkene structure, the growing chain is transferred between two tandem thiolation domains prior to DH activity.

It was hypothesized that a trans-ER of the PksE flavoprotein reduces the C22-C23 double bond.[7] At the request of Dr. Jörn Piel, University of Bonn, Germany, our goal was to isolate, overexpress, and purify the PksJ protein. Once a protein scaffold of the tandem thiolation domains and substrate analog were available, his research group hoped to demonstrate the bond reduction and perform activity characterization.

II.G.2. Results

A 535 bp region of the *pksJ* gene was amplified from *B. subtilis* 168 by PCR and successfully cloned into the pET-28a plasmid via the *Nco*I and *Xho*I restriction sites. Ligation products were transformed into competent *E. coli* DH5α cells, incubated overnight, and individual colonies used to inoculate starter cultures. Once the pET-28a/*pksJ* plasmid was purified from DH5α, it was retransformed into *E. coli* BL21(DE3) cells for the purpose of overexpression.

Positive transformants were used to inoculate starter cultures and added to 1L Luria-Bertani media. The OD_{595nm} of this large culture was monitored until it was between 0.500 and 0.600, and *pksJ* overexpression was induced with isopropyl-β-D-thiogalactopyranoside (IPTG). Following an overnight incubation, cells were recovered by centrifugation.

During the cloning process, a 6x Histidine tag was added to the protein via the pET-28a C-terminus. This allowed for *pksJ* purification by a nickel-nitrilotriacetic acid (Ni-NTA) affinity column. Once the overexpression cells were applied to the column, it

was washed with 60 mM imidazole and eluted in 1 mL fractions. Using a Bradford reagent assay at 595 nm, fractions 2 and 3 had an absorbance of 0.195; a 10% SDS-PAGE gel showed no band for either fraction.

Using PksJ protein from a new overexpression culture, purification by Ni-NTA column chromatography was repeated. Fractions 2 and 3 showed A_{595nm} of 0.518 and 0.423, respectively. A 15% SDS gel showed thick bands around the 20 kDa marker; the PksJ protein was present but samples were saturated and assumed to contain impurities. The affinity column purification procedure was repeated twice, using differing imidazole concentrations in the wash buffer solution. At 140 mM imidazole, the protein product was not present in eluted fractions and was believed to have been washed out of the column. With a 90 mM imidazole wash buffer, the eluted protein band was more concise, yet still appeared to contain impurities.

Before Dr. Piel was able to complete the scaffold synthesis and utilize the purified PksJ protein, the Christopher Walsh group demonstrated that the *trans*-ER of PksE reduces the double bond of dihydrobacillaene to produce the bacillaene antibiotic.[43] This was possibly due to the tandem thiolation domains; it was reported that kinetic analysis of this reaction will be forthcoming.

II.G.3. Experimental

Cloning*: Bacillus subtilis* 168 cells were streaked on an LB-agar plate and incubated overnight at 37ºC. One colony was used to inoculate a 5 mL LB culture that was shaken at 37ºC and 225 rpm. The following day, *B. subtilis* 168 genomic DNA was isolated with the Wizard® Genomic DNA Purification Kit from Promega.

Primers were designed to augment a 535 bp region of *pksJ*; these were a forward

primer, 5' – CAAACTAGAA<u>CCATGG</u>AAGCAGCCTTAATC – 3' with *Nco*I sequence,

and the reverse 5' – AGGGTGTCCTCGAGTTCAACGGTC – 3' with *Xho*I sequence.

Each primer was mixed with *B. subtilis* 168 genomic DNA, dNTP's, 50 mM $MgSO_4$, and

a high fidelity *Taq* polymerase. The PCR conditions were: 94.0 °C hot start for 2 min,

then 30 cycles of 30 s at 94.0°C, 30 s at 54.0°C, and 1 min at 68.0°C, then cooled to

4.0°C. A 1% agarose gel confirmed *pksJ* with a band slightly above 500 bp, and the PCR

product was purified with the QIAquick® PCR Purification Kit.

On an LB-agar plate containing 30 µg/mL kanamycin, DH5α/pET-28a cells were

grown overnight at 37°C. One colony was used to inoculate a starter culture of 5 mL LB

containing 30 µg/mL kanamycin, and shaken overnight (37°C, 225 rpm). The culture

was purified the next day with the QIAprep Spin Miniprep Kit.

The purified pET-28a plasmid and *pksJ* PCR product were both digested at 37°C

for 3.5 h, using the *Nco*I and *Xho*I restriction enzymes. Following a 65.0°C heat

deactivation for 20 min, both were purified by a 0.7% low-melting agarose gel run at

4.0°C. Each band was excised and placed in a 75.0°C water bath until the agarose had

melted. With 1µL T4 DNA ligase, 2µL buffer, and H_2O, five reactions were mixed with

different insert:plasmid ratios (in µL) of 1:1, 1:3, 3:1, 2:8, and 8:2. These were ligated in

a thermocycler by three incubation periods: 30 min at 37.0°C, 2 h at 23.0°C, then

overnight at 16.0°C.

Following heat-deactivation (20 min, 65°C) the next day, each 20µL ligation

product was mixed with 100µL competent DH5α cells. After 30 min on ice, cells were

heat shocked, 1 mL LB added, and incubated for 1 h at 37ºC. Following centrifugation (5 min, 8,000 rpm), 1 mL supernatant was removed and the remainder plated on kanamycin-agar plates. All five plates were incubated overnight at 37.0ºC; except for 1μL:1μL, all insert:plasmid ratios resulted in multiple colony growth. Four were used to inoculate cultures containing 5 mL LB with 30 μg/mL kanamycin and grown overnight. Once purified, PCR screening of all four samples confirmed the *pksJ* insert.

Overexpression: *E. coli* BL21(DE3) cells were grown overnight at 37.0ºC on an LB-agar plate. A starter culture was shaken at 225 rpm, 37.0ºC overnight; 50 μL of this was added to 5 mL LB the following morning, grown until cloudy, and made chemically competent. To 100μL competent BL21(DE3) cells, 1μL of a purified *pksJ*/pET-28a culture was added and incubated on ice for 30 min. Following a 2 min heat shock at 37.0ºC and 10 min at room temperature, 1 mL LB was added and the mixture incubated for 1 h at 37.0ºC. The sample was centrifuged at 8,000 rpm for 5 min, 1 mL supernatant was removed, and 20μL plated on agar plates containing 30 μg/mL kanamycin.

Following an overnight incubation at 37.0ºC, one colony was used to inoculate a starter culture of 5 mL LB and 30 μg/mL kanamycin, which was shaken at 220 rpm and 37 ºC. The next morning, 2mL of the starter culture was added to 1L sterile LB media, along with kanamycin to a final concentration of 30 μg/mL. The overexpression culture was shaken at 225 rpm, 37.0ºC until the OD_{595nm} reached 0.530, and then induced with 0.63 mM isopropyl-β-D-thiogalactopyranoside (IPTG). After induction, the culture was grown overnight (37.0ºC, 225 rpm). The sample was centrifuged the next day for 30 min at 6,500 rpm; supernatant was discarded and the pellet stored at -80ºC.

Purification: The pellet was resuspended in 20 mL of 1x binding buffer (5 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl, pH 7.9). This solution was sonicated for 3 min to lyse the protein, and centrifuged for 30 min (15,000 x g, 4ºC). Approximately 17 mL supernatant was syringe filtered, and loaded onto a Ni-NTA column. Once adsorbed, 20 mL of 1x binding buffer was used to wash the column; a second wash was performed with 12 mL of 1x wash buffer (60 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl, pH 7.9). The protein was eluted in 1 mL fractions with 12 mL of 1x elute buffer (1 M imidazole, 500 mM NaCl, 20 mM Tris-HCl, pH 7.9).

From each 1 mL fraction, 33 µL was added to 1 mL Bradford reagent, and absorbance recorded at 595 nm to assess protein concentration. Fractions 2 and 3 had an absorbance of 0.518 and 0.423, respectively; 20 µL was combined with 80 µL SDS buffer and run in a 15% SDS-PAGE gel. The two bands at ~20,000 Da were indicative of saturation. Column chromatography was repeated using a wash buffer with 140 mM imidazole, yielding no protein. The next step would have been to optimize the imidazole concentration of the wash step to reduce saturation and yield a pure PksJ protein.

CHAPTER III

BACILLUS TAXONOMY

III.A. Background

III.A.1. Classification of *Bacilli*

There are more than 200 known species of the *Bacillus* genus, and microbiologists have worked for decades on their taxonomic classification.[44] Historically, biochemical and morphological properties formed the basis for this before the advent of DNA sequencing. The current trend is a focus on variability in regions of the non-coding 16S rDNA gene, which is important for DNA transcription and is highly conserved among similar groups of eukaryotes and prokaryotes alike. The ability of a bacterial strain to produce specific aromatic polyketides has also been suggested as an additional means of classification.[45, 46] Horizontal gene transfer is known to occur between bacteria, and may result in two species sharing the same biosynthetic capabilities (such as homologous PKS gene clusters).

III.A.2. *B. subtilis* and *B. amyloliquefaciens*

Bacillus amyloliquefaciens is a Gram-positive bacterium that was originally mistaken for *B. subtilis*, but is now known to differ in its ability to suppress pathogenic organisms and stimulate plant growth.[7] The closely related *B. amyloliquefaciens* has also been found to produce the antibiotics difficidin and bacillaene. These are coded for by two of three large PKS gene clusters (*dif*N and *baeG*, respectively), whose sequences have been reported by Chen et al.[7] The first of these clusters, *pks1* or *bae*, is a known

ortholog of the *pksX* operon of *B. subtilis*, thought to be present due to either horizontal transfer or a common ancestor.

ATCC strains 39374 and 39320 of *B. subtilis* were shown to produce difficidin in 1987, as described earlier. While working with ATCC strain 39374, a probable *pksG* homolog was successfully amplified and sequenced. This was compared to the published genome for *B. subtilis* strain 168, and to *B. amyloliquefaciens* FZB 42. Based on the results, it was suspected that the ATCC strains had been misclassified and further analysis was conducted.

III.B. Results

III.B.1. *PksX* Presence

The probable homolog from strain 39374 was successfully amplified using *pksG* degenerate primers described in Chapter 2. This was then subcloned into pBAD via topoisomerase-TA cloning methodology and ten transformants were used to inoculate starter cultures. The following day each was purified, and PCR repeated to verify the *pksG* insert. DNA sequencing was obtained for one positive transformant and compared to both *pksG* from *B. subtilis* strain 168 and to the *baeG* gene of *B. amyloliquefaciens* FZB 42. This sequence showed an 89% coded positive identity and 98% similarity with strain 168 *pksG*, and the coded amino acid sequence was a 100% match to FZB 42 *baeG*, with a 98% DNA sequence equivalence.

A PCR was repeated using strain 39320 and the same *pksG* degenerate primers; subcloning and screening techniques were repeated and a DNA sequence was obtained. This had a 97.5% match with *baeG*, with 594 of 609 equivalent nucleotides; the

converted amino acid sequence was a 99% identity and 100% similarity match. However, there were four more mismatched amino acids when compared to *B. subtilis* 168.

A former colleague, Suzette Mills, used primers specific to the *difCNO* genes for PCR amplification with strains 168, 39374, and 39320. Agarose gel showed that all three primer sets successfully amplified a genomic region from both ATCC strains, and no products when strain 168 was the DNA template for the reaction. Since the *dif* PKS cluster is present only in *B. amyloliquefaciens*, this was considered further indication of the misclassification of these two strains as *B. subtilis*.

III.B.2. rRNA and rDNA

Ash et al. identified a conserved region of rRNA that is 98.3% similar between the *B. subtilis* and *B. amyloliquefaciens* species.[47] Within this are two hypervariable regions (nucleotides 59-79 and 625-64); classification of strains within the two species is possible based on single nucleotide polymorphisms within these regions. Primers Bsub5F and Bsub3R were used in PCRs with strains 39374 and 39320 as template DNA.[48] Agarose gel electrophoresis verified a product of ~600 bp; after subcloning into the pCR®4 vector from Invitrogen™, Inc., DNA sequencing was obtained for each sample, and showed a 99% match to *B. amyloliquefaciens* FZB 42, as compared to a 98% match to *B. subtilis* 168 (Table 4).

	NUCLEOTIDE POSITIONS										
Strain	70	126	180	185	198	202	271	285	465	472	483	GenBank
168	A	T	G	T	T	A	T	A	A	G	T	D88802
FZB 42	G	T	G	T	T	G	C	G	G	A	C	AY055221
39320	G	T	G	C	T	G	C	G	G	A	C	EU564529
39372	G	T	G	C	C	G	C	G	G	A	C	EU564530

Table 4: Sequence results for 16S polymorphisms. GenBank Accession numbers are included for each.

An additional hypervariable section of DNA, known as the intergene region (IGR), is found between the 16S and 23S ribosomal regions and varies by bacterial strain. However, the rDNA on each side of the IGR (the 16S 3' end and 23S 5' end) is highly conserved and species-specific. Multiple copies of the same IGR may be found in the same bacterium that, when amplified by PCR and run on a gel, result in a "fingerprint" pattern.[49] Taxonomic classification of individual strains may be rapidly assessed by determining the number and size of IGR copies.

An IGR fingerprint of *B. subtilis* is known to vary from related species by slight differences in the bands between 250 bp and 500 bp, and the band at ~640 bp. Whereas *B. subtilis* typically shows a single band at 430 bp, *B. amyloliquefaciens* also has bands at 320 bp, 360 bp, and 650 bp. Primers 16S-23S Up and 16S-23S Dn were designed to amplify an ~400 bp IGR region, and used for PCRs with ATCC strains 39374 and 39320
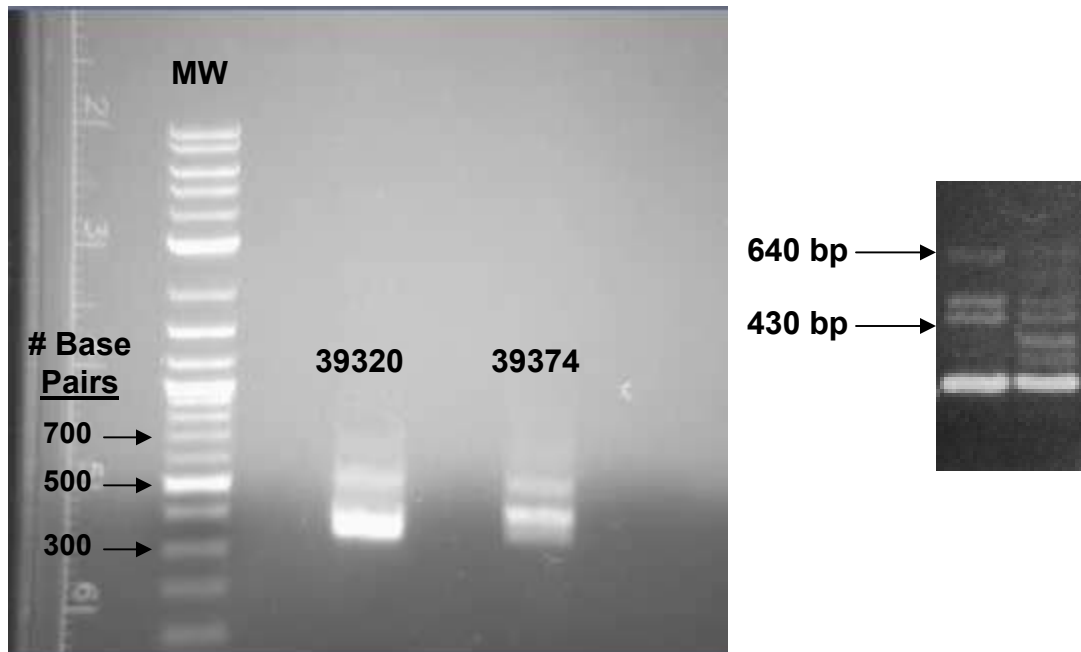
Figure 3.1. Intergene regions of strains 39374 and 39320. Gel on the left shows PCR products for ATCC strains. Gel on the right shows previously resolved fingerprint patterns of B. subtilis (left lane) and B. amyloliquefaciens (right lane).[50]

as the templates.[51, 52] Figure 3.1 shows the resulting agarose gel, with identical bands for both strains at ~350-400 bp, 500 bp, and ~650 bp. Each 350-400 bp band was excised, purified, and subcloned into a sequencing vector. Positive transformants were sent for DNA sequencing, and were a 100% match to each other. The broad "fingerprint" bands between 300-400 bp along with the absence of a band at ~640 bp, was a match for *B. amyloliquefaciens* IGRs reported by the Daniele Daffonchio research group in Milan, Italy.[50]

Six additional single nucleotide polymorphic (SNP) positions were identified in the 16S rDNA region that distinguish between the two species.[53] For ATCC strains 39374 and 39320, each nucleotide corresponded with the *B. amyloliquefaciens* FZB42

sequence rather than *B. subtilis* 168. These are shown in Table 4 as positions 202, 271, 285, 465, 372, and 483.

III.B.3. Detection of Specific Sequences

One difference between the two *Bacillus* species is the α-amylase enzyme that each produces; the *B. subtilis* gene is referred to as *amyS*, while *amyA* refers to *B. amyloliquefaciens*. The lack of homology between these two genes makes them a useful tool for distinguishing the taxonomy of individual strains.[53] The forward PCR primer ABa3 and reverse primers ABa4 and Abs3, specific for *B. amyloliquefaciens* and *B. subtilis* respectively, were used in reactions with strains 168, 39374, and 39320.[53]

PCR primers OTA1 and OTA2 were designed to amplify a second gene, *thyA*, that encodes for a thymidylate synthase found in *B. subtilis* only.[53] Agarose gel electrophoresis (Figure 3.2) showed that the *amyS* and *thyA* genes were amplified only from strain 168. No PCR product was detected for the *amyA* reactions; in addition to the SNP difference at position 185 for the ATCC strains, this indicates that they are more closely related to a related *B. amyloliquefaciens* subtype known to lack the *amyA* gene.[54]
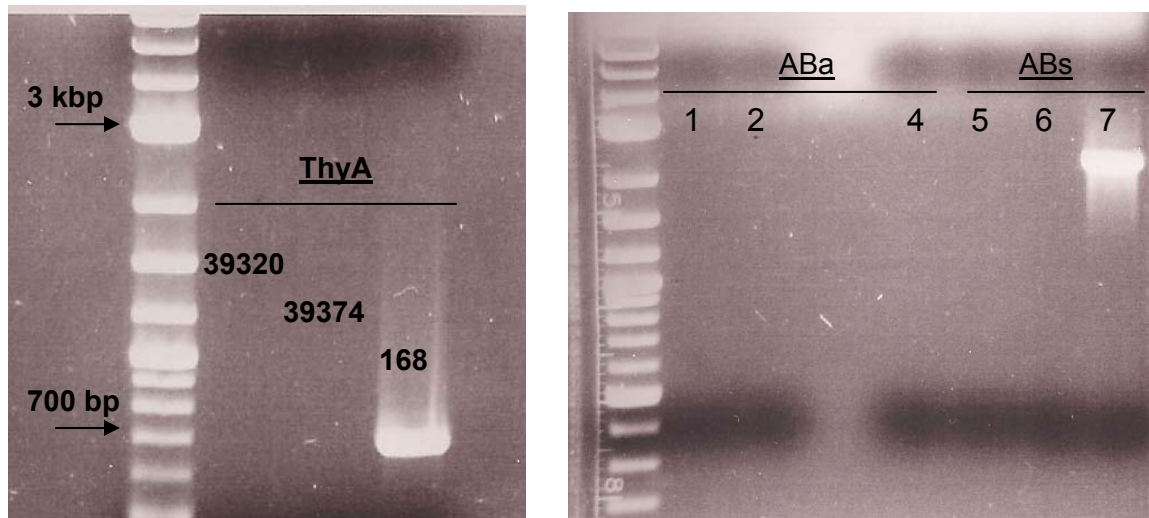
Figure 3.2. ThyA and α-amylase genes. PCR amplification of *thyA* was successful with strain 168 only (left). Neither *amyA* or *amyS* (right) were detected in strains 39320 (lanes 1 + 5) and 39374 (lanes 2 + 6). Strain 168 (lanes 4 + 7) showed presence of *amyS*.

An additional difference between the two species is the arrangement of nucleotides around the *tetB* genes that encode for an anti-porter membrane protein.[54] The reverse TetB_R primer was designed to complement the 5' end of the gene, identical for both species. PCR primers YyaO_F and YyaR_F, specific to the 3' ends of *B. subtilis* and *B. amyloliquefaciens* respectively, were used in conjunction for reactions with all three strains.[54] As expected, YyaO_F/TetB_R yielded a product with strain 168 only, while YyaR_F/TetB_R amplified a product when ATCC strain 39320 was the template.
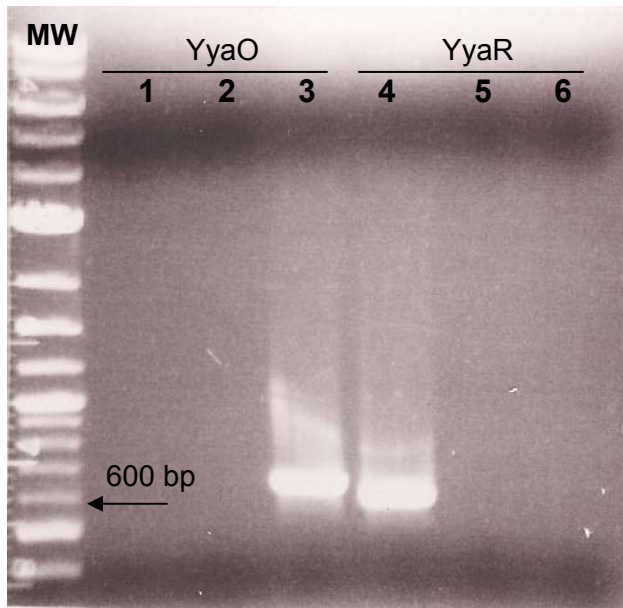
Figure 3.3. Anti-porter *tetB* gene. The *B. amyloliquefaciens*-specific YyaR primer amplified a product with strain 39320 only (lanes 1 + 4). No PCR amplification occurred with strain 39374 (lanes 2 + 5). Strain 168 (lanes 3 + 6) showed a product for *B. subtilis*-specific YyaO product.

III.C. Conclusions

Based on the *pksG* sequences, 16S rRNA and rDNA SNPs, ITG region homology, α-amylase, thymidylate synthase, and *tetB* results, we have concluded that ATCC strains 39320 and 39374 have been wrongly classified as *Bacillus subtilis*. Both of these strains are more accurately classified as *Bacillus amyloliquefaciens*, and should be regarded as such in future research endeavors.

III.D. Experimental

ATCC 39320 and 39374 *B. subtilis* strains were obtained from the American Type Culture Collection (Mannassas, VA) and grown in well-aerated tryptic soy broth. *B. subtilis* 168 genomic DNA was isolated with the Wizard Genomic DNA isolation kit from Promega, per manufacturer's instructions for Gram-positive bacteria. *E. coli* strains

TOP10 and DH5α were obtained in chemically competent form from Invitrogen, Inc. and grown in Luria-Bertani media. Dye-termination sequencing was performed at either SeqWright, Inc. (Houston, TX) or at the Institute for Health, Science, and Society (UNC Greensboro).

III.D.1. *PksX* Presence

Degenerate primers (Table 5) were used to amplify *pksG* orthologs from *B. subtilis* genomic DNA. A PCR was performed with Taq High Fidelity polymerase as follows: a single 94.0 ºC hold for 2 min, followed by 30 cycles of 94.0 ºC for 30 s, 48.0 ºC for 30 s, and 68.0 ºC for 1 min. A 1% agarose gel showed a product for each at ~650 bp, verifying *pksG* amplification. PCR clean-up was performed on both samples, each cloned, transformed into *E. coli* TOP10 cells, and plated on LB/ampicillin plates.

After overnight incubation at 37 ºC, 4-6 transformants were used to inoculate overnight starter cultures. Following purification the next day, PCR was repeated with Taq Master Mix and a 45 ºC annealing temperature. Each PCR product was run in an 1% agarose gel to verify the *pksG* insert. A sample with an agarose gel band at ~650 bp was sent for DNA sequencing.

III.D.2. rRNA and rDNA

Primers Bsub5F, Bsub3R, 16S rDNA Up/Dn, and 16S-23S Up/Dn were used to amplify regions of 600 bp, 650 bp, and 400 bp respectively. PCR was performed with ATCC strains 39374 and 39320 as template DNA, and *Taq* Master Mix polymerase. Reaction conditions were: an initial "hot-start" hold at 94.0 ºC for 2 min, followed by 30

Primer	Sequence
pksG degen Up	5' – CTN CCN TAY GAR GAY CCC GT – 3'
pksG degen DN	5' – NCK RTG NGC NCC YTT NAC CAT – 3'
Bsub5F	5' – AAG TCG AGC GGA CAG ATG G – 3'
Bsub3R	5' – CCA GTT TCC AAT GAC CCT CCC C – 3'
16S rDNA Up	5' – AAG TCG AGC GGA CAG ATG – 3'
16s rDNA Dn	5' – CCA GTT TCC AAT GAC CCT CCC G – 3'
YyaO_F	5' – GGA ACC AGT CCA CAG GGT TGT GG – 3'
YyaR_F	5' – CGA TTG AGT GGG CRA AGG AGA ATC ATT TWT GYG GT – 3'
TetB_R	5' – CCA TAT AGA GCT GTT CCA ATG AGA GG – 3'
ABa3	5' – GCA CGC TGA TGC AGT ATT TTG AAT GG – 3'
ABa4	5' – GGA GCT GTC ACC TTC CCT CGT CC – 3'
Abs3	5' – ACA GGA TCC TCC AGT CTT CAC ATC – 3'
Abs4	5' – GTT CAT CAT TAA GGA TCC TTC AGG – 3'
OTA1	5' – ATG AGA TTC GAC AAC TCA GAG G – 3'
OTA2	5' – AAA GGC AGC ATT AAA CCG C – 3'
16S – 23S Up	5' – GAC TGG GGT GAA GTC GTA AC – 3'
16S – 23S Dn	5' – TGG CTG GGT TGC CCC ATT CGG – 3'

Table 5: Primers used for taxonomy PCRs.[19, 48, 53, 55]

cycles of 94.0 ºC for 30 s, 55.0 ºC for 30 s (50.0 ºC for 16S-23S primers only), and 68.0

ºC for 1.5 min, with a final 5 min hold at 68.0 ºC. Amplification was verified by a 1%

agarose gel and PCR products were purified with the QIAquick® PCR Purification Kit

from QIAGEN Sciences. For all purified reactions, 1 μL was subcloned into the pCR®4-

TOPO vector, spread on LB/ampicillin plates, and incubated overnight at 37 ºC. The

following day, individual colonies were used to inoculate starter cultures that were

shaken overnight at 37 ºC and 225 rpm. Once these were purified the next day, they were

screened for the appropriate insert by a repeat PCR. Positive transformants were sent for

DNA sequencing.

III.D.3. Detection of Specific Sequences

Primers were obtained as designed by Idriss et al. for the α-amylase and *thy*A

genes. Three PCR reactions were run with for each primer set with Taq Master Mix

polymerase, and all three *Bacillus* strains as genomic DNA. A 1% agarose gel was run

for each PCR product to determine gene presence in the various strains.

Using the *B. amyloliquefaciens*-specific ABa3/ABa4 primers, PCR conditions

were: a 3 min hot-start at 94.0 ºC, then 30 cycles of 30 s at 94.0 ºC, 1 min at 55.0 ºC, and

1 min at 72.0 ºC, followed by a 10 min hold at 72.0ºC. For primers ABs3/ABs4 and

OTA1/OTA2, specific to *B. subtilis*, PCR was run as follows: a 3 min hot-start at 94.0 ºC,

30 cycles of 94.0 ºC for 30 s, 50.0 ºC for 1 min, and 72.0 ºC for 1 min, followed by a

single, 10 min hold at 72.0 ºC. A 1% agarose gel showed a band at ~700 bp for the *thyA*

gene with *B. subtilis* 168, with no band for either ATCC strain. Likewise, the ABs

primers resulted in a band ~2 kbp for strain 168 only.

Forward primers YyaO_F and YyaR_F were used in conjunction with the reverse primer TetB_R in the following reactions. Thermocycler settings for these PCRs were: a single 94.0 ℃ hold for 3 min, followed by 30 cycles of 94.0 ℃ for 30 s, 52.0 ℃ for 30 s, and 72.0 ℃ for 1 min, then a single 10 min hold at 72.0 ℃. With strain *B. subtilis* 168 as template DNA, a band was noted at ~650 bp for the YyaO_F/TetB_R reaction, and not for any other primer set. With the *B. amyloliquefaciens*-specific YyaR_F primer, a bright band was noted at ~600 bp for strain 39320, but not for 39374.

CHAPTER IV

INTRODUCTION TO FATTY ACID CATABOLISM

IV.A. Sporulation

The decision to sporulate is based on a number of external and internal cellular signals. Put simply, when the cell has exhausted all possibilities to acquire nutrients and its only option is starvation, the sporulation cycle begins. Bacterial endospore formation begins with chromosomal migration to the cell poles, followed by an asymmetric cell division resulting in two components: the small forespore and a larger mother cell.[56] These two cells have very different roles and fates - the chromosome-heavy mother cell exists just long enough for the full development and protection of its smaller counterpart, which is destined to be the endospore.

The differentiation process requires both specific gene metabolism and regulation. Most of these processes are controlled by four different sigma factors, two per cell, that are activated at explicit times during sporulation. Figure 4.1 shows a generalized depiction of the factors involved with different stages of division. Initial signals from the bacterium lead to phosphorylation of Spo0A, the master regulator of transcription.[56] Activation of this protein and σ^H trigger asymmetric division and transcription of three Spo proteins that encode for the remaining sigma factors.
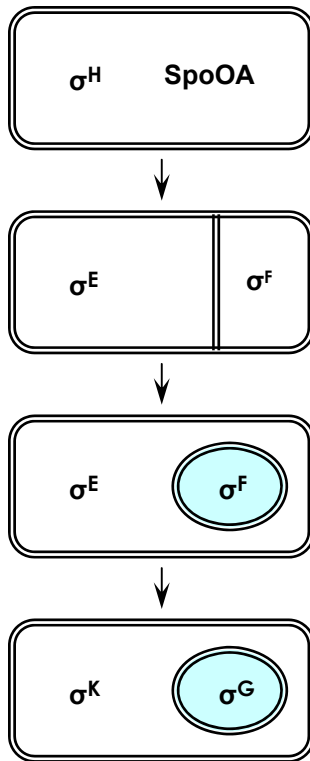
Figure 4.1. Sigma factors involved in sporulation.[56]

Immediately following cell division, the membrane-bound SpoIIE protein phosphatase initiates a cascade of events that result in proliferation of the σ^F RNA polymerase factor.[57] Active only in the forespore, σ^F is responsible for initiating transcription of many genes that are necessary for the development and survival of endospore. It also induces expression of the SpoIIR protein and another sigma factor, σ^G, which will eventually replace it in the forespore.

In the mother cell, the σ^E transcription factor is responsible for the transcription of more than 250 genes.[58] A sampling of activities that occur via some of these genes is listed in Table 6. Initially produced as an inactive, pro- σ^E protein, it is processed to

Functions of σ^E-Controlled Genes
Engulfment of the forespore by the mother cell
Inhibiting a second cell division
Down-regulation of σ^G expression
Synthesis + activation of σ^G and σ^K
Synthesis/assembly of spore cortex
Synthesis/assembly of spore coat
Fatty acid oxidation
Protein degradation
Transport of nutrients from extracellular environment

Table 6. Examples of σ^E-controlled gene functions.[58]

active form by SpoIIR protein after its σ^F –regulated expression.[56] Following conversion

from the proprotein, σ^E down-regulates the genes that led to early sporulation and takes

control of ensuing processes within the mother cell. Once the septum between the two

asymmetric cells is degraded, σ^G and σ^K take over as the forespore and mother cell sigma

factors, respectively.[56]

IV.B. Carbon Sources

All organisms typically acquire the energy needed for metabolic processes by first

breaking down glucose through the process of glycolysis. This pathway produces the

three-carbon molecule pyruvate, a precursor for the tricarboxylic acid (TCA) cycle. By

metabolizing pyruvate to acetyl-CoA, the TCA process is initiated and electrons are made

available to the electron-transport chain, the cell's major source of adenosine triphosphate

(ATP). Under conditions of low glucose concentration, cells subvert a portion of the

TCA cycle via the glyoxylate shunt, using isocitrate lyase (ICL) and malate synthase

enzymes to produce a malate intermediate. This alternative pathway serves many

purposes: it maintains electron transfer, prevents the waste of carbon atoms through CO_2 byproducts, and maintains the availability of malate and succinate.[59] These two intermediates are needed to preserve the concentration of oxaloacetate, whose importance is described later in this text.

A second cellular response to glucose deficiency is the breakdown of fatty acids to provide an alternative carbon source. These long hydrocarbon chains contain most of a cell's stored energy, and the following concerted metabolic mechanism (Figure 4.2) leads to increased acetyl-CoA concentration. Oxidation by FAD results in a double bond between C2 and C3. The second step is a stereospecific hydration reaction that results in
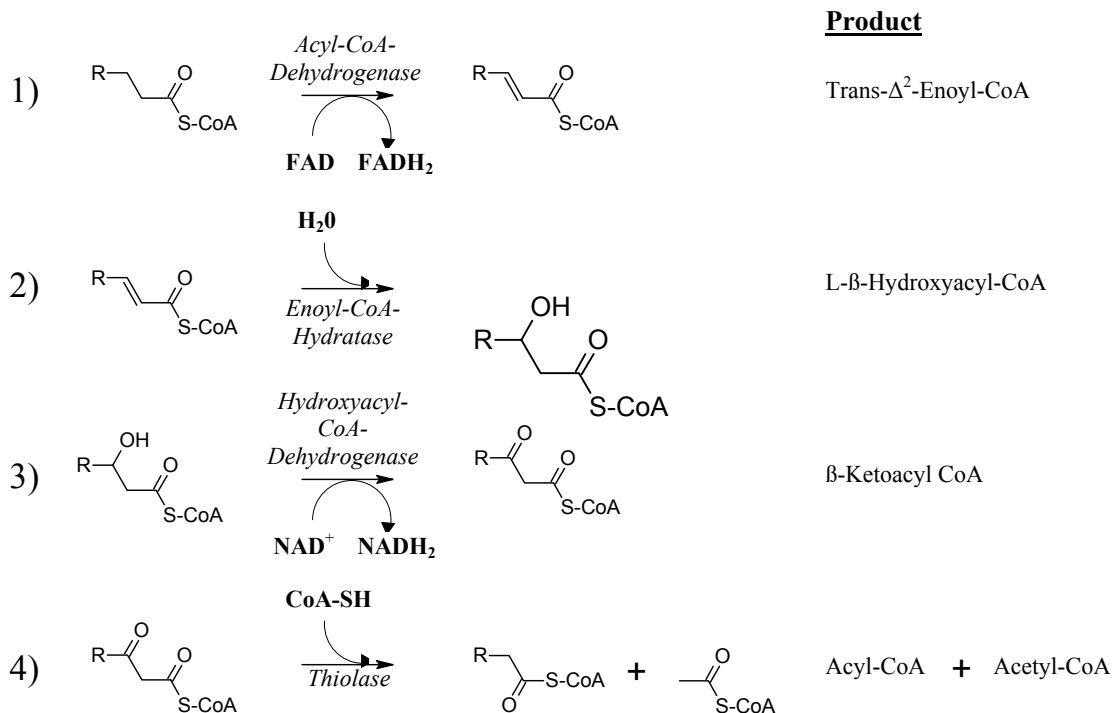

Figure 4.2. The four steps of fatty acid ß-oxidation. Oxidation of the hydrocarbon chain (1) is followed by hydration of the double bond (2), conversion of hydroxyl to a keto group via a second oxidation step(3). Cleavage between C_α and C_β (4) results in either a two- or three-carbon molecule.

a 3-hydroxyacyl-CoA molecule. Thirdly, a second oxidation reaction occurs, resulting in

the conversion of a hydroxyl to keto group. Then bond cleavage occurs between C_α and

C_β to give the two carbon acetyl-CoA.

However, if the fatty acid is branched or has an odd number of carbon atoms, β-

oxidation will ultimately yield a three carbon propionyl-CoA. At high concentrations this

compound has been implicated in cell toxicity, and so additional chemical modification is

required. Several pathways exist that are capable of metabolizing propionyl-CoA. One

of these is the methylcitric acid (MCA) cycle. After conversion of propionyl-CoA to

pyruvate, an oxidative decarboxylation yields the required two carbon acetyl-CoA.

IV.C. *Mmg* Operon

One particular group of σ^E-dependent genes was identified in 1996 by a research

group at Emory University and dubbed the *mmg* operon ("mother cell metabolic

genes").[58, 60] The six open reading frames shown (ORFs) in Figure 4.3 are capable of

encoding genes similar to those used in fatty acid metabolism and the methylcitric acid

cycle.[60] Activated during sporulation, *mmg* breakdown of fatty acids is a means for

attaining energy to drive the cell's preparation for dormancy.

Figure 4.3. The *mmg* operon. Included are five genes, the *mmgO* CRE region and the downstream *yqiQ*. The sites indicated above, *1 and *2, indicate the approximate locations for primer design, described in Chapter 5.

Each *mmg* gene is expected to play a different role in the β-oxidation of fatty

acids as listed in Table 7, with the end result being pyruvate production. Expression and

MMG OPERON	
Gene	Function
mmgA	Acetyl-coenzyme A (CoA) acetyltransferase [Thiolase]
mmgB	2-Hydroxybutyryl-CoA dehydrogenase
mmgC	Acyl-CoA dehydrogenase
mmgD	Bifunctional (methyl)citrate synthase
mmgE	*2-Methylaconitate hydratase
yqiQ	Similar to methylisocitrate lyase

Table 7. Genes of the *mmg* operon and their functions. These are as described by Eichenberger et al., with one change. The function for *yqiQ* as designated by A.L. Sonenshein is listed.[58, 61]

*The function of *mmgE* is currently being assessed in the J.J. Reddick research lab.

activity assessments for the first three genes have been performed by the Reddick

research group, and work with both *mmgD* and *mmgE* are currently underway.[62, 63] The

promoter associated with this operon is dependant on σ^E RNA polymerase and so, in the

presence of glucose, the processes of sporulation and *mmg* gene transcription are

suppressed.[60] Also within this promoter, the Bryan group discovered a catabolite-

repressive element (CRE) responsible for regulating this suppression and termed it

mmgO.[60]

IV.D. YqiQ Protein

The MCA metabolic pathway (Figure 4.4) is a mirror of the glyoxylate shunt; it is

a seven-enzyme process that metabolizes propionyl-CoA to produce pyruvate via α-

oxidation. Propionyl-CoA and oxaloacetate are joined to form 2-methylcitrate, which is
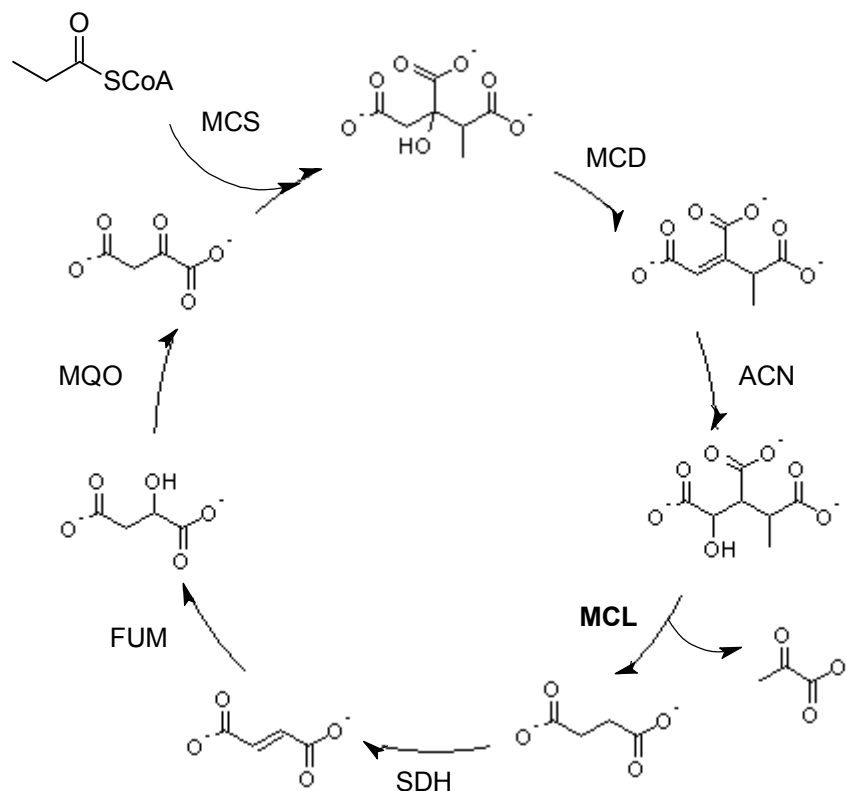
Figure 4.4. The methylcitrate cycle. Propionyl-CoA is metabolized to succinate and pyruvate by 7 enzymes: MCS (methylcitrate synthase), MCD (methylcitrate dehydrogenase), ACN (aconitase), MCL (methylisocitrate lyase), SDH (succinate dehydrogenase), FUM (fumarase), MQO (malate:quinone oxidoreductase).[64]

then converted to 2-methylisocitrate to prepare for decarboxylation. The next reaction results in a carbon-carbon bond cleavage, and a possible homolog of the enzyme responsible for this catalysis is found downstream of the *mmg* genes.

At the end of the *mmg* operon is the *yqiQ* gene, which encodes an approximately 300 amino acid sequence that Sonenshein describes as being homologous to a methylisocitrate lyase (MCL).[61] These enzymes catalyze the reaction shown in Figure 4.4, an aldol cleavage of 2-methylisocitrate to yield succinate and pyruvate.
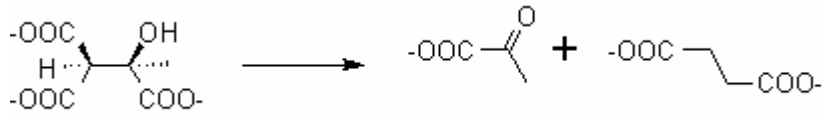
Figure 4.5. Reaction catalyzed by methylisocitrate lyase.[64]

This reaction provides pyruvate for the electron transport chain; succinate is converted first to fumarate and then oxaloacetate, replenishing the intermediates needed for the MCA cycle.

Though *B. subtilis* responds in many ways to nutrient shortage – i.e. lipid degradation, inhibition of non-critical cellular processes, sporulation – it lacks a glyoxylate shunt. Additionally, the methylcitrate cycle has never been demonstrated in the bacterium.[61] Because of this, differentiating *B. subtilis* cells that have not entered sporulation typically exhibit no growth in glucose-deficient media.

CHAPTER V

THE *MMG* OPERON OF *BACILLUS SUBTILIS*

V.A. Goals

It is our hypothesis that by turning on the *mmg* genes, the bacterium will begin to utilize the MCL cycle to break down fatty acids and harness their stored energy. These wild-type bacteria are unable to rely on propionate for energy. By activating these genes during early culture growth, our supposition is that *B. subtilis* will be capable of utilizing propionate as the primary carbon source. By developing a conditional knockout mutation of the *mmg* operon, the bacterium should exhibit *in-vitro* growth in propionate media by induced MCA cycle utilization. First, we planned to replace the *mmg* promoter with an IPTG-inducible promoter. We would then then be able to conduct growth studies using media that rely on various types of carbon sources. Our second goal was to complete the cloning, overexpression, and purification of the YqiQ protein.

V.B. Results

V.B.1. pMUTIN4 Plasmid

In order to develop a conditional knockout mutation of Mmg, the pMUTIN4 cloning vector developed by Valérie Vagner et al. was used (Figure 5.1).[65] Along with three other plasmids, this vector was constructed specifically for *B. subtilis* ORF characterization; all four are unable to replicate in the bacterium, a requirement for insertional mutagenesis.[65] Once cloned into pMUTIN4, the *mmg* promoter would be
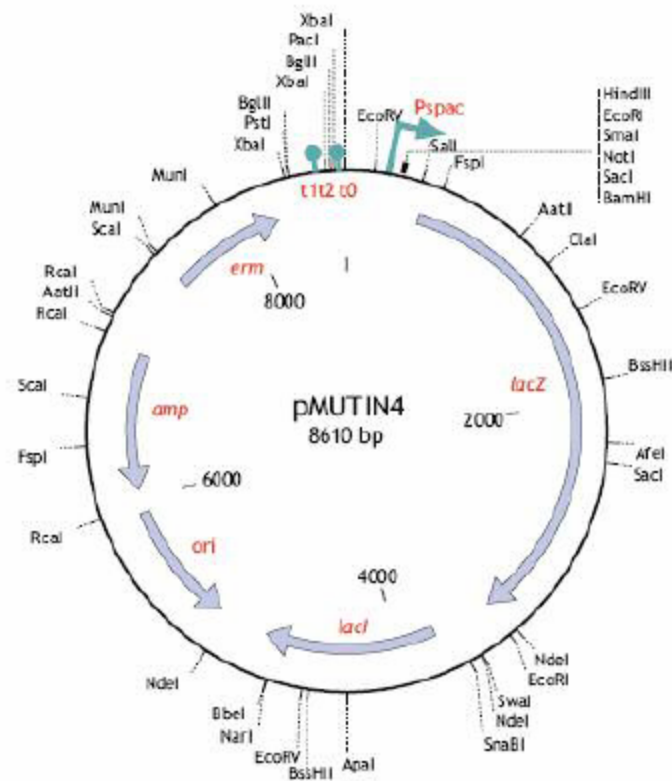
Figure 5.1. The pMUTIN4 vector.[65]

transcriptionally fused to the *lacZ* reporter gene and placed under control of an IPTG-inducible Pspac promoter. The vector contains multiple restriction endonuclease sites along with both an ampicillin and erythromycin resistance, enabling useful transformant selection.

Though it can not replicate in *B. subtilis*, a ColE1 origin site means that it can replicate in *E. coli* cells.[65] By separating the *mmg* genes from their σ^E-dependant promoter, it should be possible to inducing their transcription via Pspac using IPTG. This separation and fusion to the *lacZ* gene occur due to a single cross-over event (Figure 5.1.)
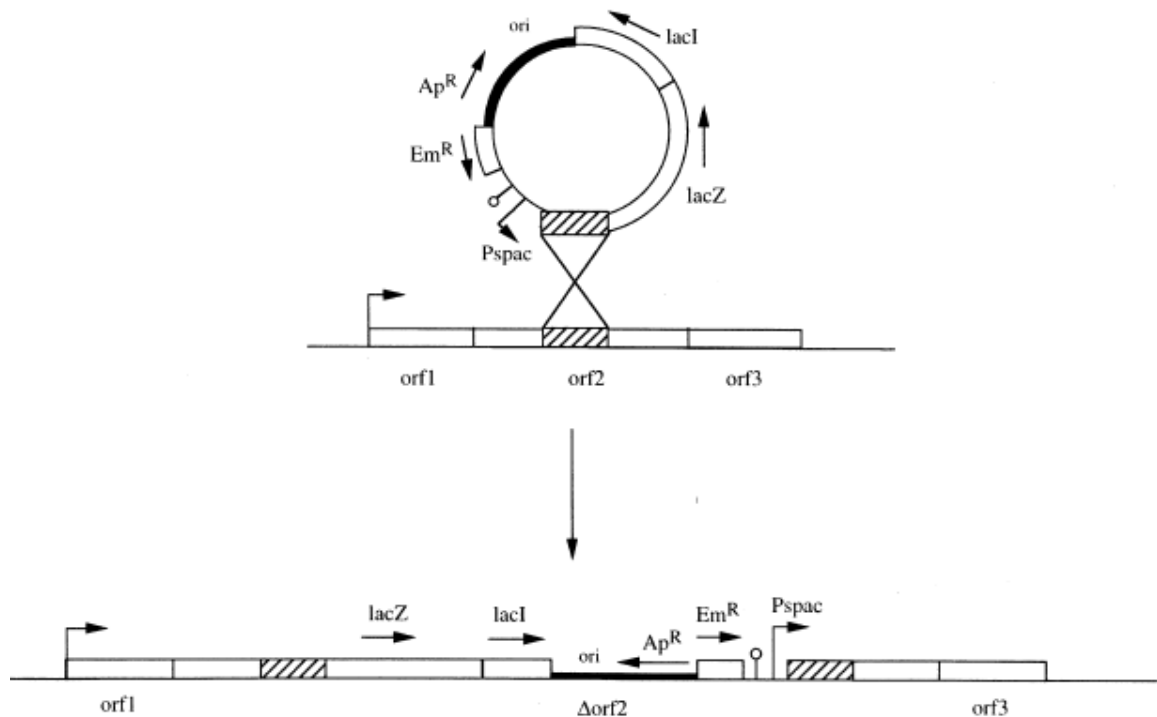
Figure 5.2. Cross-over event for pMUTIN integration.[65]

V.B.2. *Mmg* DNA Amplification and Cloning

<u>*Eco*RI restriction site:</u> Two sets of PCR primers were required for the *mmg* operon in Figure 4.3; the first set was intended to amplify a region upstream of *mmgA*, including the *mmgO* promoter, and the second set for a sequence area just before *mmgC*. Initially, primer sets pM1 and pM2 (Table 8) were designed to amplify 210 bp and 211 bp regions, respectively. Using *Taq* Master Mix polymerase and *B. subtilis* 168 as template DNA, both pM PCR reactions were successful as determined by the agarose gel in Figure 5.3.

PCR Primers	
pM1a	5'-TAG AAG CGT CTC A<u>AA GCT TTT</u> GAA GC – 3'
pM1b	5'– GGT TTC ACC T<u>GA ATT C</u>CA TTC AGA AGA TAG – 3'
pM2a	5'– GCT GAT AT<u>A AGC TT</u>G GCG ACT AG – 3'
pM2b	5'– CGC C<u>GA ATT CAT</u> CGC GAA G – 3'
M1A	5' – ACC GA<u>A AGC TT</u>C TTT CCA GAA AAG GCT TC – 3'
M1B	5' – AAC TCC G<u>GA ATT C</u>CC GAA TTT GCC AAA TG – 3'
M2A	5' – GGA TA<u>A AGC TT</u>C GTG TTT ATC AAT TAT GGA AG 3'
M2B	5' – CCG CA<u>G AAT TC</u>G CGG GAG CAA TTT C – 3'
*M1B	5' – GAG <u>GCG GCC GC</u>G TTT TGA AA – 3'
*M2b	5' – GAC GCA AA<u>G CGG CCG C</u>TT TA – 3'
G1 Forward	5' – CAG <u>AAG CTT</u> TGT TTA CCG ACT TTC CA – 3'
G1 Reverse	5' – GTT T<u>GG ATC C</u>AC ATC ATC TCC AGA GA – 3'
G2 Forward	5' – CCT <u>AAG CTT</u> AGG CCT TGG CGA TTC – 3'
G2 Reverse	5' – GAT CAG CT<u>G GAT CC</u>G AAA TTC GTC C – 3'

Table 8: Primers used for *mmg* PCRs.
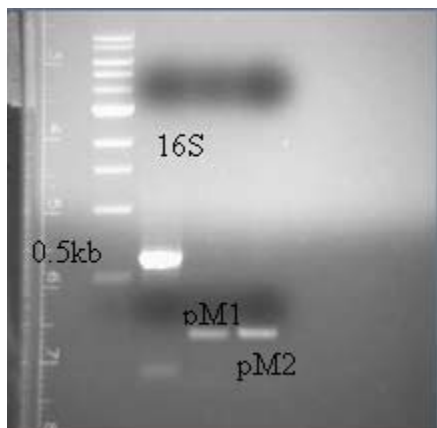
Figure 5.3. pM1 and pM2 PCR products. Primers targeting 16S rDNA was used as a positive control.

Utilizing *Hind*III and *Eco*RI restriction sites in the primers and pMUTIN4 plasmid, restriction digest was performed on both samples and the plasmid. Cloning was attempted with each PCR product, and of 16 insert:plasmid ratios four produced colony growth after ligation and transformation. For pM1, there was a single colony on plate

5:5; pM2 had colonies on plates 4:4, 3:6, and 6:3. Individual colonies were placed in overnight, ampicillin-containing cultures. These were purified and screened for the applicable inserts.

Three screening techniques were used: first, PCR using the initial conditions for amplification. Next, PCR was repeated with a lower annealing temperature of 45ºC to allow for nucleotide deletion that may have occurred due to primer design and restriction site cleavage. Lastly, a second restriction digest was performed and the products loaded directly into a 1% agarose gel to verify appropriate cleavage of plasmid and inserts. None of these yielded any evidence of the insert, though with the third technique the plasmid was apparent by gel electrophoresis.

Based on this, two more primer sets were designed. The first, M1, amplified a region in close proximity to pM1, now 231 bp and still containing both the promoter region and *mmgO*. The second, M2, consisted of a 261 bp region located just before the ribosomal binding region of *mmgC*. As seen in Table 2, these new primers placed the restriction enzyme sites much closer to the 5' end of the sequence. Both regions were amplified from *B. subtilis* strain 168 using similar PCR parameters as before, with a new annealing temperature of 52ºC to account for higher T_M's of the new primers. Agarose gel electrophoresis verified the amplification of each sample.

After purification, both reactions were transfected into retransformed pMUTIN4/*E. coli* cells. Restriction digest and ligation were performed as above, with one experimental change. During both PCR and culture purification, the final product was obtained by using 30 μL of elution buffer EB rather than the recommended 50 μL.

This was done to increase the concentration of the sample. Again, though certain

plasmid: insert ligation reactions resulted in colony growth, none of these were positive

for the M1 or M2 insert.

*Not*I restriction site: A different restriction map of the pMUTIN4 plasmid was

obtained from the company that provided the plasmid itself. This map, Figure 5.1, shows

a second *Eco*RI restriction site downstream of the *lacZ* operon. Due to the likelihood of

extra cutting during digestion procedures, primers M1B and M2B were re-designed to

contain the *Not*I restrictrion site. These were designated *M1B and *M2B (Table 7), and

all cloning, digest, and ligation were repeated with these new primers. PCRs with *Not*I-

containing reverse primers resulted in multiple bands on agarose gels. Multiple attempts

to clone PCR products into the pMUTIN4 plasmid were unsuccessful. It was then noted

that the designed restriction site was the complement of *Not*I, rather than the reverse

complementary sequence required for successful ligation.

*Bam*HI restriction site: Both primer sets were redesigned to include *Hind*III in

the forward primer, and *Bam*HI in the reverse. PCRs resulted in successful bands as

verified by agarose gel. A new stock of the pMUTIN4 plasmid was ordered, and PCR

products were cloned into both the plasmid, and retransformed pMUTIN4/*E. coli* cells.

A positive control reaction was performed, where the plasmid was cut with *Hind*III only,

ligated and transformed into DH5α cells. It was then determined that no ligation reaction

was successful.

V.C. Conclusions and Future Work

Amplification of the two *mmg* DNA regions by PCR were accomplished using a variety of primer sets. Successful reactions included the *Hind*III restriction sequence in a forward PCR primer and either *Eco*RI, *Not*I, or *Bam*HI in the reverse. One of two different pMUTIN4 maps indicates a second *Eco*RI restriction site that would cause excess cutting of the plasmid during digest, and prevent successful ligation reactions. While amplification was achieved using *Not*I, multiple bands were present for each reaction. An incorrect sequence in the primer inhibited all ligation attempts.

Buffer incompatibility of the *Hind*III and *Bam*HI restriction enzymes required a sequential digest protocol. Although this was achieved, a positive control reaction showed that no ligation was occurring between the plasmid and inserts. Future work will focus on the optimization of these reaction conditions (i.e. annealing temperature, ligase and buffer composition, etc.)

V.D. Experimental

Primer sets pM1 and pM2 were used for PCR with *B. subtilis* 168 under the following PCR conditions: a 3 minute hot start at 94ºC; 30 cycles of a 1 minute denaturation step at 94ºC, 1 minute annealing at 50ºC, and a 1 minute elongation at 72ºC; this was followed by a 10 minute hold at 72ºC, and the reactions were stored at 4ºC. Successful amplification was verified by 1% agarose gel electrophoresis, with each PCR reaction showing an appropriate band at ~200 bp. Both products were purified with the QIAquick PCR Purification Kit.

For plasmid preparation, 1 µL of lab stock pMUTIN 4 was retransformed into *Escherica coli* DH5α cells as follows: to 100 µL of competent DH5α cells, 1 µL of pMUTIN 4 was added, vortexed briefly, and incubated on ice for 30 minutes. This was followed by two incubations of 2 minutes at 37ºC, and 10 minutes at room temperature. Next, 1 mL of LB was added and the mixture incubated at 37ºC for 1.5 hours. After a 5 minute centrifugation at 8,000 rpm, 1 mL of supernatant was removed, cells were resuspended in remaining solution, and 20 µL plated on each of two LB agar plates containing 100 µg/mL of ampicillin. After overnight incubation at 37ºC, both plates showed growth of numerous colonies. Each subsequent time the plasmid was used for restriction digest procedures, a single colony was placed in an overnight culture containing ampicillin, and purified the next day.

Both PCR products, along with the purified pMUTIN 4 plasmid, were digested at 37ºC for 3.5 hours using *Hind*III and *Eco*RI restriction enzymes, and *Eco*RI buffer. They were then deactivated at 65ºC for 20 minutes. The plasmid digest was loaded into a 0.7% low-melting agarose gel and run at 120 Volts; the resulting band was excised and placed in a 75ºC water bath until the agarose had melted. The two insert digests, PCR products pM1 and pM2, were again purified with the QIAquick® PCR Purification Kit. Using T4 DNA ligase, and with 16 different plasmid: insert ratios, ligation was performed overnight with thermocycler settings of: 30 minutes at 37ºC, 2 hours at 23ºC, and 16ºC for the remainder of the night.

The following morning, all reactions were heat-deactivated for 20 minutes at 65ºC, and transformed into DH5α as described above, using 20 µL of the ligation product

with 100 µL of competent DH5α cells. The resulting mixtures were plated on LB-agar

plates containing 100 µg/mL of ampicillin, and observed the following day for growth.

Of the 16 insert:plasmid ratios attempted, four produced colony growth after

ligation and transformation. For pM1, there was a single colony on plate 5:5; pM2 had

colonies on plates 4:4, 3:6, and 6:3. Individual colonies were placed in overnight,

ampicillin-containing cultures. These were purified and screened for the applicable

inserts. Three screening techniques were used: PCR using the conditions described

above, PCR with a lower annealing temperature of 45ºC, and restriction digest with the

products loaded directly into a 1% agarose gel to verify appropriate cleavage of plasmid

and inserts. No insert was detected, but the plasmid was successfully cut by

endonucleases.

Using primer sets M1 and M2, PCRs were repeated as above, with an annealing

temperature of 52ºC to account for higher T_M's of the new primers. Agarose gel

electrophoresis verified the amplification, and restriction digest, ligation and

transformation were attempted as before. Again, despite some colony growth following

transformation, neither *mmg* insert was detected.

A new stock of the pMUTIN4 plasmid was ordered, and primer sets G1 and G2

were designed to contain the *Hind*III and *Bam*HI restriction sites. Using *B. subtilis* 168

as the template, PCR conditions were: one 2 min hold at 94.0 ºC followed by 30 cycles of

30 s at 94.0 ºC, 30 s at 56.0 ºC, and 1 min at 68.0 ºC, with a final 5 min hold at 68.0 ºC.

Agarose gel electrophoresis verified amplification of ~340 bp for both reactions, and

PCR products were purified.

A restriction map of the pMUTIN4 plasmid was done by incubating a 25 µL reaction at 37.0 ºC for 1 hour. The mixture contained 14.0 µL of a purified plasmid starter culture, 7.2 µL H₂0, 0.3 µL BSA, 1.0 µL of *Hind*III or *Bam*HI, and 2.5 µL of buffer 2 or *Bam*HI buffer, respectively. Both were run in an agarose gel, along with a control reaction containing no enzyme, and showed successful cutting of the plasmid by both enzymes in 1 hour.

Sequential restriction digest was performed with the plasmid and G1/G2 PCR products. For pMUTIN4, an initial 10 µL reaction (8.5 µL plasmid), 1.0 µL buffer 2, 0.3 µL BSA, 0.5 µL *Hind*III) was incubated at 37.0 ºC for 1 hr, then heat-deactivated for 20 min at 65.0 ºC. Each was mixed with 33.2 µL ddH₂0, 5.0 µL BamHI buffer, 0.3 µL BSA, and 1.5 µL *Hind*III, and incubated at 37.0 ºC for 1 hr.

For the G1 and G2 purified PCR products a larger reaction volume was necessary for increased concentration of the inserts. The initial digest volume was 25 µL, and was purified with the QIAquick PCR Purification Kit following heat-deactivation. A second 25 µL digest reaction was performed with *Bam*HI. All three products were run in a 1% agarose gel, the bands excised and purified.

Two ligation procedures were performed on the purified restriction digest products. The Quick Ligation™ Kit from New England BioLabs®, Inc. requires a 5 min incubation at room temperature to ligate a plasmid with a 3 molar excess of insert. Alternatively, multiple plasmid:insert ratios were incubated with T4 DNA ligase at thermocycler settings of: 30 min at 37 ºC, 2 hr at 23.0 ºC, and 16.0 ºC overnight. The next day all reactions were heat-deactivated for 20 min at 65.0 ºC.

All ligation reactions were transformed into TOP10 cells per manufacturer's instructions. As a positive control, pMUTIN4 was digested with *Hind*III and ligated with both techniques. G1/G2 transformation reactions were unsuccessful, and an excess of the positive control reaction yielded only one transformant.

The pMUTIN4 plasmid was retransformed into *E. coli* DH5α cells, plated on LB/ampicillin plates, and incubated overnight at 37 ºC. One colony was used to inoculate an overnight starter culture and the plasmid was purified the following day. A single restriction digest reaction with pMUTIN4 and *Hind*III was repeated as a positive control reaction, and the product transformed into both TOP10 and competent DH5α cells. Using an excess volume of the reaction again yielded only a single transformant; from this it was determined that the ligation step, rather than transformation, was unsuccessful.

V.E. Cloning, Overexpression and Purification of the YqiQ Protein

A second objective of this research was to complete the cloning, overexpression, and purification of the YqiQ protein. From lab stock, the BL21(DE3) plasmid containing the *yqiQ*/pET-28a plasmid (previously prepared in the Reddick laboratory by Suzette Mills) was streaked on an LB agar plate containing kanamycin antibiotic at a 25 µg/mL concentration. One colony from the plate was grown overnight in a kanamycin-containing culture; the following day, 2 mL of the culture was added to 1 liter LB and shaken at 225 rpm (37ºC) until an Absorbance of 0.5 was obtained at 595 nm. The culture was then induced with 0.2382 grams of isopropyl ß-D-1-thiogalactopyranoside (IPTG), inactivating the *lac* repressor and promoting the synthesis of the YqiQ protein.

After shaking overnight at 225 rpm and 37ºC, the culture was centrifuged at 6,500 rpm for 30 minutes, supernatant decanted, and the cells stored at -80ºC.

Cells were resuspended in 20 mL of a binding buffer containing Tris (20 mM), NaCl (0.5 M), and imidizole (5 mM), at pH 7.9. This was applied to a Ni^{2+}-affinity column, and washed with 60 mM imidizole. Using 1 M imidizole, 1 mL elutant fractions were collected and a protein concentration was determined for each using the Bradford method. Fractions 2 and 3 contained the highest concentration, ~2.4 and 2.0 mg/mL respectively, and were run on an SDS-PAGE gel to verify the YqiQ protein (Figure 4). Each showed a very heavy band at ~25-30 kDa, and were combined for dialysis to further purify the samples. This was achieved overnight in a 25 mM Tris buffer, pH 8.0.

The retentate was highly precipitated and so filtered the following day, leaving ~2.5 mL of product. The Bradford assay of this sample gave a protein concentration of 7mg/mL, indicating that the supernatant still contained the protein. For storage, 250 µL of 80% glycerol was added to give a final glycerol-weight concentration of 10%, and the samples stored at -80ºC. Activity characterization will be carried out in the future by a member of the Reddick research group.

REFERENCES

1.   Seow, K. T.; Meurer, G.; Gerlitz, M.; WendtPienkowski, E.; Hutchinson, C. R.; Davies, J., 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. *Journal of Bacteriology* **1997,** 179, (23), 7360-7368.

2.   Decker, T.; Hipp, S.; Ringshausen, I.; Bogner, C.; Oelsner, M.; Schneller, F.; Peschel, C., Rapamycin-induced G(1) arrest in cycling B-CLL cells is associated with reduced expression of cyclin D3, cyclin E, cyclin A, and survivin. *Blood* **2003,** 101, (1), 278-285.

3.   Schwecke, T.; Aparicio, J. F.; Molnar, I.; Konig, A.; Khaw, L. E.; Haydock, S. F.; Oliynyk, M.; Caffrey, P.; Cortes, J.; Lester, J. B.; Bohm, G. A.; Staunton, J.; Leadlay, P. F., The Biosynthetic Gene-Cluster for the Polyketide Immunosuppressant Rapamycin. *Proceedings of the National Academy of Sciences of the United States of America* **1995,** 92, (17), 7839-7843.

4.   Schupp, T.; Toupet, C.; Cluzel, B.; Neff, S.; Hill, S.; Beck, J. J.; Ligon, J. M., A Sorangium-Cellulosum (Myxobacterium) Gene-Cluster for the Biosynthesis of the Macrolide Antibiotic Soraphen a - Cloning, Characterization, and Homology to Polyketide Synthase Genes from Actinomycetes. *Journal of Bacteriology* **1995,** 177, (13), 3673-3679.

5.   Bibb, M. J.; Biro, S.; Motamedi, H.; Collins, J. F.; Hutchinson, C. R., Analysis of the Nucleotide-Sequence of the Streptomyces-Glaucescens Tcml Genes Provides Key Information About the Enzymology of Polyketide Antibiotic Biosynthesis. *Embo Journal* **1989,** 8, (9), 2727-2736.

6.   Lee, S. K.; Basnet, D. B.; Choi, C. Y.; Sohng, J. K.; Ahn, J. S.; Yoon, Y. J., The role of erythromycin C-12 hydroxylase, EryK, as a substitute for PikC hydroxylase in pikromycin biosynthesis. *Bioorganic Chemistry* **2004,** 32, (6), 549-559.

7.   Chen, X. H.; Vater, J.; Piel, J.; Franke, P.; Scholz, R.; Schneider, K.; Koumoutsi, A.; Hitzeroth, G.; Grammel, N.; Strittmatter, A. W.; Gottschalk, G.; Sussmuth, R. D.; Borriss, R., Structural and functional characterization of three polyketide synthase gene clusters in Bacillus amyloliquefaciens FZB 42. *Journal of Bacteriology* **2006,** 188, (11), 4024-4036.

8.   Castoe, T. A.; Stephens, T.; Noonan, B. P.; Calestani, C., A novel group of type I polyketide synthases (PKS) in animals and the complex phylogenomics of PKSs. *Gene* **2007,** 392, (1-2), 47-58.

9.   Pereda, A.; Summers, R. G.; Stassi, D. L.; Ruan, X. A.; Katz, L., The loading domain of the erythromycin polyketide synthase is not essential for erythromycin biosynthesis in Saccharopolyspora erythraea. *Microbiology-Sgm* **1998,** 144, 543-553.

10.   Jackowski, S.; Rock, C. O., Metabolism of 4'-phosphopantetheine in Escherichia coli. In 1984; Vol. 158, pp 115-120.

11.   Carreras, C.; Frykman, S.; Ou, S.; Cadapan, L.; Zavala, S.; Woo, E.; Leaf, T.; Carney, J.; Burlingame, M.; Patel, S.; Ashley, G.; Licari, P., Saccharopolyspora erythraea-catalyzed bioconversion of 6-deoxyerythronolide B analogs for production of novel erythromycins. *Journal of Biotechnology* **2002,** 92, (3), 217-228.

12.   Mironov, V. A.; Sergienko, O. V.; Nastasyak, I. N.; Danilenko, V. N., Biogenesis and regulation of biosynthesis of erythromycins in Saccharopolyspora erythraea. *Applied Biochemistry and Microbiology* **2004,** 40, (6), 531-541.

13.   Rawlings, B. J., Type I polyketide biosynthesis in bacteria (Part A - erythromycin boisynthesis). *Natural Product Reports* **2001,** 18, (2), 190-227.

14.   Weissman, K. J.; Bycroft, M.; Staunton, J.; Leadlay, P. F., Origin of starter units for erythromycin biosynthesis. *Biochemistry* **1998,** 37, (31), 11012-11017.

15.   Rawlings, B. J., Type I polyketide biosynthesis in bacteria (Part B). *Natural Product Reports* **2001,** 18, (3), 231-281.

16.   Scotti, C.; Piatti, M.; Cuzzoni, A.; Perani, P.; Tognoni, A.; Grandi, G.; Galizzi, A.; Albertini, A. M., A Bacillus-Subtilis Large Orf Coding for a Polypeptide Highly Similar to Polyketide Synthases. *Gene* **1993,** 130, (1), 65-71.

17.   Straight, P. D.; Fischbach, M. A.; Walsh, C. T.; Rudner, D. Z.; Kolter, R., A singular enzymatic megacomplex from Bacillus subtilis. *Proceedings of the National Academy of Sciences of the United States of America* **2007,** 104, (1), 305-310.

18.	Chang, Z. X.; Sitachitta, N.; Rossi, J. V.; Roberts, M. A.; Flatt, P. M.; Jia, J. Y.; Sherman, D. H.; Gerwick, W. H., Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium Lyngbya majuscula. *Journal of Natural Products* **2004,** 67, (8), 1356-1367.

19.	Edwards, D. J.; Marquez, B. L.; Nogle, L. M.; McPhail, K.; Goeger, D. E.; Roberts, M. A.; Gerwick, W. H., Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium Lyngbya majuscula. *Chemistry & Biology* **2004,** 11, (6), 817-833.

20.	El-Sayed, A. K.; Hothersall, J.; Cooper, S. M.; Stephens, E.; Simpson, T. J.; Thomas, C. M., Characterization of the mupirocin biosynthesis gene cluster from Pseudomonas fluorescens NCIMB 10586. *Chemistry & Biology* **2003,** 10, (5), 419-430.

21.	Rosenberg, E. R., Eliora; Orr, Elisha; Paitan, Yossi Antibiotic TA biosynthesis gene cluster of Myxococcus xanthus. 2000.

22.	Wu, J.; Hothersall, J.; Mazzetti, C.; O'Connell, Y.; Shields, J. A.; Rahman, A. S.; Cox, R. J.; Crosby, J.; Simpson, T. J.; Thomas, C. M.; Willis, C. L., In vivo mutational analysis of the mupirocin gene cluster reveals labile points in the biosynthetic pathway: the "Leaky Hosepipe" mechanism. *Chembiochem* **2008,** 9, (9), 1500-1508.

23.	Calderone, C. T.; Kowtoniuk, W. E.; Kelleher, N. L.; Walsh, C. T.; Dorrestein, P. C., Convergence of isoprene and polyketide biosynthetic machinery: Isoprenyl-S-carrier proteins in the pksX pathway of Bacillus subtilis. *Proceedings of the National Academy of Sciences of the United States of America* **2006,** 103, (24), 8977-8982.

24.	Butcher, R. A.; Schroeder, F. C.; Fischbach, M. A.; Straightt, P. D.; Kolter, R.; Walsh, C. T.; Clardy, J., The identification of bacillaene, the product of the PksX megacomplex in Bacillus subtilis. *Proceedings of the National Academy of Sciences of the United States of America* **2007,** 104, (5), 1506-1509.

25.	Stein, T., Bacillus subtilis antibiotics: structures, syntheses and specific functions. *Molecular Microbiology* **2005,** 56, (4), 845-857.

26.	Wilson, K. E.; Flor, J. E.; Schwartz, R. E.; Joshua, H.; Smith, J. L.; Pelak, B. A.; Liesch, J. M.; Hensens, O. D., Difficidin and Oxydifficidin - Novel Broad-Spectrum

Antibacterial Antibiotics Produced by Bacillus-Subtilis .2. Isolation and Physicochemical Characterization. *Journal of Antibiotics* **1987,** 40, (12), 1682-1691.

27.   Zimmerman, S. B.; Schwartz, C. D.; Monaghan, R. L.; Pelak, B. A.; Weissberger, B.; Gilfillan, E. C.; Mochales, S.; Hernandez, S.; Currie, S. A.; Tejera, E.; Stapley, E. O., Difficidin and Oxydifficidin - Novel Broad-Spectrum Antibacterial Antibiotics Produced by Bacillus-Subtilis .1. Production, Taxonomy and Antibacterial Activity. *Journal of Antibiotics* **1987,** 40, (12), 1677-1681.

28.   Zweerink, M. M.; Edison, A., Difficidin and Oxydifficidin - Novel Broad-Spectrum Antibacterial Antibiotics Produced by Bacillus-Subtilis .3. Mode of Action of Difficidin. *Journal of Antibiotics* **1987,** 40, (12), 1692-1697.

29.   Albertini, A. M.; Caramori, T.; Scoffone, F.; Scotti, C.; Galizzi, A., Sequence around the 159-Degrees Region of the Bacillus-Subtilis Genome - the Pksx Locus Spans 33-Center-Dot-6 Kb. *Microbiology-Uk* **1995,** 141, 299-309.

30.   Calderone, C. T.; Bumpus, S. B.; Kelleher, N. L.; Walsh, C. T.; Magarvey, N. A., A ketoreductase domain in the PksJ protein of the bacillaene assembly line carries out both alpha- and beta-ketone reduction during chain growth. *Proceedings of the National Academy of Sciences of the United States of America* **2008,** 105, (35), 12809-12814.

31.   Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A. M.; Alloni, G.; Azevedo, V.; Bertero, M. G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Borriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S. C.; Bron, S.; Brouillet, S.; Bruschi, C. V.; Caldwell, B.; Capuano, V.; Carter, N. M.; Choi, S. K.; Codani, J. J.; Connerton, I. F.; Cummings, N. J.; Daniel, R. A.; Denizot, F.; Devine, K. M.; Dusterhoft, A.; Ehrlich, S. D.; Emmerson, P. T.; Entian, K. D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S. Y.; Glaser, P.; Goffeau, A.; Golightly, E. J.; Grandi, G.; Guiseppi, G.; Guy, B. J.; Haga, K.; Haiech, J.; Harwood, C. R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M. F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; KlaerrBlanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S. M.; Levine, A.; Liu, H.; Masuda, S.; Mauel, C.; Medigue, C.; Medina, N.; Mellado, R. P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; Oreilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S. H.; Parro, V.; Pohl, T. M.; Portetelle, D.; Porwollik, S.; Prescott, A. M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlan, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S. J.; Serror, P.; Shin, B. S.; Soldo, B.;

Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H. F.; Zumstein, E.; Yoshikawa, H.; Danchin, A., The complete genome sequence of the Gram-positive bacterium Bacillus subtilis. *Nature* **1997,** 390, (6657), 249-256.

32. Zimmerman, S. B. S., NJ), Wilson, Kenneth E. (Westfield, NJ), Monaghan, Richard L. (Somerset, NJ), Del Val, Sagrario M. (Madrid, ES), Fernandez I, Maria M. (Madrid, ES), Hensens, Otto D. (Red Bank, NJ), Flor, James E. (Bridgewater, NJ), Deriso, Cheryl (Avenel, NJ) Difficidin and derivative antibacterials. 1985.

33. Mootz, H. D.; Finking, R.; Marahiel, M. A., 4 '-phosphopantetheine transfer in primary and secondary metabolism of Bacillus subtilis. *Journal of Biological Chemistry* **2001,** 276, (40), 37289-37298.

34. Nakano, M. M.; Corbell, N.; Besson, J.; Zuber, P., Isolation and Characterization of Sfp - a Gene That Functions in the Production of the Lipopeptide Biosurfactant, Surfactin, in Bacillus-Subtilis. *Molecular & General Genetics* **1992,** 232, (2), 313-321.

35. Quadri, L. E. N.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T., Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **1998,** 37, (6), 1585-1595.

36. Tsuge, K.; Ano, T.; Shoda, M., Isolation of a gene essential for biosynthesis of the lipopeptide antibiotics plipastatin B1 and surfactin in Bacillus subtilis YB8. *Archives of Microbiology* **1996,** 165, (4), 243-251.

37. Moldenhauer, J.; Chen, X. H.; Borriss, R.; Piel, J., Biosynthesis of the antibiotic bacillaene, the product of a giant polyketide synthase complex of the trans-AT family. *Angewandte Chemie-International Edition* **2007,** 46, (43), 8195-8197.

38. Wawrik, B.; Kerkhof, L.; Zylstra, G. J.; Kukor, J. J., Identification of unique type II polyketide synthase genes in soil. *Applied and Environmental Microbiology* **2005,** 71, (5), 2232-2238.

39.   Miller, D. N.; Bryant, J. E.; Madsen, E. L.; Ghiorse, W. C., Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Applied and Environmental Microbiology* **1999,** 65, (11), 4715-4724.

40.   Zhou, J. Z.; Bruns, M. A.; Tiedje, J. M., DNA recovery from soils of diverse composition. *Applied and Environmental Microbiology* **1996,** 62, (2), 316-322.

41.   Brady, S. F., Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. *Nature Protocols* **2007,** 2, (5), 1297-1305.

42.   Spear, J. R.; Ley, R. E.; Berger, A. B.; Pace, N. R. In *Complexity in natural microbial ecosystems: The Guerrero Negro experience*, 2003; 2003; pp 168-173.

43.   Bumpus, S. B.; Magarvey, N. A.; Kelleher, N. L.; Walsh, C. T.; Calderone, C. T., Polyunsaturated Fatty-Acid-Like Trans-Enoyl Reductases Utilized in Polyketide Biosynthesis. *Journal of the American Chemical Society* **2008,** 130, (35), 11614-11616.

44.   Euzeby, J. P., Bacterial Nomenclature - Changes of Veterinary Interest During 1993. *Revue De Medecine Veterinaire* **1994,** 145, (1), 3-31.

45.   Metsa-Ketela, M.; Halo, L.; Munukka, E.; Hakala, J.; Mantsala, P.; Ylihonko, K., Molecular evolution of aromatic polyketides and comparative sequence analysis of polyketide ketosynthase and 16S ribosomal DNA genes from various streptomyces species. *Applied and Environmental Microbiology* **2002,** 68, (9), 4472-4479.

46.   Metsa-Ketela, M.; Salo, V.; Halo, L.; Hautala, A.; Hakala, J.; Mantsala, P.; Ylihonko, K., An efficient approach for screening minimal PKS genes from Streptomyces. *Fems Microbiology Letters* **1999,** 180, (1), 1-6.

47.   Ash, C.; Farrow, J. A. E.; Wallbanks, S.; Collins, M. D., Phylogenetic Heterogeneity of the Genus Bacillus Revealed by Comparative-Analysis of Small-Subunit-Ribosomal Rna Sequences. *Letters in Applied Microbiology* **1991,** 13, (4), 202-206.

48.   Wattiau, P.; Renard, M. E.; Ledent, P.; Debois, V.; Blackman, G.; Agathos, S. N., A PCR test to identify Bacillus subtilis and closely related species and its application to

the monitoring of wastewater biotreatment. *Applied Microbiology and Biotechnology* **2001,** 56, (5-6), 816-819.

49.   Daffonchio, D.; Cherif, A.; Brusetti, L.; Rizzi, A.; Mora, D.; Boudabous, A.; Borin, S., Nature of Polymorphisms in 16S-23S rRNA Gene Intergenic Transcribed Spacer Fingerprinting of Bacillus and Related Genera. In 2003; Vol. 69, pp 5128-5137.

50.   Daffonchio, D.; Borin, S.; Consolandi, A.; Mora, D.; Manachini, P. L.; Sorlini, C., 16S-23S rRNA internal transcribed spacers as molecular markers for the species of the 16S rRNA group I of the genus Bacillus. *FEMS Microbiology Letters* **1998,** 163, (2), 229-236.

51.   Ni, Y. Q.; Wan, D. S.; He, K. Y., 6S rDNA and 16S-23S internal transcribed spacer sequence analyses reveal inter- and intraspecific Acidithibacillus phylogeny. *Microbiology-Sgm* **2008,** 154, 2397-2407.

52.   Sagredo, B.; Jedlicki, E.; Orellana, O. In *Organization of the 16s-23s Intergenic Spacer Region of the 2 Rna Operons from Thiobacillus-Ferrooxidans*, 1992; 1992; pp 239-247.

53.   Idriss, E. E.; Makarewicz, O.; Farouk, A.; Rosner, K.; Greiner, R.; Bochow, H.; Richter, T.; Borriss, R., Extracellular phytase activity of Bacillus amyloliquefaciens FZB45 contributes to its plant-growth-promoting effect. *Microbiology-Sgm* **2002,** 148, 2097-2109.

54.   Reva, O. N.; Dixelius, C.; Meijer, J.; Priest, F. G., Taxonomic characterization and plant colonizing abilities of some bacteria related to Bacillus amyloliquefaciens and Bacillus subtilis. *Fems Microbiology Ecology* **2004,** 48, (2), 249-259.

55.   Lampel, K. A.; Dyer, D.; Kornegay, L.; Orlandi, P. A., Detection of Bacillus spores using PCR and FTA filters. *Journal of Food Protection* **2004,** 67, (5), 1036-1038.

56.   Piggot, P. J.; Hilbert, D. W., Sporulation of Bacillus subtilis. *Current Opinion in Microbiology* **2004,** 7, (6), 579-586.

57.   Hilbert, D. W.; Piggot, P. J., Novel spoIIE mutation that causes uncompartmentalized sigma(F) activation in Bacillus subtilis. *Journal of Bacteriology* **2003,** 185, (5), 1590-1598.

58.   Eichenberger, P.; Jensen, S. T.; Conlon, E. M.; van Ooij, C.; Silvaggi, J.; Gonzalez-Pastor, J. E.; Fujita, M.; Ben-Yehuda, S.; Stragier, P.; Liu, J. S.; Losick, R., The sigma(Epsilon) regulon and the identification of additional sporulation genes in Bacillus subtilis. *Journal of Molecular Biology* **2003,** 327, (5), 945-972.

59.   Gould, T. A.; de Langemheen, H. V.; Munoz-Elias, E. J.; McKinney, J. D.; Sacchettini, J. C., Dual role of isocitrate lyase 1 in the glyoxylate and methylcitrate cycles in Mycobacterium tuberculosis. *Molecular Microbiology* **2006,** 61, (4), 940-947.

60.   Bryan, E. M.; Beall, B. W.; Moran, C. P., A sigma(E)-dependent operon subject to catabolite repression during sporulation in Bacillus subtilis. *Journal of Bacteriology* **1996,** 178, (16), 4778-4786.

61.   Sonenshein, A. L., The Krebs Citric Acid Cycle. In *Bacillus subtilis and Its Closes Relatives: from Genes to Cells*, A. L. Sonenshein, e. A., Ed. ASM Press: Washington, D.C., 2002; pp 151-162.

62.   Reddick, J. J.; Antolak, S. A.; Raner, G. M., PksS from Bacillus subtilis is a cytochrome P450 involved in bacillaene metabolism. *Biochemical and Biophysical Research Communications* **2007,** 358, (1), 363-367.

63.   Reddick, J. J.; Williams, J. K., The mmgA gene from Bacillus subtilis encodes a degradative acetoacetyl-CoA thiolase. *Biotechnology Letters* **2008,** 30, (6), 1045-1050.

64.   Upton, A. M.; McKinney, J. D., Role of the methylcitrate cycle in propionate metabolism and detoxification in Mycobacterium smegmatis. *Microbiology-Sgm* **2007,** 153, 3973-3982.

65.   Vagner, V.; Dervyn, E.; Ehrlich, S. D., A vector for systematic gene inactivation in Bacillus subtilis. *Microbiology-Sgm* **1998,** 144, 3097-3104.