ABSTRACT

NATASHA LEE
Complementation of Chromosomal Deletions in *Mycobacterium tuberculosis*
Under the direction of DR. RUSSELL KARLS

Tuberculosis (TB) is an infectious disease caused by *Mycobacterium tuberculosis* (*Mtb*). The World Health Organization estimates that one third of the world's population is latently infected with *Mtb*; these individuals have a 10% risk of developing TB. To prevent, diagnose, and treat TB, it is imperative to understand the mechanisms this pathogen uses to survive in a host. Defining roles that different *Mtb* genes play in virulence is typically achieved by characterizing mutants deleted for targeted gene(s). To ensure that the resulting phenotypes are associated with the deleted genes, the wild type gene(s) must be re-introduced. This is typically performed by PCR amplification of the genes and cloning them onto plasmids that can replicate in *Mtb*. Amplification of large regions of DNA often results in mutations due to the DNA polymerase used in vitro. The focus of my research has been to complement a 16-kb region of the *Mtb* genome. To help define functions of genes within this region, a series of plasmids were created containing portions of the deleted region obtained from the parent strain of the mutant. Difficulty with amplifying the entire region led us to pursue an alternate strategy of obtaining this region from the chromosome using plasmid-mediated recombination. The first step in developing a plasmid for this purpose was completed. This work is anticipated to allow the generation of complementation plasmids with a low risk of PCR-induced mutations. The plasmids generated will aid vaccine research by facilitating studies to identify which *Mtb* genes encode virulence factors.

INDEX WORDS: *Mycobacterium tuberculosis*, Complementation, Cobalamin Biosynthesis, Homologous Recombination, Bacterium, Plasmid

COMPLEMENTATION OF CHROMOSOMAL DELETIONS IN *MYCOBACTERIUM TUBERCULOSIS*

by

NATASHA LEE

A Thesis Submitted to the Honors council of the University of Georgia

in Partial Fulfillment of the Requirements for the Degree

BACHELOR OF SCIENCE

in MICROBIOLOGY

with HONORS

and CURO SCHOLAR DISTINCTION

Athens, Georgia

2010

COMPLEMENTATION OF CHROMOSOMAL DELETIONS IN

MYCOBACTERIUM TUBERCULOSIS

by

NATASHA GI-YOON LEE

Approved:

Dr. Russell Karls — May 6th, 2010

Dr. Russell Karls — Date
Faculty Research Mentor

Approved:

Benjamin Grosse-Siestrup — May 6th, 2010

Benjamin Grosse-Siestrup — Date
Reader

Approved:

Dr. David S. Williams — May 7th, 2010

Dr. David S. Williams — Date
Director, Honors Program, Foundation Fellows and
Center for Undergraduate Research Opportunities

Approved:

Dr. Pamela B. Kleiber — May 7th, 2010

Dr. Pamela B. Kleiber — Date
Associate Director, Honors Program and
Center for Undergraduate Research Opportunities

ACKNOWLEDGEMENTS

I would like to first and foremost, thank Dr. Karls and Dr. Quinn for their guidance and encouragement over the years. I would also like to thank the following people of the Quinn/Karls lab: Mr. Benjamin Grosse-Siestrup for making me learn how to do things on my own; Ms. Kari Fine for her entertaining stories; and Ms. Shelly Helms for making me feel like a part of the lab from the very beginning.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1
INTRODUCTION

Mycobacterium tuberculosis is an aerobic bacterium that is the causative agent of the infectious disease tuberculosis. The bacterium is characterized, due to the high content of mycolic acids within the cell surface, as an acid-fast bacterium. Mycolic acids make the bacterium difficult to be identified by Gram staining. Four other mycobacteria are in the *M. tuberculosis* complex: *M. bovis*, *M. africanum*, *M. canetti*, and *M. microti* (van Soolingen et al. 1997). *Mycobacterium tuberculosis* has a very low infectious dose; less than 10 organisms are needed to cause infection (Behr et al. 1999). In 2006, 9.2 million people were diagnosed with the disease and 1.7 million deaths were recorded (Park 2008). The World Health Organization approximates that one third of the world's population harbors the bacilli as an asymptomatic, or latent, form and 10% of those infected have the risk of developing the active disease in their lifetime. If left untreated, the death rate for those infected with the active form of the disease is greater than 50% (Mayo Clinic 2009).

Tuberculosis is a bacterial infection that most commonly affects the lungs. Humans typically acquire the infection by inhalation of infectious aerosols. The mycobacteria gain access to the pulmonary alveoli where they invade and replicate within alveolar macrophages. Once the host immune system is stimulated, macrophages, fibroblasts, T-lymphocytes, and B-lymphocytes accumulate around the bacilli and form a structure called a granuloma (Agarwal et al. 2005). Though the intended function of a granuloma is to blockade the bacilli so that the immune cells can destroy it, some bacilli are not eliminated. In some instances, the bacilli are thought to enter

a dormant stage within a granuloma, leading to the latent stage of the tuberculosis infection (Mayo Clinic, 2009). It is possible for a granuloma to break down over time allowing the bacilli to exit out of the dormant stage and begin replicating. This results in the active stage of the infection. The granuloma can also lead to caseous necrosis, cell death within the center of the granuloma (Kaufmann 2002). In addition to infecting the lungs, the bacilli are also capable of infecting other areas of the body when transported through the bloodstream.

Treatment of both latent and active form of tuberculosis is difficult due to the complex structure of the cell wall (Narasimh, Acharya, & Goldman 1970). Many antibiotics are slow to enter into the cell and require long periods of treatment to kill the bacterium due to its slow growth rate. If the bacilli are persistent against frontline antibiotics such as rifampicin and isoniazid, additional drugs must be administered. Antibiotic resistance has led to drug-resistant , which is growing health concern among immunocompromised patients. Currently, the bacillus Calmette-Guérin (BCG) vaccine is given in many parts of the world. However, its use is limited due to its variable effectiveness at preventing pulmonary TB (0-80%; Comstock 1994). Efforts to create more effective vaccines are in progress and include creating augmented BCG vaccines, recombinant peptide vaccines, DNA vaccines, genetically-modified vaccinia viruses, and attenuated *M. tuberculosis* strains (Oplinger 2004).

Perhaps the most-promising approach is to create live-attenuated *M. tuberculosis* bacilli as vaccine. A live-attenuated vaccine is a form of the disease-causing agent with genetic defects. These defects are ones where specific genes with roles in virulence have been modified or deleted. Such modifications ideally render the bacteria avirulent, but typically multiple genes must be modified to severely attenuate the bacteria. Such live-attenuated vaccines are

advantageous because the bacteria retain many surface characteristics for recognition by the host immune system to elicit a cell-mediated immune response. To create a live-attenuated vaccine for TB, it is useful to understand the functions of the *M. tuberculosis* genes involved in virulence.

In previous studies, a mutant *M.tuberculosis* strain lacking a functional *sigC* gene, encoding sigma factor C, was shown to be attenuated in mice (Sun, et al. 2003) and in guinea pigs (Karls, et al. 2006). The *sigC* gene is located within a cluster of genes, some of unknown function and some with homology to cobalamin biosynthesis genes in other bacteria. Cobalamins, such as vitamin B_{12}, are known to be cofactors for important metabolic pathways within a cell (Herbert 1999). Cobalamin biosynthesis genes are also conserved in other pathogenic strains of mycobacteria. It is hypothesized that the cobalamin biosynthesis cluster is important for virulence of *M. tuberculosis*.

A 16-kb region of the *M. tuberculosis* genome including *sigC* and several cobalamin biosynthesis genes was deleted by graduate student Benjamin Grosse-Siestrup in the Karls laboratory (unpublished data). To help determine the roles of the genes located within the cluster, the *M. tuberculosis* mutant lacking the cobalamin biosynthesis region will be complemented with various plasmids, each containing unique portions of the deleted region. These plasmids will be used to determine specific gene function and to identify genes that potentially play a role in virulence of the bacteria. Such information will be useful in the development of a live-attenuated vaccine to prevent tuberculosis.

CHAPTER 2
MATERIALS AND METHODS

The following methods described have been used to produce multiple plasmids that will aid in complementation of deleted genes within the *M. tuberculosis* cobalamin biosynthesis mutant. Sequences of primers used in PCR reactions are listed in Table 1. Specific features of plasmids are highlighted in Table 2.

Preparation of pMV306

Plasmid pMV306, which is able to integrate into the *M. tuberculosis* genome, enables single-copy expression of genes inserted on the plasmid (Kong and Kunimoto). Plasmid pMV306 was prepared to receive insert DNA by digestion with restriction endonucleases *Hin*dIII and *Xba*I to obtain a 3,962-bp fragment. The DNA from the digestion reaction was cleaned and purified using the Zymoclean DNA Clean and Concentrator kit (ZymoResearch, Inc.) and eluted with 10 μL dH_2O. The digested product was then de-phosphorylated with Calf Intestinal Alkaline Phophatase ("CIAP"). The product was separated on a 1% agarose gel by electrophoresis, stained with ethidium bromide, visualized by exposure to ultraviolet light, and compared against DNA size standard (DNA of bacteriophage λ DNA digested with *Bst*EII) to determine the band size and concentration. The concentration of the pMV306/*Hin*dIII/*Xba*I/CIAP product was determined to be 110 ng/μL. The DNA was diluted to a final concentration of 50 ng/μL.

Construction of pGSblaC

Plasmid pGSblaC was created by inserting the *blaC* gene from the *M. tuberculosis* cobalamin biosynthesis region into the vector pMV306. To obtain the ~1,300-bp *blaC* fragment, a Polymerase Chain Reaction (PCR) reaction with primers p1136 and p1135, H37Rv genomic DNA, and Pfu turbo DNA polymerase was created (see Table 1 for primer sequences). The PCR parameters used are as follows:

Temperature (°C)	97°	;	97°	,	57°	,	72°	;	97°	,	65°	,	72°	;	72°	;	4°
Time (min:sec)	4:00	;	0:30	,	0:30	,	0:25	;	0:30	,	0:30	,	0:25	;	3:00	;	hold
			2 cycles						23 cycles								

The PCR product was purified using the Zymoclean DNA Clean and Concentrator kit and eluted twice with 9 µL dH_2O. The product was digested with *Hin*dIII and *Xba*I and purified with the Zymoclean DNA Clean and Concentrator kit and eluted twice with 5 µL dH_2O. The concentration of the fragment was first determined by λ-*Bst*EII DNA comparison, but on the gel, the band was brighter than the standards. As this is outside the linear range, the DNA concentration was then determined using a spectrometer measuring absorbance at 260 nm and calculated to be 278 ng/µL. A ligation reaction was then prepared to insert the *blaC* fragment into the prepared vector, pMV306/*Hin*dIII/*Xba*I/CIAP. To ligate *blaC* into pMV306, 50 ng of the vector was mixed with 18 ng of the *blaC* fragment along with 2 µL 10X ligase buffer and 0.5 µL T4 DNA ligase enzyme (NEB) for a final volume of 20 µL. The ligation reaction was incubated at room temperature for ~12 minutes before being transformed. The ligated DNA was then

transformed into TAM1 chemically-competent *E. coli* cells. The cells were plated on Luria-Bertaini (LB) agar plates supplemented with 50 µg/mL of kanamycin. Colonies that grew on the plates expressed kanamycin resistance. Four candidates (A, B, C, D) were selected from the transformation plates and streaked for isolation. An individual isolated colony for each of the four candidates was grown in 5 mL LB broth containing 50 µg/mL kanamycin. The plasmid was extracted from the *E. coli* cells using a QIAprep Spin Miniprep kit and the DNA was eluted with 50 µL dH_2O. The plasmid DNA was then digested with *Age*I to screen for insertion of the *blaC* gene. The digested products were observed following 1% agarose gel electrophoresis, staining for 5 minutes in ethidium bromide (25 µl of a 10 mg/ml stock diluted into 100 ml dH_20), destaining for 5 minutes in dH_20, exposure to ultraviolet light, and image capture on an Alpha Imager. Proper insertion of the *blaC* fragment should yield 4310- bp and 851-bp bands. Candidate D displayed this pattern and was grown n 100 mL LB broth with 100 µL of a 50 mg/mL kanamycin stock. The DNA was extracted and purified with a QIAgen Midiprep kit and eluted with 1 mL dH_2O for a final concentration of 48 ng/µL. A *Bam*HI digest was performed to confirm the plasmid. Candidate D exhibited the correct *Bam*HI digestion band pattern, 3574 bp and 1816 bp. Candidate D was sent to Genewiz, Inc for DNA sequencing. No mutations were detected in the inserted fragment.

Construction of pGSblaC-sigC

The plasmid pGSblaC-sigC was created by inserting a 1795-bp fragment, which contained the genes *blaC* and *sigC* of the *M.tuberculosis* genome, into the vector pMV306. To

obtain the *blaC-sigC* fragment, primers p1134 and "HindIII-92downsigC" were used in a PCR reaction with the following cycle with a 50°- 60° temperature gradient:

Temperature (°C)	97°	;	97°	, 50°-60°,	72°	;	97°	, 66°	, 72°	;	72°	; 4°
Time (min:sec)	4:00	;	0:30	, 0:30	, 0:26	;	0:30	, 0:30	, 0:26	;	3:00	; ∞
			2 cycles				23 cycles					

After observing the results from the temperature gradient, the fragment was amplified by running another PCR reaction with the following cycle:

Temperature (°C)	97°	;	97°	, 58.5°	, 72°	;	97°	, 66°	, 72°	;	72°	; 4°
Time (min:sec)	4:00	;	0:30	, 0:30	, 0:26	;	0:30	, 0:30	, 0:26	;	3:00	; ∞
			2 cycles				23 cycles					

The DNA product was visualized as described previously. The product was purified using a Zymoclean DNA Clean and Concentrator and eluted twice with 9 µL dH_2O. The purified product was then digested with *Hin*dIII and *Xba*I. The digested DNA was separated on a 1% agarose gel and the concentration was determined by comparing the fragment DNA to λ-*Bst*EII DNA standards. The final concentration of the fragment was estimated to be 35 ng/µL. A ligation reaction was prepared to insert the *blaC-sigC* fragment into the pMV306 vector. To ligate *blaC-sigC* into pMV306, 50 ng of pMV306*Hin*dIII/*Xba*I/CIAP was mixed with 22 ng of the *blaC-sigC* fragment along with 2 µL 5X Ligase buffer and 0.5 µL T4 DNA ligase enzyme (Invitrogen) for a final volume of 20 µL. The ligated DNA transformed into TAM1 chemically-competent *E.coli* cells. The cells were plated on LB agar plates supplemented with 50 µg/mL of kanamycin. Four different candidates (A, B, C, D) were streaked for isolation on fresh plates. Two candidates (C,

D) from the isolation plates were grown in 5 mL LB + 50µg/mL kanamycin. The plasmid was extracted from the *E. coli* cells using a QIAprep Spin Miniprep kit; the DNA was eluted with 50 µL dH_2O. The plasmid DNA was then digested with *Bam*HI to screen for insertion of the *blaC-sigC* fragment. The digested products were observed on a 1% agarose gel. Proper insertion of the *blaC-sigC* fragment should yield 3579-bp and 2180-bp bands. Candidate D exhibited the correct banding pattern and the plasmid was amplified in 150 mL LB broth with 50 µg/ mL kanamycin. The DNA was extracted and purified with a QIA Midiprep kit (Qiagen, Inc.) and eluted with 600 µL dH_2O for a final concentration of 91 ng/µL and sent to Genewiz, Inc for DNA sequencing. The sequence files showed that there was a mutation within the fragment. Due to this mutation, the procedure for construction was repeated and a final candidate was sent to be sequenced. No mutations were noted on the second candidate sequenced.

Construction of pGScob

The plasmid pGScob should contain the entire cobalamin region and is intended for complementation of the 16-kb cobalamin deletion mutant of *M. tuberculosis.* The plasmid should contain 4 different fragments comprising the cobalamin deletion region and flanking regulatory sequences: fragment 1 (2463 bp), fragment 2 (5941 bp), fragment 3 (7831 bp), and fragment 4 (2945 bp). Fragment 1, 2 and 3 were already contained within plasmid pGScob123. However, pGScob123 did not contain gene *int*, which encodes for the gene for the L5 att/int region to enable the plasmid to integrate into the *M. tuberculosis* genome.

To create pGScob, a DNA fragment containing cob regions 1, 2, and 3 was generated by digesting plasmid pGScob123with *Avr*II and *Xba*I. Fragment 123, 15,867bp in size, was excised

following 1% agarose gel electrophoresis and purified from the gel using an Invitrogen Gel purification kit. The DNA was eluted with 50 µL TE buffer at 65°C. The final concentration of fragment123 was 40 ng/µL. Fragment 4 was obtained by PCR of *M. tuberculosis* genomic DNA with primers p1166 and p1167 and the following cycling conditions:

Temperature (°C)	97°	;	97°	,	56°	,	72°	;	97°	,	65°	,	72°	;	72°	;	4°
Time (min:sec)	3:00	;	0:30	,	0:30	,	0:45	;	0:30	,	0:30	,	0:45	;	5:00	;	∞
			2 cycles						23 cycles								

Following electrophoresis, a band of the correct size, 2945 bp, for fragment 4 was detected and the PCR product was purified with the Zymoclean DNA Clean and Concentrator kit and eluted twice with 13 µL dH_20. Fragment 4 was then digested with *Avr*II, dephosphorylated with CIAP, purified with the Zymoclean DNA Clean and Concentrator kit, and eluted twice with 10 µL dH_2O. Fragment 4/*Avr*II/CIAP was then digested with *Hin*dIII. The concentration of fragment 4 was determined by comparing the DNA to λ-*Bst*EII DNA standards. Efforts to transform the ligation products of fragment 123 and fragment 4 were unsuccessful; likely due to the large size of the fragment 123 (15,867bp). Construction of pGScob was terminated and another method for complementation was discussed.

Sequence determination of pMB351

The plasmid pMB351 was created by another laboratory (Beggs, et al. 1995), but the entire sequence of this plasmid was not published. This plasmid may be useful for complementation studies because it uses the LR7 origin which is different than the pAL5000 origin that is routinely used for maintaining plasmids in *M. tuberculosis*. Various portions of the

plasmid were sequenced using primers p1205, p1206, p1211, and p1219 (see Table 1 for primer sequences), some of which were designed as a result of primer walking (designed from sequencing results generated to learn the sequence of an adjacent region). The final sequence of the plasmid, including the *kan* gene, was determined.

Construction of pAR7-P_{lacZ}

Plasmid pAR7-P_{lacZ} is being created to insert the *lacZ* promoter from the *E. coli lac* operon into the plasmid pAR7 such that the promoter can drive expression of the *sacB* gene. Plasmid pAR7 was created by Akanksha Rajeurs, another undergraduate in the lab, and was digested with *Pme*I. The DNA product was purified using a Zymoclean DNA Clean and Concentrator and eluted twice with 8 µL dH_2O. The plasmid was de-phosphorylated with CIAP. The concentration of the plasmid was determined by comparing the plasmid DNA to a 1-kb DNA ladder on a 1% agarose gel. The concentration of pAR7 was 269 ng/µL. To obtain the *lacZ* promoter, plasmid pUC19spf was digested with *Pvu*II and *Hin*dIII. The digested DNA was purified using a Zymoclean DNA Clean and Concentrator kit and eluted twice with 18 µL of 65°C dH_2O. The digested DNA product was observed by staining a 3% low-melting-temperature agarose gel with SYBR safe dye and exposed to black light. The correct band size representing the *lacZ* fragment (179 bp) was extracted from the gel. The DNA was purified using an Invitrogen Gel Extraction kit and eluted twice with 8 µL dH_2O. T4 DNA polymerase (NEB) and deoxynucleotides were then added to the *lacZ* promoter fragment to create blunt ends. The DNA product was purified using a Zymoclean DNA Clean and Concentrator kit and eluted twice with 8 µL dH_2O. The concentration was determined by comparing the DNA product against a 1-kb

DNA ladder. The final concentration of the fragment was 8.4 ng/µL. A ligation reaction was prepared to insert the *lacZ* promoter fragment into the *Pme*I site of pAR7. For ligation, 50 ng of the vector was mixed with 1 ng of the *lacZ* promoter fragment and 5 µL 5X Ligase buffer and 1 µl T4 DNA ligase enzyme (Invitrogen). The ligation reaction was incubated at 4°C overnight. The reaction was then cleaned with the Zymoclean DNA Clean and Concentrator kit and eluted twice with 6 µL dH_2O. The ~10 µL of DNA recovered was then digested with *Pme*I to ensure that the only plasmids with inserts remained circular. A portion (3 µL) of digested DNA was then transformed into TAM1 chemically-competent *E. coli* cells. The cells were plated on LB agar plates supplemented with 200 µg/mL hygromycin. Multiple colonies from the plates were selected to be screened for proper insertion of the *lacZ* promoter fragment. These candidates were grown in 5 mL LB broth with 200 µg/ml hygromycin to amplify the plasmid DNA. The DNA was extracted and purified with a QIAgen Miniprep kit and eluted in 50 uL dH_2O. The plasmid DNAs were then digested with *Ase*I and *Not*I to determine the orientation of the *lacZ* promoter fragment. The digested products were observed on a 3% agarose gel stained with ethidium bromide. Proper insertion of the fragment should yield the following banding pattern: 5223 bp, 1597 bp, 1566 bp, and 50 bp. The 50-bp band is diagnostic for the promoter being in the correct orientation. A correct candidate was selected and efforts are in progress to sequence the pAR7-P_{lacZ} plasmid for confirmation.

CHAPTER 3
RESULTS AND DISUSSION

The objective of this project was to develop plasmids that will aid in complementation of gene within a 16-kb genomic region deleted in a *M. tuberculosis* cobalamin biosynthesis cluster mutant (Figure 1). By re-introducing a plasmid that contains a specific gene or gene cluster into the mutant bacterium, it may be possible to restore the wild-type phenotype and identify genes that are associated with that phenotype.

The plasmids that will be used for complementation (pGSblaC, pGSblaC-sigC) were designed by inserting a specific gene or gene cluster into the vector pMV306 (Figure 2). The vector pMV306 is a desirable vector to use for complementation because it contains many important features. The vector contains an *int* gene, which encodes the mycobacteriophage L5 integrase and *attP* site. The integrase enzyme catalyzes the insertion of the plasmid into the *M. tuberculosis* chromosome at the *attB* site. The pMV306 vector also contains an *aph* gene which encodes for the enzyme aminophosphohydrolase; this enzyme confers resistance to the aminoglycoside antibiotic kanamycin. Plasmid pMV306 also contains an origin of replication for *E. coli* which allows the plasmid to replicate in *E. coli*. Varying regions of the cobalamin biosynthesis region were inserted into unique *Hind*III and *XbaI* restriction sites of plasmid pMV306.

Several plasmids containing different portions of the cobalamin biosynthesis region were constructed (Figure 3). The plasmids pGSblaC and pGSblaC-sigC were created by inserting the

blaC gene and *blaC-sigC* gene cluster, respectively, into the pMV306 vector (Figure 3). The insert DNAs were obtained through PCR of *M. tuberculosis* genomic DNA. Both fragments were digested with *Xba*I and *Hind*III and ligated with the pMV306 vector previously digested with the same enzymes and de-phosphylated. The most likely insertions would either be a single insertion or three inserts ligated together. Proper insertion of the *blaC* region or the *blaC-sigC* region was confirmed by restriction analysis following digestion with *Bam*HI. Once a plasmid with the correct banding pattern was obtained, it was amplified and sent for DNA sequencing. During the sequence analysis of a pGSblaC-sigC candidate, a single-nucleotide mutation was observed in the sequence file. This mutation was likely PCR induced. In addition to pGSblaC and pGSblaC-sigC, two additional plasmids were constructed by another lab member and will be used for complementation. Plasmid pGScobK-Rv2077a contains the downstream-most portion (*cobK-Rv2077a*) of the deleted region. The plasmid pGS1551sigC+ ("pGSsigC") contains the *sigC* gene. In the future, these plasmids can individually be inserted into the *M. tuberculosis* cobalamin biosynthesis mutant to determine if the *blaC* gene (gene encoding for the enzyme beta-lactamase), the *blaC-sigC* cluster (cluster containing beta-lactamase and an alternate sigma factor), or the *sigC* gene are sufficient to restore resistance to beta-lactam antibiotics or production of SigC (See Figure 3).

The plasmid pGScob was designed with the intention of complementing the entire 16-kb deleted region with a 19-kb region that includes the deleted region and an additional ~3 kb of flanking DNA that may provide regulatory signals. The entire complementing region was designed to be obtained from four separate PCR reactions from *M. tuberculosis* genomic DNA to generate fragments 1, 2, 3, and 4. Fragments 1, 2, and 3 were ligated together in a plasmid named

pGScob123 which was created by graduate student Benjamin Grosse-Siestrup. This plasmid does not contain the bacteriophage L5 att/int region necessary for integration of the plasmid into the *M. tuberculosis* genome. Efforts were undertaken to extract the cob123 region containing *cob* fragments 1, 2, and 3 from plasmid pGScob123 and insert it, with fragment 4 into the pMV306 vector. Obtaining the cob123 fragment and fragment 4 did not produce any difficulties. However, the difficulty in creating the pGScob plasmid was in ligation of the two fragments with the pMV306 vector and in obtaining successful recombinants following transformation of *E. coli*. Presumably the desired plasmid, which should have been ~23 kb in length, was too large to be successfully transformed. It is know that transformation efficiency decreases with plasmid size.

Due to the unsuccessful cloning attempts to create the plasmid containing the entire cobalamin biosynthesis region and because PCR-induced mutations increase with the length of PCR product generated, another approach was undertaken to obtain the *Rv2059c-cobK* region by recombining it onto an episomal plasmid by homologous recombination. The requirements for the recombination system are i) origins for replication in both *M. tuberculosis* and *E. coli*, ii) an antibiotic resistance gene (*hyg*) to select for transformants receiving the plasmid, iii) a counter-selectable gene (*sacB*) that will function in *E. coli*, and iv) regions of homology that flank the gene or genes to be recombined onto the plasmid. Plasmid pAR7 is being used as the starting point for creating the recombination vector (Figure 4). Plasmid pAR7 has an origin of replication for *E. coli* (oriE), and origin of replication for *M. tuberculosis* (oriM), resistance genes for hygromycin (hyg) and kanamycin (kan), the *gfp* gene (encoding production of green fluorescent protein) and the counterselectable gene *sacB* under control of the mycobacterial Hsp60 promoter.

The *sacB* gene encodes enzyme levansucrase that converts sucrose into a toxic polymer that kills bacteria expressing the gene in the presence of sucrose. Plasmid pAR7 will receive several modifications to produce plasmid pNL1 (Figure 4). First, the *E. coli* lacZ promoter (P_{lacZ}) will be inserted upstream of the *sacB* gene so that this gene can be expressed in *E. coli*. As described in the methods, the *lacZ* promoter fragment was isolated from plasmid pUC19spf as a *Pvu*II/ blunted *Hin*dIII fragment. The fragment was ligated into pAR7 digested with *Pme*I. Following transformation of *E. coli* TAM1 cells, candidates were screened for plasmids having the *lacZ* promoter oriented in the same direction as *sacB*. Candidates were screened by digestion with *Ase*I and *Not*I (Figure 5). Based on the available sequence for pAR7, the following band sizes are expected when pAR7 is digested with both enzymes: 5099 bp, 1597 bp, and 1566 bp. Unexpectedly, four bands were detected instead of three (Figure 5). The 5099 bp band is present, but instead of having two bands in the 1566 to 1597 range, only one band of that size appears along with bands of ~1200 and ~350 bp. Therefore, it appears that the plasmid has an unknown restriction site that cleaves either the 1566 bp or 1597 bp band. As the insertion site for the *lacZ* promoter is within the 5099-bp band, these enzymes can still be diagnostic for determining the orientation of the promoter fragment. In the desired orientation, the *lacZ* fragment increases the 5099-bp band to 5223 bp and produces a second band of 50 bp. Candidate #6 appears to produce the 50-bp band (Figure 5). If the insert went in the opposite orientation, then larger band would be 5143 bp and the smaller band 130 bp. Candidate #3 appears to produce the 130-bp band (Figure 5). Therefore, candidate #6 was sent for DNA sequencing to confirm the orientation and the absence of mutations.

After the promoter fragment is confirmed in the correct orientation, the next step will be to insert the *M. tuberculosis Rv2059* genomic region into the *Not*I site located upstream of the Hsp60 promoter and the *M. tuberculosis cobK* genomic region into the *BstZ*17I site located downstream of the s*acB* gene (Figure 5). Introduction of these regions will provide homologous DNA sequences to enable recombination between the plasmid and the genome upon transformation of plasmid pNL1 into wild type *M. tuberculosis*.

The steps to obtain plasmids containing the entire *M. tuberculosis* genomic region from *Rv2059* to *cobK* are as follows: First plasmid pNL1 is introduced into strain *M. bovis* BCG, which has similar genomic sequences in this region as those in wild type *M. tuberculosis*. Transformants will be obtained by plating on Middlebrook 7H10 agar supplemented with OADC, hygromycin, and kanamycin. Transformants will be streaked for isolation on fresh 7H10 OADC plates containing both antibiotics. Isolated colonies will then be cultured in Middlebrook 7H9 broth supplemented with OADC until stationary phase is reached. In bacteria in which the genomic DNA and the plasmid DNA undergo recombination in both regions of identity, the plasmid will now contain the chromosomal DNA and the chromosome will contain the *sacB* gene. Specifically, in pNL1 the homologous DNA that is upstream of the Hsp60 promoter and downstream of the *sacB* gene corresponds to the *Rv2059* gene and the *cobK* gene on the *M. tuberculosis* chromosome, respectively. If recombination between pNL1 and the *M. tuberculosis* chromosome is successful, the recombined pNL1 will contain the region between *Rv2059* and *cobK*. Plasmid DNA will then be isolated from all transformants. The DNA will then be transformed into *E. coli* TAM1 cells and plated on LB agar containing hygromycin and 10% sucrose. If recombination has occurred successfully in BCG, then *E. coli* transformants that lack

a functional *sacB* gene should grow. Two types of transformants without functional *sacB* genes are expected. First are those we want: *E. coli* receiving plasmids which have previously undergone recombination in BCG in the *Rv2059* and *cobK* regions between the plasmid and the chromosome resulting in exchange of the intervening regions (as these will have lost the *sacB* gene and should survive growth on the sucrose-containing medium). Second are *E. coli* transformants that received plasmid with spontaneous mutations within the *sacB* gene (as these mutations would not make functional levansucrase and would not convert sucrose to a toxic product). To screen transformants for those containing plasmids that have exchanged the *sacB* gene for the *Rv2059-cobK* region, colony PCR will be used. Primers to amplify the *sigC* gene will be used, as the *sigC* gene should only be present in transformants that have the *Rv2059-cobK* region.

Once the plasmid is obtained, it can be used in complementation studies. To evaluate complementation of the entire 16-kb cobalamin biosynthesis mutant, the kanamycin resistance gene will need to be removed to enable this plasmid to be transformed into the *M. tuberculosis* cobalamin biosynthesis mutant along with integrating plasmid pGScobK-Rv2077a.

Because the expected recombination frequency to obtain recombination in both the *Rv2059* region and the *cobK* region is approximately one in a million bacteria, a plasmid encoding the recombinase gene *recA* could be used to increase the recombination frequency necessary in BCG to obtain pNL1 recombinants receiving the *Rv2059-cobK* region. For this to be possible, a plasmid with a different origin of replication than used in pNL1 is needed. Plasmid pMB351 has the pLR7 origin of replication that is different than the pAL5000 origin of replication in pNL1. Plasmid pMB351 also contains a *kan* gene encoding resistance to

kanamycin and an origin of replication site for *E. coli*, oriE. Plasmid pMB351 was created and described by another laboratory (Beggs, et al. 1995). Only the pLR7 origin sequence was published along with a drawing of the plasmid. To use this plasmid, we began by digesting the plasmid with restriction enzymes, but obtained unexpected banding patterns, suggesting that the published plasmid map may be incorrect. Therefore, we decided to sequence the plasmid. Based upon the published drawing of pMB351, the *kan* gene was located upstream of oriE and downstream of the pLR7 origin. The sequencing results for the primers p1205 and p1206 showed that the two primers read into each other rather than reading into an unknown gene. This proved that the *kan* gene was not in the location described by the authors. A second sequencing reaction using primer p1211 was used to read into the oriE region away from the LR7 region. This sequencing result showed a region that was not homologous to the oriE region downstream from a region that showed homology to the LR7 region. Because the sequencing result did not show any homology to the oriE region, a third sequencing reaction was needed to locate the end of the *kan* gene. The third sequencing reaction utilized the primer p1219 and read from the *kan* gene into the oriE region. This sequencing result showed a region that was not homologous to either the pLR7 origin or to the region upstream of oriE. After a BLAST search with both sequencing results from p1219 and p1211, the *kan* gene showed a 97% match to a region of a kanamycin resistance gene in plasmid pSCR1001. Using the four sequencing reactions, the full sequence of the *kan* gene as well as the full sequence of the pMB351 plasmid was determined. A drawing of the revised plasmid map for pMB351 is shown in Figure 6. The complete genomic sequence is shown in Figure 7.

Future Studies

The plasmids pGSblaC, pGSblaC-sigC, pGScobK-Rv2077a, and pNL1 will be used to determine which gene and/or genes are important in restoring the wild-type phenotype of the *M. tuberculosis* cobalamin biosynthesis deletion mutant. The deleted portion contains many genes and each plasmid contains a unique region of the deleted portion, allowing the role and importance of the gene in relation to virulence easier to observe. This can be done by transforming the mutant with one of the plasmids and observing to see if the mutant phenotype is restored to that observed in a nonmutated strain. Such phenotypic analyses will likely include examination of virulence in infection studies. Because the deleted region contains several genes, utilizing plasmids that contain specific portions of the deleted region will be able to narrow down the cluster into specific genes that play a role associated with each phenotype.

As described above, efforts to create the pNL1 plasmid for recombination are still in progress. Remaining steps are the insertion of the *Rv2059* and *cobK* regions into pNL1. If use of pNL1 does not lead to recombinants containing the *Rv2059-cobK* region, then a pMB351 derivative will be created that expresses the *recA* gene from a strong promoter in BCG. Because plasmid pMB351 also encodes resistance to kanamycin, the kanamycin resistance gene in pNL1 will need to be removed to enable both plasmids to be transformed into BCG at the same time. This should increase the recombination frequency and increase the number of plasmids in the pool extracted from BCG, which should increase the frequency of *E. coli* transformants that have obtained these sacB-deleted plasmids.

The plasmids generated in this project will be useful in defining the gene functions within the cobalamin biosynthesis region. The incorrect published plasmid map for pMB351 may explain why this plasmid was not used in any other publication since this plasmid was described. Now that we have determined the complete sequence for the plasmid, our laboratory and others will now be able to confidently use this plasmid to express genes in mycobacteria.

Table 1: PCR Primers Utilized

Primer Name	***Primer Sequence (5' → 3')***
P1129	GTCAAGCTTCGGTGAAGACCTATCGGTC
P1134	GTGTCTAGACCGATCCCGTGTCGCTCG
P1135	GTGTCTAGATCGGGTGCGGGTACAAC
P1136	GTGAAGCTTGACGTGGACGGATGTGATCG
P1166	TATAAGCTTGGTGAGCGTGCAGGTTTCG
P1167	CCAGCAGAGGTTCGACGGTGCA
P1205	CACGTTAAGGGATTTTGGTCATGAG
P1206	CCCTTCTAAACGCCCATCG
P1211	CTGACCGTCAAGTCAAGATCC
P1219	CAGGAGCAAGGTGAGATGAC
HindIII-92downsigC	CGCAAGCTTCGGTGGTCATGATAGC

Table 2: Plasmids and Relevant Features

Plasmid Name	***Important/Revelant Features***
pAR7	oriE, oriM, P_{hsp60}*sacB*, P_{L5}*gfp mut-2*, *hyg*, *kan* (vector template for pNL1)
pAR7P_{lacZ}	Plasmid created from pAR7 by insertion of P_{lacZ} promoter upstream of *sacB*
pMB351	oriE, LR7 ori, *kan*
pMV306	oriE, *aph*, *int*,
pNL1	Plasmid to be created from pAR7 by insertion of P_{lacZ} promoter upstream of *sacB*, and 500-bp regions of homology to *Rv2059* and *cobK* of
pUC19spf	Source of P_{lacZ} fragment for pNL1

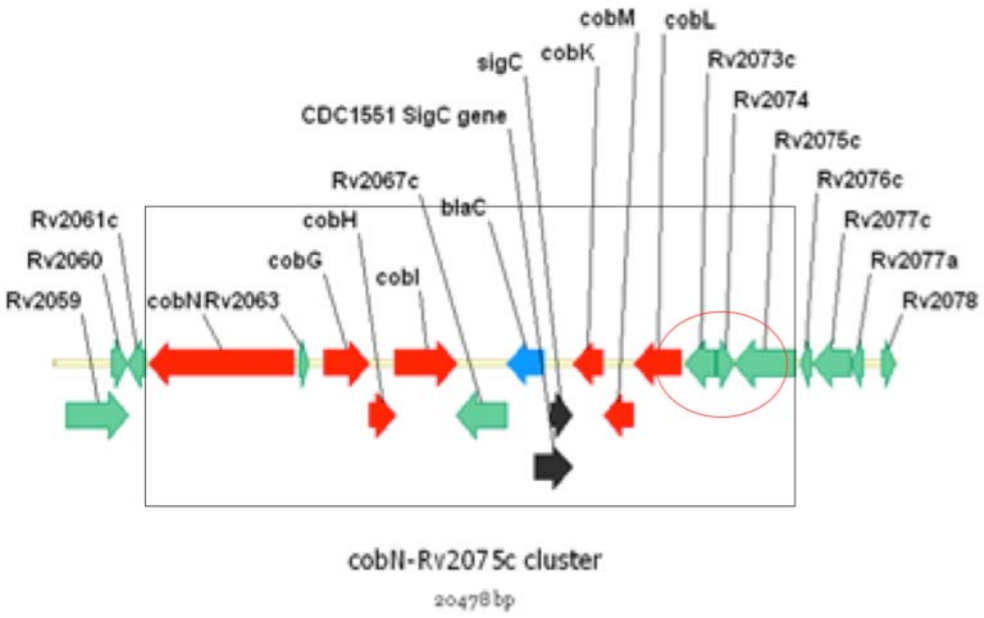


Figure 1: Cobalamin Biosynthesis Region of *M. tuberculosis*: The cobalamin deletion mutation removed genes *cobN* through *Rv2075c* (boxed region). The arrows indicate open reading frames (ORFs) and their orientation. ORFs predicted to encode cobalamin biosynthesis enzymes are shown (red arrows). The *sigC* gene with two possible translational start sites are shown (black arrows). The beta-lactamase gene (*blaC*) is shown in blue. Genes of unknown function are shown in green. The RD9 region present in *M. tuberculosis*, but absent in *M. bovis* BCG, is circled in red.

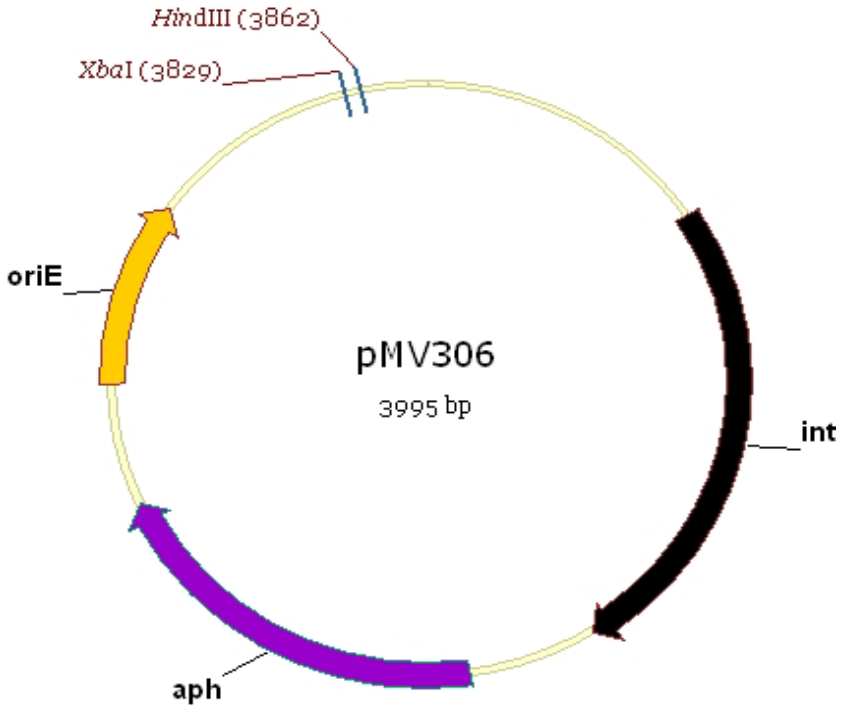


Figure 2: Plasmid pMV306: The basic vector used for complementation in *M. tuberculosis* is plasmid pMV306. This vector contains an *int* region which encodes the mycobacteriophage L5 integrase and *attP* site. The integrase enzyme catalyzes the insertion of the plasmid into the *M. tuberculosis* chromosome at the *attB* site. The vector pMV306 also encodes the amino-phospho-hydrolase gene (*aph*) which confers resistance to kanamycin. The presence of an origin of replication for *E. coli* (oriE) allows the plasmid to replicate in *E. coli*. Unique restriction sites used in this project for insertion of genes in pMV306 are shown.

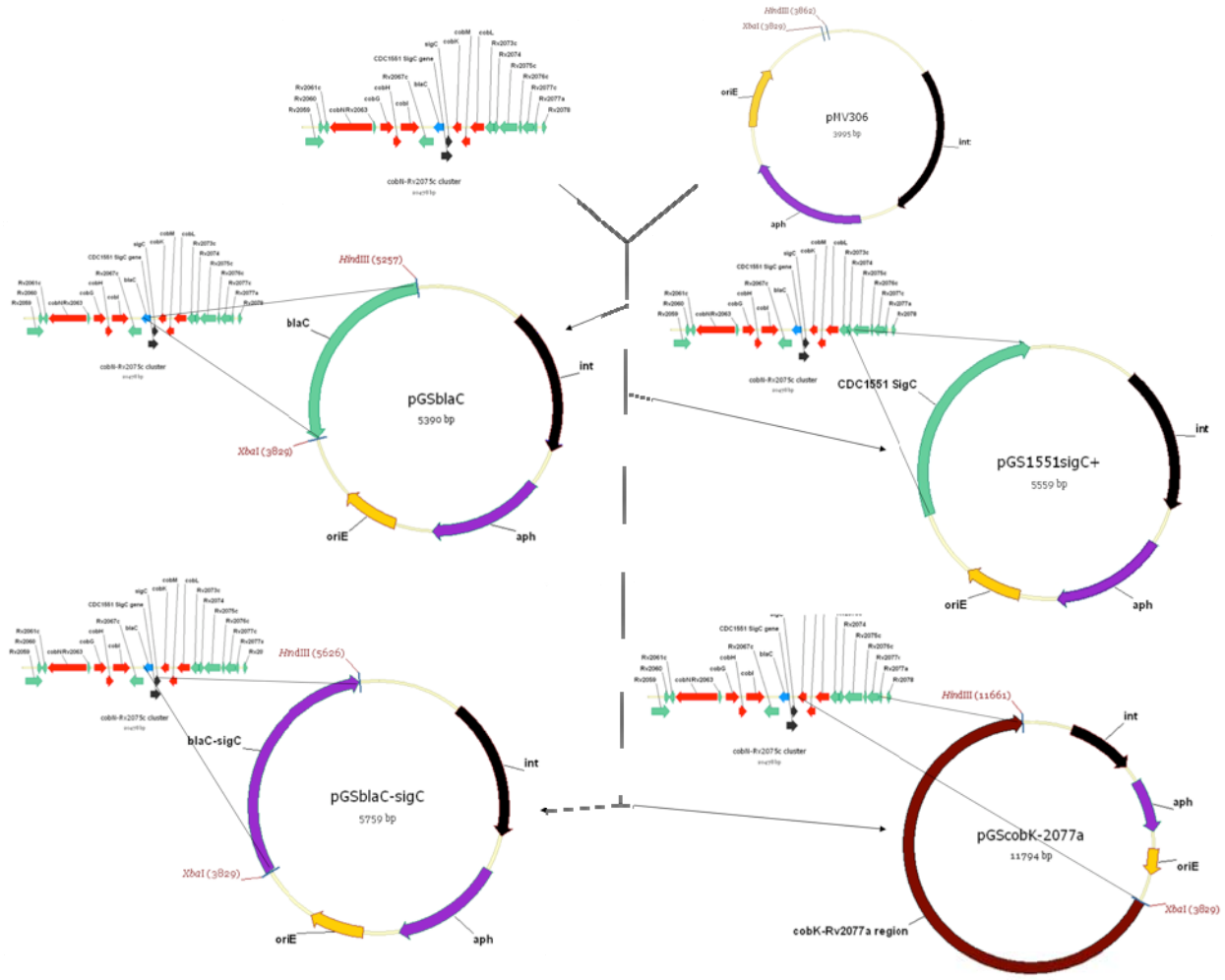


Figure 3: Construction of complementation plasmids: Plasmids generated contain different portions of the cobalamin biosynthetic cluster:

1. pGSsigC contains the ~560-bp *sigC* gene and the 400-bp region upstream of *sigC.*
2. pGSblaC contains the 1.4-kb *blaC* gene and the 400-bp region upstream of *blaC*. The *blaC* fragment was digested with *Xba*I and *Hind*III and ligated into vector pMV306 digested with the same enzymes.
3. pGScobK-2077a contains the ~6.8 kb region between the *cobK* gene and the *Rv2077a* gene. This fragment was digested with *Xba*I and *Hind*III and ligated into vector pMV306 digested with the same enzymes.
4. pGSblaC-sigC contains a ~1.8 kb region which includes the *blaC* and the *sigC* gene. This fragment was digested with *Xba*I and *Hind*III and ligated into vector pMV306 digested with the same enzymes.

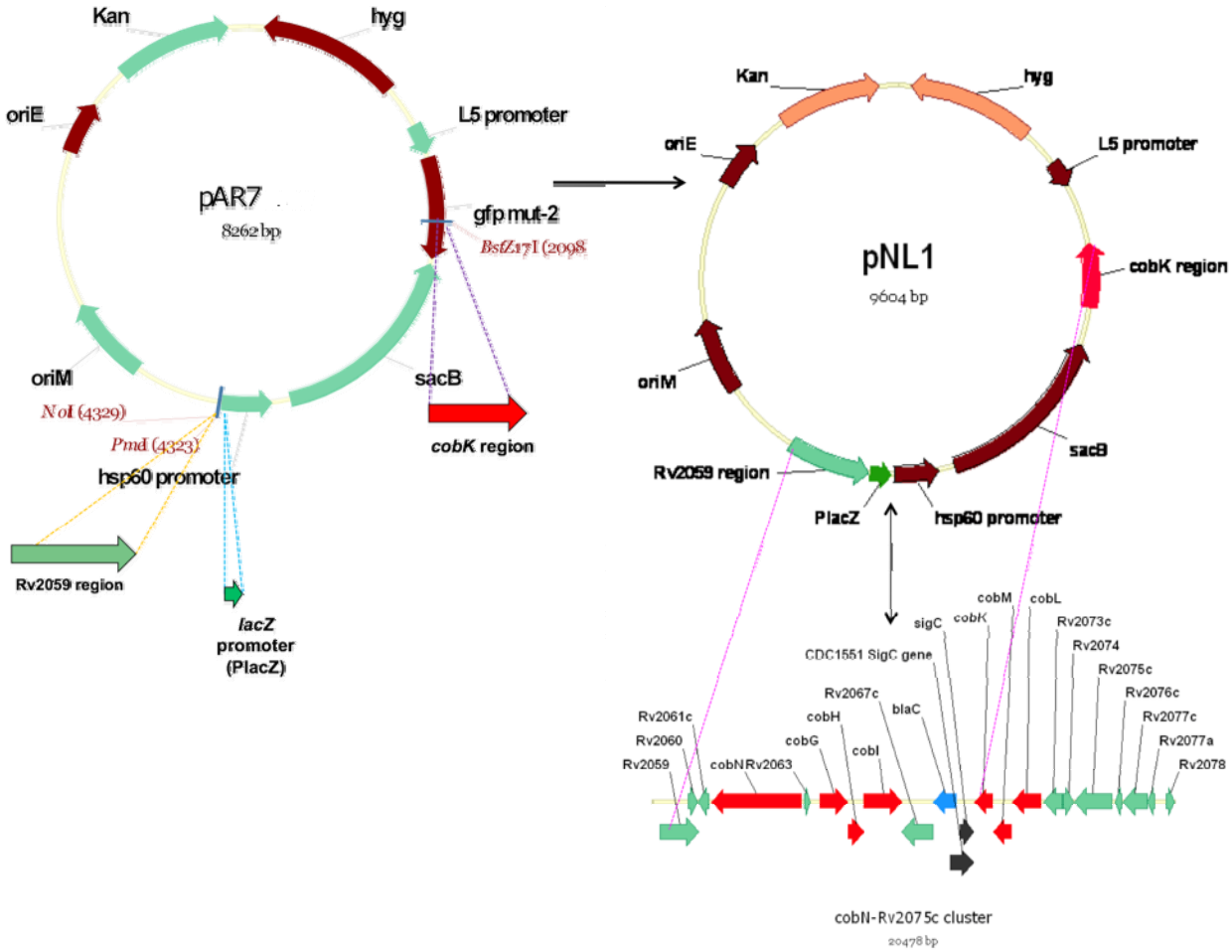


Figure 4: Recovery of the *Rv2059-cobK* region from *M. tuberculosis* by homologous recombination: Plasmid pAR7 contains the following features: an origin for replicating in mycobacteria (*oriM*), an origin for replication in *E. coli (oriE), kan* and *hyg* genes that confer resistance to antibiotics kanamycin and hygromycin, respectively; the *gfp mut-2* gene that encodes for green florescent protein under control of the mycobacteriophage L5 promoter, and the mycobacterial Hsp60 promoter controlling expression of *sacB*. Efforts are in progress to create pNL1. First, the *E. coli* promoter (P_{lacZ}) will be inserted upstream of the *sacB* gene. Second, ~500-bp regions of *Rv2059c* and *cobK* will be inserted such that they flank the *sacB* gene. The *sacB* gene encodes enzyme levan sucrase, which converts sucrose to a toxic product.

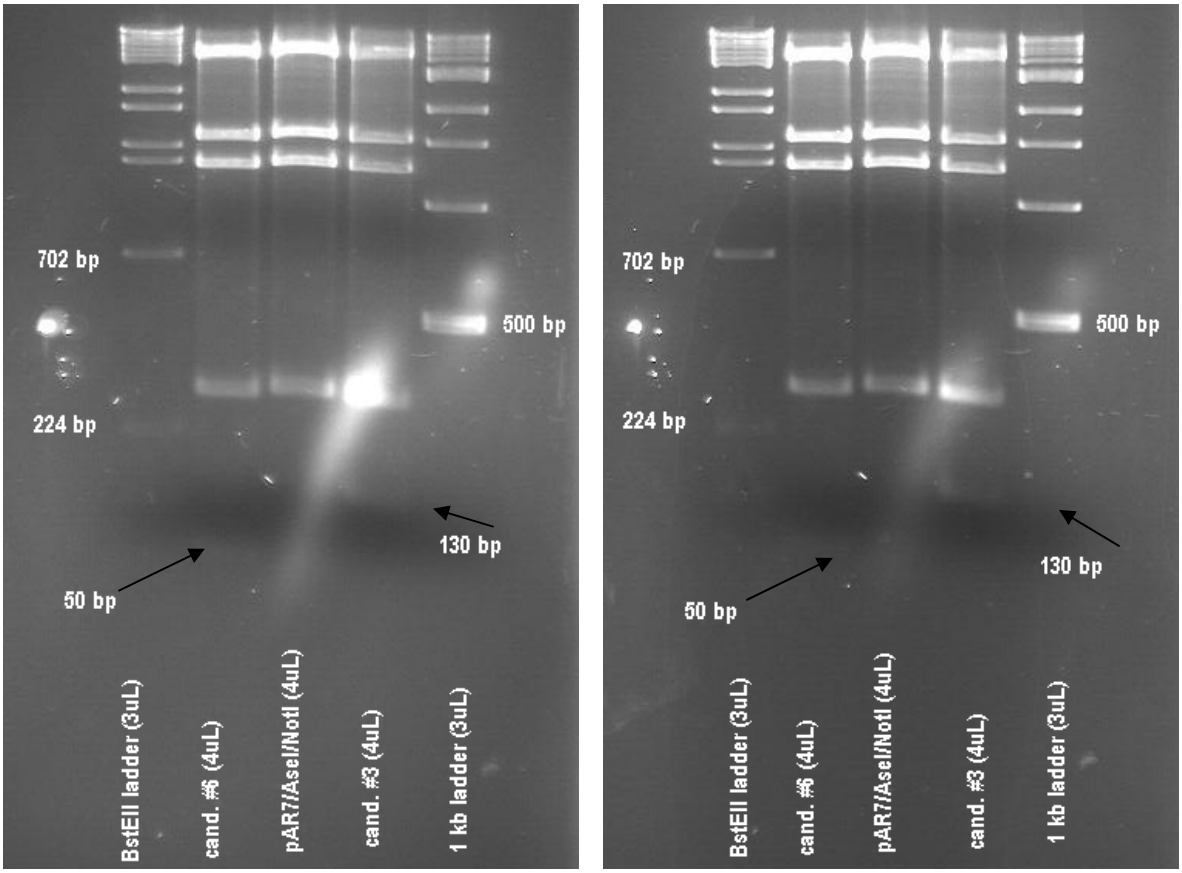


Figure 5: Screening transformants for insertion of P_{lacZ} into the *Pme*I site of pAR7. Plasmid pAR7 and plasmid DNA isolated from transformants were digested with *Ase*I and *Not*I and resolved on a 3% agarose gel. The DNA loaded onto the gel is indicated beneath each lane. The sizes of the lower molecular weight DNA standards are indicated. The expected bands for the pAR7 are 5099 bp, 1597 bp, and 1566 bp. The third band that correlates to approximately 300 bp in all three lanes may be due to a secondary restriction site that was not noted in the pAR7 sequence file. The expected banding pattern for a plasmid receiving the *lacZ* promoter oriented in the same direction as *sacB* in pAR7 is 5223 bp, 1597 bp, 1566 bp, and 50 bp, with the 50 bp band being diagnostic. Digestion of candidate #6 produced the 50-bp band. Digestion of a pAR7 having the *lacZ* promoter in the opposite orientation should produce a 130-bp band, which is evident in candidate #3.

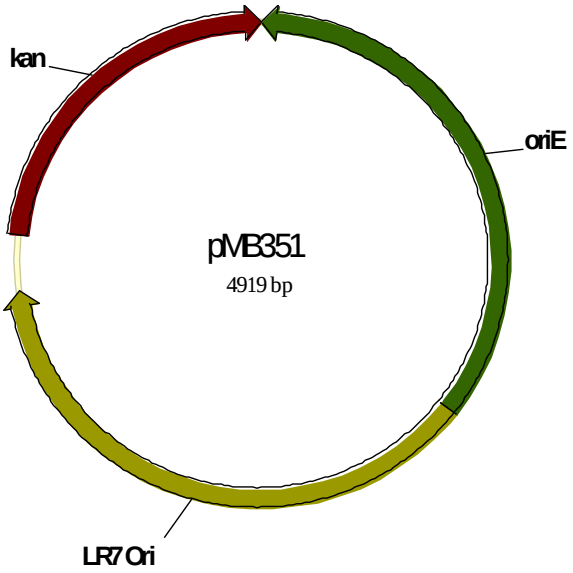
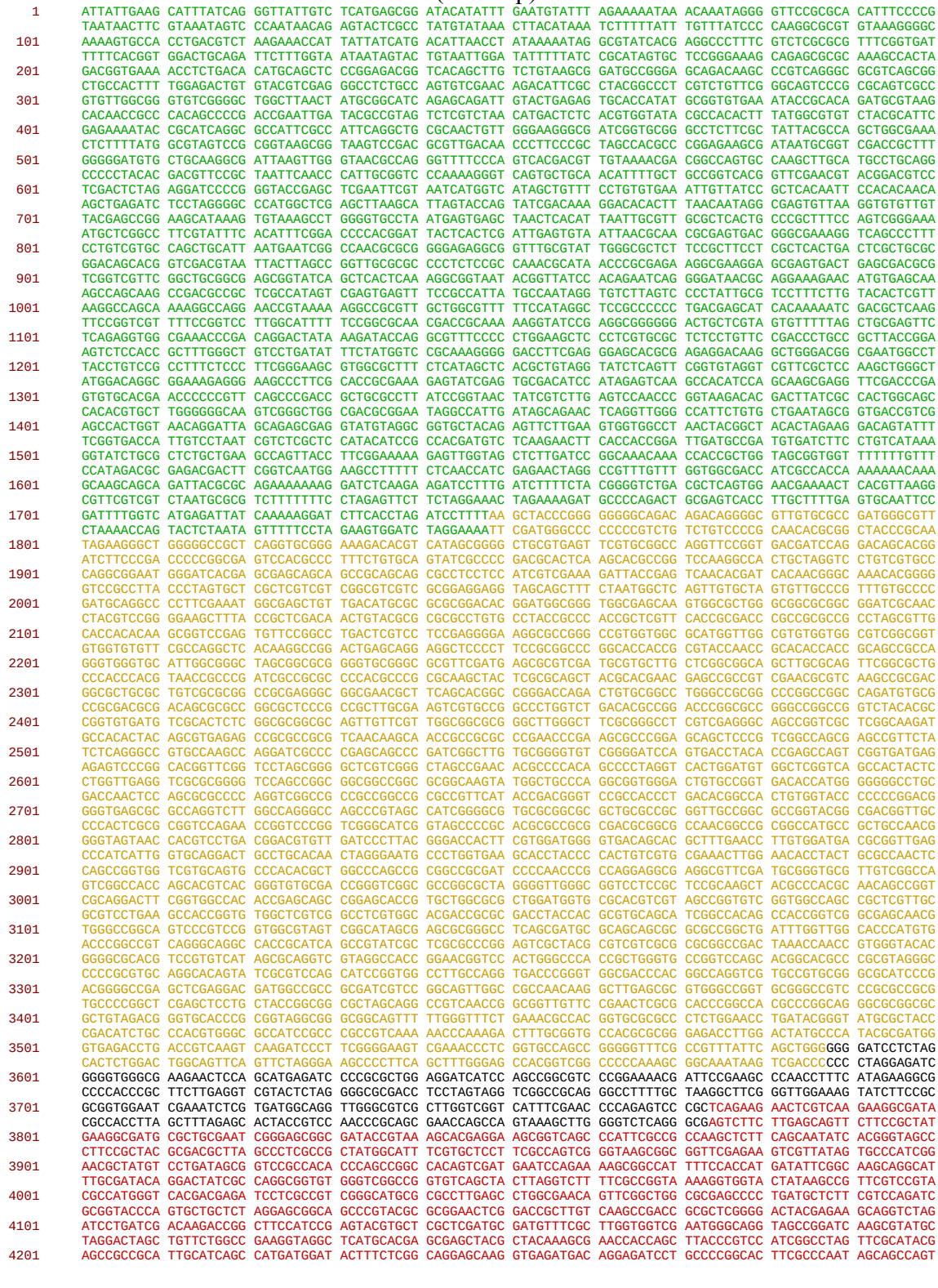


Figure 6: Plasmid drawing of pMB351 resulting from assembling DNA sequencing reads: Shown is plasmid pMB351 with the correct orientation and location of the *kan* gene. The plasmid contains an origin of replication for *E.coli* (oriE), an origin of replication for *M. tuberculosis* (LR7Ori), and a gene which confers resistance to kanamycin (*kan*).

pMB351
(4919 bp)



```
        TCGGCGGCGT AACGTAGTCG GTACTACCTA TGAAAGAGCC GTCCTCGTTC CACTCTACTG TCCTCTAGGA CGGGGCCGTG AAGCGGGTTA TCGTCGGTCA
4301    CCCTTCCCGC TTCAGTGACA ACGTCGAGCA CAGCTGCGCA AGGAACGCCC GTCGTGGCCA GCCACGATAG CCGCGCTGCC TCGTCCTGCA GTTCATTCAG
        GGGAAGGGCG AAGTCACTGT TGCAGCTCGT GTCGACGCGT TCCTTGCGGG CAGCACCGGT CGGTGCTATC GGCGCGACGG AGCAGGACGT CAAGTAAGTC
4401    GGCACCGGAC AGGTCGGTCT TGACAAAAAG AACCGGGCGC CCCTGCGCTG ACAGCCGGAA CACGGCGGCA TCAGAGCAGC CGATTGTCTG TTGTGCCCAG
        CCGTGGCCTG TCCAGCCAGA ACTGTTTTTC TTGGCCCGCG GGGACGCGAC TGTCGGCCTT GTGCCGCCGT AGTCTCGTCG GCTAACAGAC AACACGGGTC
4501    TCATAGCCGA ATAGCCTCTC CACCCAAGCG GCCGGAGAAC CTGCGTGCAA TCCATCTTGT TCAATCATGC GAAACGATCC TCATCCTGTC TCTTGATCAG
        AGTATCGGCT TATCGGAGAG GTGGGTTCGC CGGCCTCTTG GACGCACGTT AGGTAGAACA AGTTAGTACG CTTTGCTAGG AGTAGGACAG AGAACTAGTC
4601    ATCTTGATCC CCTGCGCCAT CAGATCCTTG GCGGCAAGAA AGCCATCCAG TTTACTTTGC AGGGCTTCCC AACCTTACCA GAGGGCGCCC CAGCTGGCAA
        TAGAACTAGG GGACGCGGTA GTCTAGGAAC CGCCGTTCTT TCGGTAGGTC AAATGAAACG TCCCGAAGGG TTGGAATGGT CTCCCGCGGG GTCGACCGTT
4701    TTCCGGTTCG CTTGCTGTCC ATAAAACCGC CCAGTCTAGC TATCGCCATG TAAGCCCACT GCAAGCTACC TGCTTTCTCT TTGCGCTTGC GTTTTCCCTT
        AAGGCCAAGC GAACGACAGG TATTTTGGCG GGTCAGATCG ATAGCGGTAC ATTCGGGTGA CGTTCGATGG ACGAAAGAGA AACGCGAACG CAAAAGGGAA
4801    GTCCAGATAG CCCAGTAGCT GACATTCATC CGGGGTCAGC ACCGTTTCTG CGGACTGGCT TTCTACGTGT TCCGCTTCCT TTAGCAGCCC TTGCGCCCTG
        CAGGTCTATC GGGTCATCGA CTGTAAGTAG GCCCCAGTCG TGGCAAAGAC GCCTGACCGA AAGATGCACA AGGCGAAGGA AATCGTCGGG AACGCGGGAC
4901    AGTGCTTGCG GCAGCGTGA
        TCACGAACGC CGTCGCACT
```

Figure 7: Complete sequence of pMB351: Below is the complete sequence of pMB351. The full sequence was determined using primers p1205, p1206, p1211 and p1219 (see Table 1 for primer sequences). The three features of pMB351 (see Figure 6) are color-coded as follows: the oriE is labeled in green, the LR7 Ori is labeled in gold, and the *kan* gene is labeled in red.

Works Cited

Agarwal, R., & Malhotra, P. . (2005, April 27). Tuberculous dilated cardiomyophathy: an under-recognized entity?. *BMC Infectious Disease*, *5*(29), Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1090580/?tool=pmcentrez

Beggs, M.L., Crawford, J.T. (1995). Isolation and sequencing of the replication region of mycobacterium avium plasmid pLR7. *Journal of Bacteriology, 177*(17), 4836-4840.

Behr, MA, Warren, SA, & Salamon, H. (1999). Transmission of mycobacterium tuberculosis from patients smear-negative for acid-fast bacilli. *Lancet*, *353*(9151), 444-449.

Comstock, G. (1994). The International tuberculosis campaign: a pioneering venture in mass vaccination and research. *Clinical Infectious Diseases*, *19*(3), 528-540.

Herbert, V. (1999). Potential harmful effets of folate food fortification without added cobalamin. *Harrisonis Principles of Internal Medicine*, (14), Retrieved from http://www.victorherbert.com/cv812.htm

Karls, R. K., & Guarner, J. (2006). Examination of mycobacterium tuberculosis sigma factor mutants using low-dose aerosol infection of guinea pigs suggests a role for sigC in pathogenesis. *Microbiology*, *152*, 1591-1600.

Kaufmann, S. (2002). Protection against tuberculosis: cytokines, t cells, and macrophages. *Annals of the Rheumatic Disease*, *61*(2), ii54-ii58.

Kong, D., & Kunimoto, K.Y. (1995). Secretion of human interleukin 2 by recombinant mycobacterium bovis bcg. *Infection and Immunity*, *63*(3), Retrieved from http://iai.asm.org/cgi/content/abstract/63/3/799

Mayo Clinic staff, Initials. (2009, January 28). *Tuberculosis: causes*. Retrieved from http://www.mayoclinic.com/health/tuberculosis/DS00372/DSECTION=causes

Narasimh Acharya, P.V., & Goldman, D.S. . (1970). Chemical composition of the cell wall of the h37ra strain of mycobacterium tuberculosis. *Journal of Bacteriology*, *102*(3), 733-739.

Oplinger, A.A. (2004, January 26). *First u.s. tuberculosis vaccine trial in 60 years begins*. Retrieved from http://www.nih.gov/news/pr/jan2004/niaid-26.htm

Park, A. (2008, October 02). Tuberculosis: an ancient disease continues to thrive. *Time*, Retrieved from http://www.time.com/time/health/article/0,8599,1846698,00.html

Sun, R., Converse, P.J. (2003). Mycobacterium tuberculosis ECF sigma factor sigC is required for lethality in mice and for the conditional expression of a defined gene set. *Molecular microbiology, 52*(1), 25-38.

van Soolingen, D. (1997). A novel pathogenic taxon of the Mycobacterium tuberculosis complex, canetti: characterization of an exceptional isolate from africa. *International Journal of Systematic Bacteriology*, *47*(4), 1236-1245.