CONSTRUCTION AND USE OF A TRANSPOSON FOR IDENTIFICATION OF ESSENTIAL GENES IN MYCOBACTERIA

By

Sarah Danielle Riggs

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE
in
BIOLOGICAL SCIENCES

Zhaomin Yang, Chair

Joseph O. Falkinham, III

David L. Popham

April, 2011

Blacksburg, Virginia

Keywords: *Mycobacterium smegmatis*, Tuberculosis, essential genes, transposon, genetics, TetR

Construction and Use of a Transposon for Identification of Essential Genes in Mycobacteria

By Sarah D. Riggs

ABSTRACT

The continuing emergence of multi-drug resistant *Mycobacterium tuberculosis* is threatening the ability to treat tuberculosis (TB) worldwide. The development of new anti-TB drugs requires new approaches and new drug targets. In this study, a *mariner*-based transposon, *TnQuoVadis*, was constructed to identify essential genes as potential drug targets. This transposon has an outward-facing anhydrotetracycline (ATc)-inducible promoter at each end. A mutant with *TnQuoVadis* inserted upstream of an essential gene may display normal growth in the presence of ATc, but exhibit no growth or severely diminished growth in the absence of ATc. *TnQuoVadis* was placed onto a vector with a temperature sensitive replication origin for more efficient mutagenesis of mycobacteria. In a preliminary genetic screen using the model organism *Mycobacterium smegmatis*, 13 mutants with ATc-dependent growth were identified. Identification of the insertion sites by cloning and sequencing indicated that there were nine genetic loci containing transposon insertions upstream of essential gene candidates in *M. smegmatis*. Further analysis of these genes indicated that many were previously known essential in both *M. smegmatis* and *M. tuberculosis.* These results demonstrate that *TnQuoVadis* and its delivery system can be utilized for the identification of essential genes in mycobacteria.

ACKNOWLEDGEMENTS

I would like to especially thank Dr. Zhaomin Yang for giving me the opportunity to work under his guidance, encouraging me to aim higher and inspiring me to accept nothing less than excellence. I would also like to thank Dr. Joseph Falkinham, III and Dr. David Popham for their time and valuable suggestions in my research.

I would like to thank all my colleagues in Dr. Yang's lab for making it a place filled with fun, enthusiasm and adventures. Special thanks to Dr. Wesley Black who has always been there to nudge me in the right direction. I would like to acknowledge the following people for their friendship: Regina Wallace, Chengyun Li, and Andrew Schneider, also to the administrative staff in the Department of Biological Sciences. I am thankful to Sue Rasmussen for her help and to Catherine Webb and Katherine Rodgers for a rewarding teaching experience.

I would like to thank the following undergraduates that worked very hard assisting me on my project: Cordelia Weiss and Chelsea Calloway. I am indebted to Dr. Manli Yang Davis and Lee Davis for their hospitality and valuable friendship. As always, my main support system has been my loyal family who I appreciate and love dearly. Mom, you're the best! Most importantly I am thankful for the spiritual strength afforded to me by my Lord and Savior Jesus Christ.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1:

Chapter 2:

LIST OF TABLES

Chapter 2:

CHAPTER 1

Literature Review

General introduction to the mycobacteria

Mycobacteria are rod-shaped and non-motile bacteria that belong to the phylum *Actinobacteria* which include bacteria with high G/C content (Ventura *et al.,* 2007). Due to the complexity of the mycobacterial cell wall that is rich in waxy lipids, mycobacteria cannot be effectively Gram-stained (White, 2007). Instead, all mycobacteria are stained by a technique called acid-fast staining. This is in part due to lengthy chains of mycolic acids, between 60 and 90 carbons long (Gangadharam & Jenkins, 1998), that results in a formidable protective envelope. Mycolic acids are a class of lipids that are known to form thick hydrophobic layers due to a complex mixture of branched, long-chain fatty acids (White, 2007; Alahari *et al*., 2007). Only a handful of other bacteria are known to have mycolic acids contained in their cell walls; these bacteria form a monophyletic taxon called the CMN group named after its members: *Corynebacterium*, *Mycobacterium*, and *Nocardia* (Ventura *et al.*, 2007). Of this select group, mycobacteria have the longest mycolic acid chains making this property unique to the *Mycobacterium* genus (Gangadharam & Jenkins, 1998).

Opportunistic nature of the mycobacteria

Members of the mycobacteria inhabit diverse habitats, many of which live in soil (Wolinsky & Rynearson, 1968), or aquatic environments (Collins *et al*., 1984). Many mycobacteria are typically free-living and feed on dead organic matter (Falkinham *et* al., 1980), however, members of the genus are considered to be opportunistic pathogens. Non-tuberculosis infection may occur from the environment possibly resulting in skin lesions or bacterial colonization in bodily secretions (Bicmen *et al*., 2010). There are more than 120 mycobacteria species known to cause a broad spectrum of disease in humans (Tortoli, 2006). Medically

important species include *M. leprae,* the causative agent of leprosy, and *M. tuberculosis*, the causative agent of tuberculosis. These well known human pathogens are members of the slowly growing mycobacteria (SGM) group which require more than 7 days to produce visible colonies on solid media, sometimes requiring weeks (Pfyffer *et al*., 2003).

The tuberculosis threat

Tuberculosis (TB), a pandemic disease, kills approximately two million people annually and infects one-third of the world population (WHO Report, 2010). Its causative agent is the notorious mycobacteria species *Mycobacterium tuberculosis.* The disease is highly contagious and especially dangerous to the immune compromised. Due to the lack of an effective vaccine for adult pulmonary tuberculosis (Fine *et al*., 1999), antibiotic administration is the main strategy to combat TB (Guo *et al*., 2007). The treatment of patients requires a continuous antibiotic administration for 6-18 months (Frieden *et al*., 2006) which is further complicated by the emergence of multidrug-resistant TB strains. There are two classes of drug resistant *M. tuberculosis*: multi drug resistant (MDR) strains which are resistant to first-line drug treatments, and extensively drug resistant (XDR) strains that are resistant to at least three second-line drug treatments (Alahari *et al*., 2007). Multi-drug resistant TB strains poses a serious threat to human health worldwide which compromises health security, challenges the control of infectious diseases, and jeopardizes health care that significantly increases costs (WHO Report, 2010), thus there is an urgent need for the development of new anti-TB drugs.

Hindrances to TB research

Part of the difficulty of developing new anti-TB drugs is due to the infectious nature of *M. tuberculosis* and its lengthy generation time. A laboratory interested in working with *M. tuberculosis* must have access to biosafety level 3 containment as defined by the Centers for Disease Control and Prevention (CDC). The organism belongs to the SGM group with doubling times around 24 hours that yields colonies between three and four weeks after incubation (Hatfull & Jacobs, 2000). An additional challenge to TB research is the lack of tools for efficient genetic manipulations in mycobacteria. This has become an obstacle for the identification of essential genes in TB. For these reasons, working directly with *M. tuberculosis* tends to be problematic. Instead, much advancement has been made by using related mycobacteria as a model species and correlating discoveries to *M. tuberculosis.* One such organism that tends to be attractive to work with is *M. smegmatis.*

Model organism *M. smegmatis*

To overcome some hindrances associated with TB research, strategies may be developed using the model organism *Mycobacterium smegmatis.* *M. smegmatis* is a favored organism due to being a member of the rapidly growing mycobacteria (RGM) group. RGM species require less than 7 days to produce visible colonies on solid media (Ade´kambi and Drancourt, 2004). The average doubling time of *M. smegmatis* is between three to four hours with visible colonies forming within two to four days, thus *M. smegmatis* can be grown easily and quickly as compared to *M. tuberculosis*. Another benefit includes the need of only a biosafety level 2 containment as defined by the CDC. Due to the approachability of *M. smegmatis* research, many labs have been able to contribute important scientific finds to the community.

M. smegmatis strain mc²155

The fast growing *Mycobacterium smegmatis* species has been used as a model organism for mycobacteria studies. However, transforming the species proved to be extremely difficult and only recently has *M. smegmatis* been efficiently transformed with plasmid vectors due in part to the creation of highly transformation efficient strains derived from the *M. smegmatis* strain $mc^2 6$ (Snapper *et al*., 1990). One such strain named mc²155 is a mutant with the ability to be transformed with efficiencies that are four to five orders of magnitude greater than the wild type strain $mc^2 6$ (Snapper *et al*., 1990). After its discovery, mc²155 has helped advance the understanding of mycobacteria by making development of genetic tools for regulated gene expression more accessible.

Regulated gene expression in mycobacteria

The first mycobacteria genetic tool developed for use in regulating gene expression was an inducible promoter that uses the regulatory elements adapted from the acetamidase operon (Ehrt & Schnappinger, 2006; Wei & Rubin, 2008). Acetamidase allows *M. smegmatis* to utilize amides as the sole carbon source. Expression of acetamidase is induced by acetamide (Ehrt & Schnappinger, 2006) and is regulated by three proteins and four promoters (Roberts *et al.*, 2003) with two positive regulators and one negative regulator (Ehrt & Schnappinger, 2006). Unfortunately, controlling gene expression using the acetamidase operon promoter is difficult due to its complexity and the regulatory mechanism of the system has yet to be completely understood (Ehrt & Schnappinger, 2006). Despite these shortcomings, it has been successfully used in a number of mycobacteria research studies and applications such as over-expression of

heterologous proteins, and creation of gene expression vectors and regulation experiments (Ehrt & Schnappinger, 2006; Parish *et al.,* 2001; Roberts *et al.,* 2003; Triccas *et al*., 1998). Recently, a simpler mycobacteria gene expression system was developed from a tetracycline export system (Fig. 1-1).

The tetracycline (Tet) export system (Fig. 1-1) is controlled by two operator sites (*tetO*), a repressor (*tetR*) and a gene encoding a tetracycline efflux pump (*tetA*) (Ehrt & Schnappinger, 2006). Tetracycline is capable of inhibiting growth and can freely diffuse across cell membranes (Argast & Beck, 1984). In order to remove the presence of tetracycline, the cell utilizes the Tet system to export tetracycline out of the cytoplasm (Ehrt & Schnappinger, 2006). The two *tetO* operator sites overlap two promoters that drive the expression of the *tetR* and *tetA* genes. These operator sites are the target of TetR which binds and prevents transcription of both the *tetR* and *tetA* genes (Carroll *et al.,* 2005). When tetracycline is present, it binds to TetR which causes a conformational change leading to the repressor disassociating from *tetO* thus initializing transcription of both *tetR* and *tetA* (Carroll *et al.,* 2005) (Fig. 1-1). The TetA protein is inserted into the cell membrane where it acts as a channel to export tetracycline from the cell. Tet controlled systems have been modified in many ways and constructed into plasmids to allow for analysis of essential genes in mycobacteria, most notably *M. smegmatis* and *M. tuberculosis* (Ehrt & Schnappinger, 2006) in which the *tetA* gene was removed so that genes of interest could be inserted.

There are a total of three separate tetracycline (Tet) controlled systems (Fig. 1-2) that have been developed for mycobacteria that differ in repressors and promoters (Ehrt & Schnappinger, 2006). Tetracycline is an antibiotic that is known to inhibit growth in

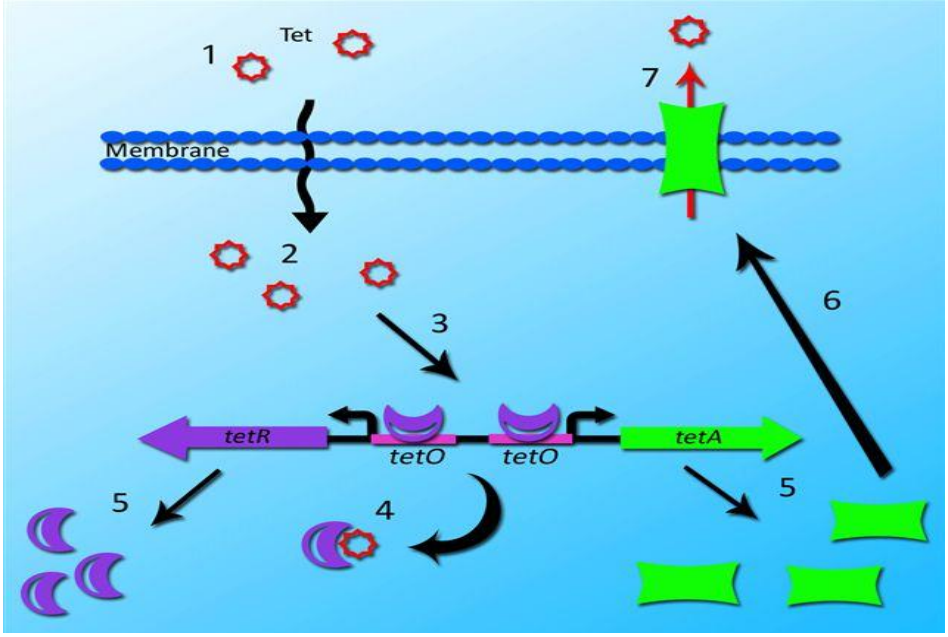

Figure 1-1. Tetracycline export system. Tetracycline present in the environment (1) diffuses across the membrane (2), where it interacts and binds to the TetR repressor protein (3), causing a conformational change in which TetR displaces from the operator (4). This frees up the promoter regions for both *tetR* and *tetA* allowing transcription which leads to the production of TetR and TetA protein (5). The TetA proteins are embedded into the membrane (6), which are active antiport proteins that export tetracycline by passing H+ ions in via the proton motive force (not shown). Once tetracycline has been fully transported from the inside of the cell, TetR will bind unhindered to the operators and restrict transcription.

M. smegmatis but all three Tet systems developed can be induced by very low concentrations of tetracycline that will not inhibit mycobacteria growth. The three systems are as followed: the TetR(Z)-derived system, and the two TetR(B)-derived systems (Ehrt & Schnappinger, 2006). TetR(Z) was isolated from *Corynebacterium glutamicum*, but the system is active in mycobacteria (Ehrt & Schnappinger, 2006) (Fig. 1-2). The two TetR(B) systems are both *Tn10*-derived. They differ in the number of *tetO* operators used, with one system using one operator and the other using two (Ehrt & Schnappinger, 2006). They also differ in the organism the promoter was taken from. The TetR(B)1 promoter controls expression by a modified *Bacillus subtillis* xylose promoter and the TetR(B)2 system uses a mycobacterial promoter (Ehrt & Schnappinger, 2006). The mycobacterial promoter used in TetR(B)2 is a strong promoter that is a 303bp long DNA fragment taken from *M. smegmatis* (Kaps *et al*., 2001). This promoter was modified by inserting Tet operators upstream and downstream of the -35 region using oligonucleotide-directed PCR mutagenesis (Ehrt *et al.,* 2005). These Tet systems are particularly useful in analyzing essential genes in mycobacteria as the Tet promoter containing the operators can be controlled by tetracycline. Both the acetamidase and the tetracycline-regulated promoters have been used successfully in gene expression and regulation experiments. This is done by replacing the wild-type promoters for genes of particular interest with one of the regulated promoters causing those genes to be transcribed only in the presence of the inducers (Wei & Rubin, 2008). This method is becoming widely used due to its ability to test for essential genes by growing strains both in the presence and absence of the inducers, as well as gene function by depleting the strain of the inducer and looking for phenotypic changes (Wei & Rubin, 2008).

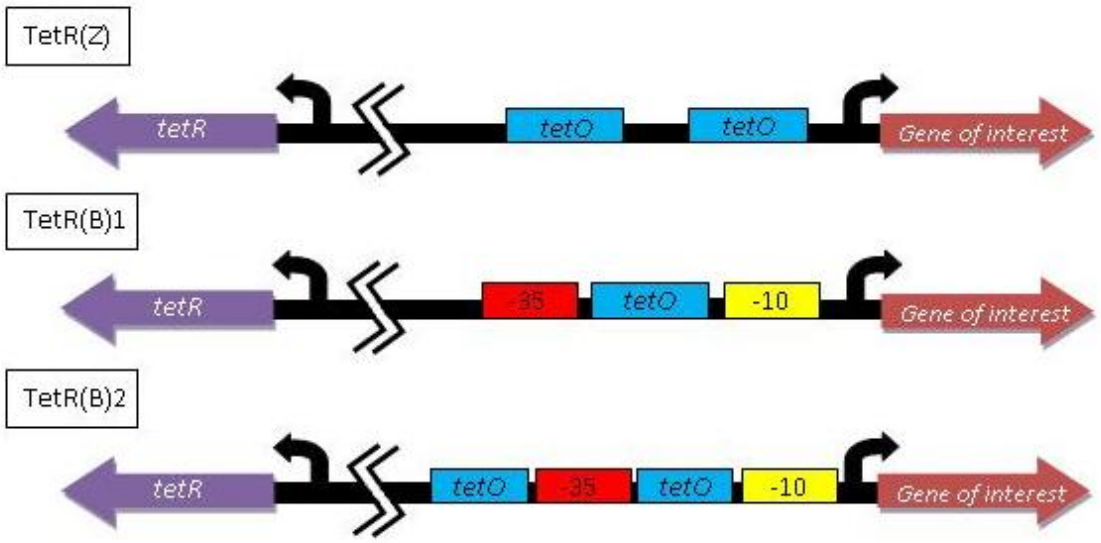

Figure 1-2. Tet-derived mycobacterial expression systems. Depiction of the three Tet controlled expression systems that have been developed for mycobacteria. The TetR(Z) derived system controls expression of the gene of interest from an unidentified promoter containing two *tetO* operators. The TetR(B)1 derived system controls expression by a *Bacillus subtillis* xylose promoter that had *tetO* inserted between the -35 and -10 sites. The TetR(B)2 system controls expression by an adapted *M. smegmatis* promoter that contains two *tetO* regions flanking the -35 site. Modified from reference: (Ehrt & Schnappinger, 2006).

Previous attempts at identifying essential genes in mycobacteria

One of the obstacles in the studies of mycobacteria is the lack of tools for efficient identification of essential genes. The major challenges are growing strains with mutations in essential genes and mapping these mutations once they are available. These difficulties have hindered the study of mycobacterial biology and the progress in combating mycobacterial infections. Methods have been developed in order to address these issues. One such method relies on large-scale random transposon mutagenesis to locate essential genes (Hutchison *et al*., 1999). Transposition causes insertion mutations by introducing large fragments of DNA, thus disrupting the original sequence. It is understood that when a transposon inserts into a gene whose expression is essential for the survival of the cell, that cell will fail to grow or even die, but if a transposon inserts into a non-essential gene, the cell will continue grow. Large numbers of mutants can be analyzed by identifying transposon insertions in genes via DNA sequencing. Sequencing produces a collection of insertion sites representing the non-essential genes and those genes that are not disrupted are potentially the essential genes (Snyder and Champness, 2007). The more mutants that are sequenced, the more accurate the prediction of the essential genes, and the clearer the final picture becomes. Transposon mutagenesis has been used in this manner to detect potential essential genes in mycobacteria (Sassetti *et al*., 2003); however, this method is not exact. This is a time-consuming and costly practice with no guarantee that the genes without transposon insertions are essential, which leads to the need for more experimentation to test the essentiality of the genes.

Another way to identify essential genes is to create temperature-sensitive conditional mutants by introducing point mutations by chemical mutagenesis (Sambrook and Russell, 2001). The protein products containing point mutations may function under normal temperatures but

may not fold correctly with elevated temperatures. Once point mutations are introduced, they are grown in both permissive (30ºC) and non-permissive (37ºC) temperatures to identify those that have the temperature-sensitive mutation. Screening conditional mutants created by chemical mutagenesis has been used to detect potential essential genes in mycobacteria (Liu and Nikaido, 1999); however, similar to transposon mutagenesis, identifying essential genes by creating temperature-sensitive mutants is a time-intensive process. To identify the essential genes, vectors containing fragments of wild-type genomic DNA are introduced into the mutant to attempt to restore the ability to grow in elevated temperatures. Once a DNA fragment is identified that yielded a successful restoration, the mutation is mapped by amplifying the genomic region corresponding to the DNA fragment.

Due to the increasing availability of fully sequenced genomes, comparative genomics has become a favorable method for identifying putative essential genes. In this method, computer-aided genome comparisons identify essential genes that have been highly conserved across multiple species which are compared against well documented model organisms (Mushegian and Koonin, 1996). Although useful, this method is limited by the availability of sequenced genomes.

Research objectives

The goal of this project is to construct a transposon and a temperature-sensitive delivery vector that will allow for the identification of essential genes in the model organism *M. smegmatis* quickly and efficiently. Our method uses aspects from all three approaches described above. 1) We aimed to utilize transposon mutagenesis to randomly insert an engineered transposon across the *M. smegmatis* genome. 2) We used a temperature-sensitive suicide plasmid

to deliver the transposon. 3) We compared the predicted essential regions in *M. smegmatis* with homology to regions in *M. tuberculosis*.

In order to reach our goal, the following specific aims were accomplished:

a. Construction of the transposon and temperature-sensitive delivery vector

b. Testing in *M. smegmatis*

c. Selection of mutations

d. Isolation and identification of regulated sequences

e. Comparing identified ORFs of *M. smegmatis* with *M. tuberculosis*

Rationale of the experimental design

Locating essential genes can prove to be difficult since essential gene disruption results in cell death. By utilizing a gene regulation system, it is possible to control gene expression by dictating when the gene is 'on' or 'off'. The rationale was to construct a transposon with inducible *tetO* promoters at both ends that drive transcription outwards. When the transposon is inserted upstream of an essential gene, the mutant will display inducer-dependent growth. That is, such mutants will grow in the presence of the inducer to allow their isolation and maintenance. In the absence of the inducer, they will not grow which allows their differentiation from strains with mutations in non-essential genes. The identification and mapping of these mutations is relatively straightforward because they are marked by transposon insertions.

The tetracycline regulation system has successfully been used in controlling mycobacteria gene expression (Blokpoel *et al.,* 2005; Ehrt *et al*., 2005; Guo *et al*., 2007) making

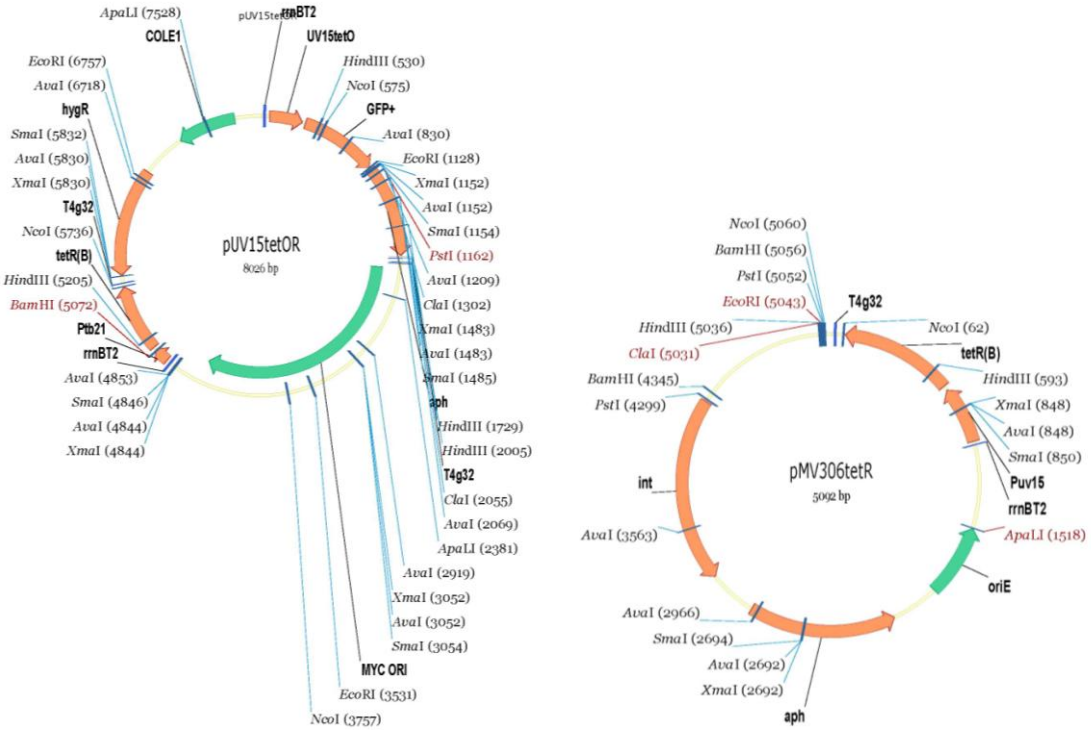

Figure 1-3. Plasmids pUV15*tetO*Rm and pMC1s. Plasmid pUV15*tetO*Rm contains a *M. smegmatis* promoter containing two *tetO* operators from the TetR(B)2 system, labeled here as UV15*tetO*. Plasmid pMC1s (pMV306tetR) is integrative in mycobacteria and contains the *tetR* gene, labeled here as tetR(B). Both plasmids from reference: (Ehrt *et al*., 2005)

this an ideal system for our purposes due to its simplicity and ease of induction. For experimentation, we chose plasmids pUV15*tetO*Rm and pMC1s (Ehrt *et al*., 2005) which carry the *tetR* gene and the Tet operators (Fig. 1-3). Plasmid pUV15*tetO*Rm contains a modified *M. smegmatis* promoter containing two *tetO* operator sites [TetR(B)2] that was used as the inducible promoters in the engineered transposon. The integrative plasmid pMC1s was chosen to insert the *tetR* gene into *M. smegmatis,* creating a strain that constitutively expressed TetR. This TetR protein would be available to bind to the *tetO* operators contained within the transposon. Anhydrotetracycline (ATc), an analogue of tetracycline, was used as the inducer instead of tetracycline due to its relatively reduced toxicity and high affinity for TetR (Ehrt & Schnappinger, 2006). When present, ATc binds to TetR, causing a conformational change which displaces TetR from the *tetO* operators. In our system, if a transposon promoter is regulating an essential gene, the cell will be viable in the presence of ATc but not in the absence of ATc.

For delivery of the transposon into the TetR *M. smegmatis* strain, we utilized a vector with a mycobacterial thermosensitive origin of replication. The most well studied mycobacteria temperature-sensitive origins of replication were derived from the mycobacteria plasmid pAL5000 using hydroxylamine to introduce mutations (Guilhot *et al.,* 1992). The pAL5000 *M. fortuitum* plasmid is a small 4837bp replicon with five open reading frames (Ranes *et al.,* 1990) that has been used extensively in mycobacteria research. We chose to modify the mycobacterial thermosensitive plasmid pMycomar to contain *TnQuoVadis* for efficient delivery in our parental strain.

CHAPTER 2

Construction and Use of a Transposon for Identification of Essential Genes in Mycobacteria

INTRODUCTION

Mycobacterium tuberculosis, the causative agent of tuberculosis (TB), is a globally prominent human pathogen that has been estimated to infect one-third of the world population and causes approximately 4,600 deaths each day (WHO Report, 2010). Control of TB depends almost entirely on antibiotic administration due to the lack of an effective vaccine for the prevention of the disease in adults (Fine *et al*., 1999). The emergence of multidrug-resistant TB strains has severely complicated treatment and poses a serious threat to human health worldwide. With the numbers of new cases of drug-resistant TB increasing at an alarming rate, there is an urgent need for the development of new anti-TB drugs to prevent the further deterioration of health security.

One of the obstacles of mycobacteria research is the lack of tools for efficient genetic manipulations of mycobacteria and for the identification of essential genes. These difficulties have hindered the study of mycobacterial biology and the progress in combating mycobacterial infections. Here, we outline our efforts to combine molecular biology and microbial genetics to develop tools to assist in the systematic identification of essential genes in mycobacteria. Our basic idea was to construct a transposon with inducible *tetO* promoters at both ends that drive transcription outwards (Fig. 2-1). We utilized the Tet-controlled expression system TetR(B)2 (Ehrt *et al*., 2005) due to its simplicity and ease of induction. When the transposon is inserted upstream of an essential gene, the mutant will display inducer-dependent growth. That is, such mutants will grow in the presence of the inducer to allow their isolation and maintenance. In the absence of the inducer, they will not grow which allows their differentiation from mutants in non-essential genes (Fig. 2-1).

We demonstrate here the construction of our transposon, *TnQuoVadis*, and a vector for efficient mutagenesis in mycobacteria. A preliminary proof-of-concept experiment indicated that these new tools are effective for their intended purpose. The hope is that our tool will help to identify the unique set of essential genes in mycobacteria to facilitate the development of new anti-TB drugs and a better understanding of mycobacterial biology.

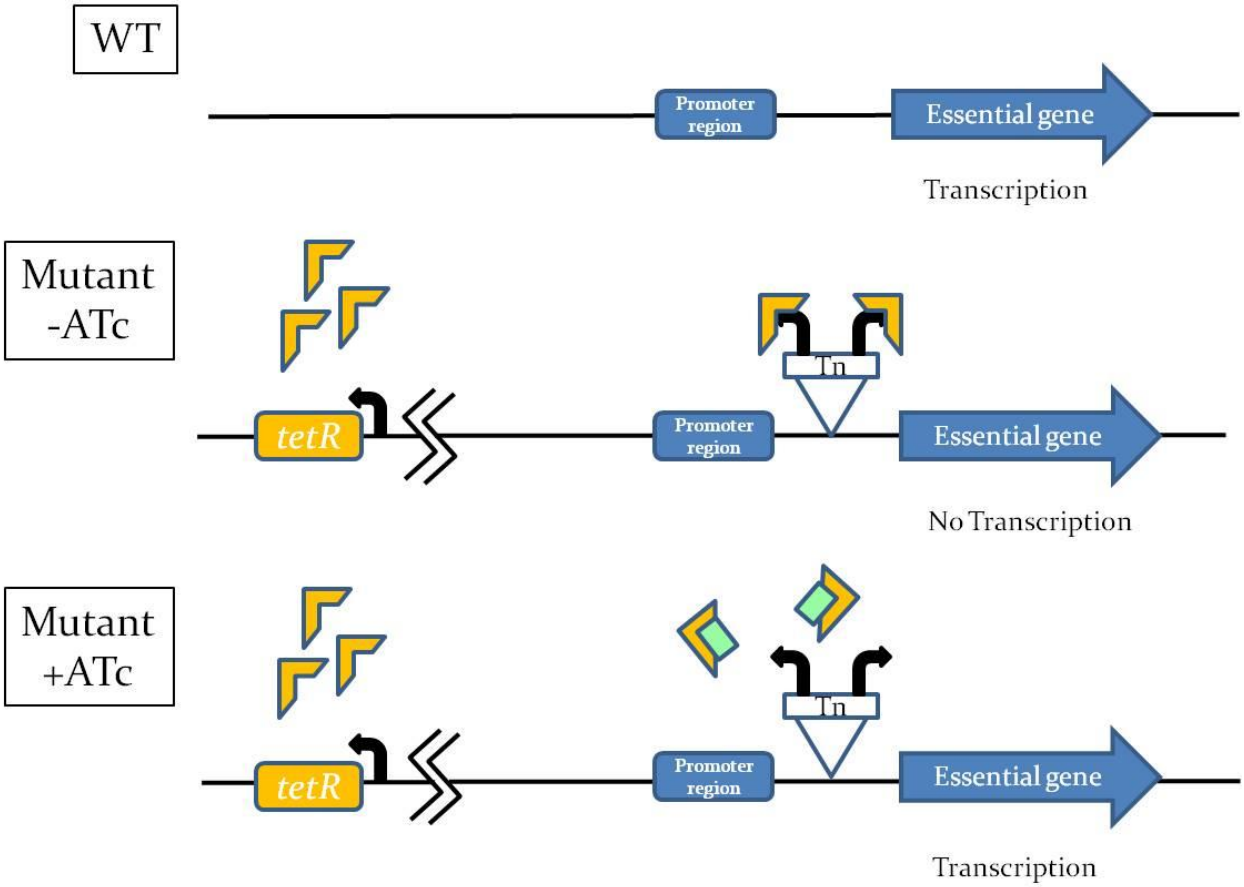

Figure 2-1. *TnQuoVadis* Function. The repressor gene (*tetR*), which is integrated on the chromosome, is constitutively expressed. If the transposon (Tn) inserts upstream of an essential gene, the TetR protein (shown in yellow) binds to the inducible *tetO* promoters within the transposon, thus this gene becomes regulated by the inducer. The inducer anhydrotetracycline (ATc) is shown as a green box.

MATERIALS AND METHODS

Bacterial strains and growth conditions

All strains and plasmids used in this study are listed in Table 2-1. The *Escherichia coli* strains DH5αMCR (GibcoBRL) and DH5*αλpir* (Rubin *et al*., 1999) were grown and maintained at 37ºC on Miller Luria-Bertani (LB) agar plates or in LB liquid media (Miller, 1972). *M. smegmatis* strains were grown and maintained in LB liquid media supplemented with 0.05% Tween 80 or on LB agar at 30ºC, 37ºC or 39 ºC as specified. All plates contained 1.5% agar and all liquid cultures were aerated by shaking at 250 rpm on a rotary shaker.

When required for selection, kanamycin (Km) and hygromycin (Hyg) were used at 100 µg/ml, and tetracycline was used at 15 µg/ml for *Escherichia coli*. Km and Hyg were used at 50 µg/ml and gentamycin (Gm) at 15 µg/ml for selection in *M. smegmatis*. For induction in *M. smegmatis*, an anhydrotetracycline (ATc) concentration of 50 ng/µl was chosen as this concentration was previously found to be ideal in *M. smegmatis* TetR induction that utilized $P_{UV15tetO}$ (Ehrt *et al*., 2005).

Overlapping PCR

All PCR primers, listed in Table 2-2, were designed using PrimerQuest^{SM} software (http://www.idtdna.com/SCITOOLS/Applications/PrimerQuest/ Default.aspx), and purchased through Integrated DNA Technologies. The 25 µl PCR reactions consisted of 5× Phusion® HF buffer (Finnzymes), 200 µM of each dNTP, 0.5 µM of each primer, ~50-200 ng template DNA and 1U of Phusion® DNA polymerase (Finnzymes). PCR products were purified using the QiAquick PCR purification kit.

Table 2-1. *M. smegmatis* strains and plasmids used in this study.

Strains/Plasmids	Genotype	Reference
Strains:		
E. coli		
DH5αMCR	Cloning host	GibcoBRL
DH5*αλpir*	Cloning host, recognizes R6Kγ origin	Rubin *et al*., 1999
M. smegmatis		
mc²155	Electroporation-proficient mutant of $mc^2$6	Snapper *et al*., 1990
VTY123	mc²155 derivative, pSR125 integrated	This study
Plasmids:		
pMycoMar	*magellan4* mutagenesis vector, Kan^R, $Gent^R$	Rubin *et al*., 1999
pMyr6k	pMycoMar parent, Amp^R	Rubin *et al*., 1999
pWB300	Cloning vector, Tet^R	Black, 2005
pMC1s	*tetR* constitutively expressed, integrative at L5 *attB*, Kan^R	Ehrt *et al*., 2005
pSE100	T4 transcription terminator, Hyg^R	Guo *et al*., 2007
pUV15*tetO*Rm	Inducible mycobacterial promoter ($P_{UV15tetO}$), Hyg^R	Ehrt *et al*., 2005
pBlueScript KS	Cloning vector, Amp^R	Stratagene
pMe1ml1	*lacZ* reporter plasmid, Hyg^R	Ehrt *et al*., 2005
pZerO-2	Cloning vector, Kan^R	Invitrogen
pSR114	pWB300 derivative, *XbaI-SpeI magellan4* fragment from pMyr6k	This study
pSR117	pMe1mL1 derivative, Δ*tetR*	This study
pSR118	pSE100 derivative, Kan^R region isolated from pZerO-2	This study
pSR119	pBluescript derivative, Right transposon fragment	This study
pSR120	pBluescript derivative, Left transposon fragment	This study
pSR121	pSR119 derivative, *TnQuoVadis* construct	This study
pSR122	pBluescript derivative, transposase fragment isolated from pMycoMar	This study
pSR123	pSR121 derivative, *TnQuoVadis*/transposase construct	This study
pSR125	pMC1s derivative, ΔKan^R, Hyg^R region isolated from pSE100	This study
pSR126	pSR117 derivative, ΔHyg^R, Kan^R region isolated from pZerO-2	This study
pSR129	pMycoMar derivative, *TnQuoVadis* mutagenesis vector	This study

Plasmid construction

Plasmids used in this study are listed in Table 2-1. To construct a plasmid with a T4 transcription termination sequence downstream of a kanamycin (Km) selection region, the 1.2kb Km selection region was amplified from pZerO-2 with primers pZerO_kanR_F and pZerO_kanR_R (Table 2-2) resulting in a product in which the reverse primer contained a *HindIII* restriction site. This fragment was digested with *HindIII* and cloned into *EcoRV* and *HindIII* digested pSE100 (Guo *et al*., 2007) adjacent to the T4 transcription termination sequence to generate pSR118.

To construct an integrative vector constitutively expressing TetR to be used with *TnQuoVadis* mutagenesis, the Km region was excised from pMC1s by digestion with *NsiI* and *BspHI* and the resulting liner vector was treated with T4 DNA polymerase. The Hyg selection marker from pSE100 was excised by digestion with *BspHI* and *SmaI*, treated with T4 DNA polymerase and cloned into the linerized pMC1s vector backbone to yield pSR125.

In order to analyze TetR activity of the parental strain VTY123, a reporter plasmid pSR126 was constructed as follows. The TetR region was deleted from reporter plasmid pMe1ml1 (Ehrt *et al*., 2005) by *NotI* digestion and re-ligated to generate pSR117. The hygromycin (Hyg) region in pSR117 was excised by *SacII* digestion and the vector backbone was treated with T4 DNA polymerase. This prepared the vector to receive the Kan region amplified from pZerO-2 (Invitrogen) with primers pZerO_kanR_F and pZerO_kanR_R (Table 2-2) that generated pSR126. This vector was used to analyze the TetR activity of strain VTY123 (data not shown).

To construct the mycobacteria mutagenesis vector carrying *TnQuoVadis*, the native transposon and transposase were removed from pMycoMar by digesting with *ScaI* and *BamHI* and the resulting vector backbone was treated with T4 DNA polymerase to prepare it to receive the *TnQuoVadis*/transposase cassette from pSR123 (Fig. 2-3B). Plasmid pSR123 was digested with *SacII* and *BamHI* and treated with T4 DNA polymerase, which yielded an insert containing the *TnQuoVadis*/transposase. These digestions prepared the vector and insert with compatible ends for ligation which yielded pSR129 (Fig. 2-3B).

All plasmid constructs were transformed into DH5αMCR (GibcoBRL) to increase yield and harvested. Chemically-competent DH5*α*MCR cells were prepared by the Inoue Method (Sambrook & Russell, 2001). All plasmids constructed in this study were confirmed by restriction digestion and/or DNA sequencing.

Transposon mutagenesis of *M. smegmatis*

Electro-Competent VTY123 cells, prepared as previously described (Parish & Brown 2008), were transformed with pSR129 (Fig. 2-3B) using electroporation by applying conditions as previously described (Hatfull & Jacobs, 2000). After a 4 hour recovery in LB broth at 30ºC, cells were plated on LB medium containing Gm and incubated at 30ºC for 4 days which allowed pSR129 replication. An estimated 88,000 transformants were harvested from plates by scrapping and were resuspended in liquid LB with 0.05% Tween 80 via syringe (22G1 ½) at approximately $1.2x10^7$ cells per ml. A solution of 50% glycerol was added to the cell suspension for a total concentration of 15% glycerol. 350 µl aliquots of this cell suspension were put into cryogenic tubes and stored at -80°C for later use in transposon mutagenesis. To identify transposon mutants in or near essential genes, approximately $1.2x10^4$ cells from the stocks were spread on individual

agar plates (100mm × 15mm standard petri dish) with LB containing Km and ATc and incubated for 3 days at 39°C, the non-permissive temperature for myco ts-ori (Guilhot *et al*., 1992).

TnQuoVadis mutant screening

Mutant colonies obtained at 39°C in the presence of ATc were replica-plated onto plates with and without ATc using velveteen squares (Scienceware), and grown at 39°C for 2 days. Colonies displaying ATc-dependent growth were patched consecutively onto plates with and without ATc to verify the mutant phenotype. Transposon insertion mutants were also tested for gentamycin sensitivity to ensure the loss of the plasmid. To examine growth by serial dilution with the spot assay, reproducible ATc-dependent mutants were grown in LB liquid media supplemented with 0.05% Tween 80 and harvested during the exponential growth phase. Cell suspensions of each mutant were standardized by measurement at OD_{600}, and 10 µl of cell suspension measuring an OD of 0.1, 0.01, and 0.001 were spotted on plates with and without ATc. The approximate amount of cells spotted were 500,000, 50,000 and 5,000 cells respectively. Results were documented after 2 days incubation. The phenotypes of the first 13 are shown in Figure 2-5.

Genomic DNA isolation

Mycobacteria genomic DNA preparation was carried out as previously described (Parish & Brown, 2008). The site of *TnQuoVadis* insertion in the identified mutants (Fig. 2-6) was determined as follows. Genomic DNA was digested with *SacII* which does not cut within the transposon and the resulting fragments were self-ligated. The 20 µl genomic DNA digestion reactions consisted of 10× Buffer 4 (New England Bioline), 1 µl of *SacII* at 20U/µl (New England Bioline) and ~1µg genomic DNA. The following 25 µl ligation reactions consisted of

10× ligase buffer (New England Bioline), 0.5 µl T4 DNA ligase (New England Bioline) and 10 µl of the digestion reaction mix. A 5-10 µl volume of the ligation mix was transformed into DH5*αλpir* which is capable of propagating plasmids with the R6Kγ origin. Recovered plasmids from the transformants, containing *TnQuoVadis* flanked by genomic DNA from its insertion site, were purified by the QIAprep protocol and sequenced using primer Mariner_SP_R3 that binds to the right end of *TnQuoVadis* (Fig. 2-2). The 10 µl sequencing reaction consisted of 5× BigDye® sequencing buffer (Applied Biosystems), 1 µl BigDye® reaction mix (Applied Biosystems), 3 µM of the primer, and ~250-500 ng double stranded plasmid template. Cycling parameters were as follows: 95°C for 1 minute, followed by 30 cycles of 95°C for 15 seconds, 50°C for 15 seconds, and 60°C for 4 minutes.

Bioinformatic analyses

To identify the *TnQuoVadis* insertion locations, the genomic DNA flanking the *TnQuoVadis* insertion sites were individually compared to the complete genome of *M. smegmatis* mc^2155 using the NCBI BLASTn algorithm to identify matching sequences (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi). In each of the 13 mutants, a matching sequence with close to 100% identity was identified. The ORFs corresponding to potential essential genes as well as other nearby genes in the genetic neighborhood were further compared to the complete genome of *M. tuberculosis* H37Rv using the NCBI BLASTp (Altschul *et al*., 1997) algorithm to identify protein product similarities (Table 2-3) and conserved regions.

RESULTS

Design of *TnQuoVadis*

A *mariner*-based transposon, designated *TnQuoVadis* (Fig. 2-2), was designed for mutagenesis of essential genes in mycobacteria. The *mariner* element was chosen due to its broad host range and relaxed site specificity (Lampe *et al*., 1996). The starting template was *magellan4,* a *mariner*-derivative used successfully in mycobacteria (Rubin *et al*., 1999); this transposon contained the *E. coli* R6Kγ origin of replication and a kanamycin (Km) resistance gene. Our major modification was the addition of the $P_{UV15tetO}$ inducible promoters immediately inside the inverted repeats (IR) (Fig. 2-2); $P_{UV15tetO}$ is an engineered mycobacterial promoter repressed by TetR-*tetO* interaction and induced by anhydrotetracycline (ATc) (Ehrt *et al*., 2005). The two $P_{UV15tetO}$ are positioned in *TnQuoVadis* to drive transcription outward from both ends of the transposon. A minor modification was the insertion of the transcription terminator of gene 32 of phage T4 (T4g32T) (Guo *et al*., 2007; Kaps *et al*., 2001) after the Km resistance gene to prevent read-through from the Km promoter (Fig. 2-2). If *TnQuoVadis* transposes onto the chromosome of a TetR-expressing mycobacterial host strain, the genes adjacent to the transposon in either orientation could be regulated by ATc. If the insertion occurs in the promoter or the 5' end of an essential gene, the transposon mutant may be viable only when ATc is present in the media.

Construction of *TnQuoVadis*

For the actual construction of *TnQuoVadis*, the left and right segments were assembled separately, followed by their ligation at a *SmaI* site (Fig. 2-2). The fragment on the left of *SmaI*

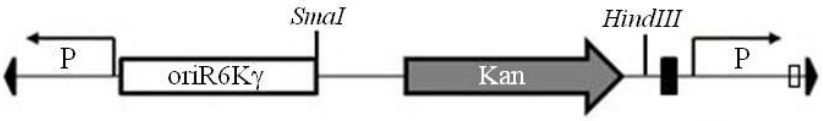

Figure 2-2. Features of the *TnQuoVadis* transposon. Depiction of the 2.8 kb *TnQuoVadis* element. *TnQuoVadis* has the inverted repeats (IRs) and the R6Kγ origin from *magellen4*. Immediately inside the IRs, depicted as black triangles, are two identical mycobacterial promoters ($P_{UV15tetO}$), labeled P, that are inducible by the addition of ATc and drive outward transcription from the transposon. The kanamycin resistance marker is for selection of the transposon and the transcription terminator, depicted as a black rectangle, prevents read-through from the kan promoter. The *E. coli* R6Kγ origin of replication facilitates the cloning and identification of transposon insertions. The *SmaI* and *HindIII* site used for constructing the transposon are indicated. The site for primer Mariner_SP_R3, depicted as a white rectangle, is used to sequence sites of *TnQuoVadis* insertion in genomic DNA.

contains the R6Kγ origin whereas the fragment on the right contains the Km gene (Fig. 2-2). Both fragments also contain the $P_{UV15tetO}$ promoter which was obtained by PCR using primers P_F1 and P_R1 (Table 2-2) with the plasmid pUV15*tetO*Rm (Ehrt *et al*., 2005) as the template.

For the assembly of the R6Kγ-containing fragment (Fig. 2-2), *magellan4* was first liberated from pMyr6k (Rubin *et al*., 1999) by *XbaI* and *SpeI* digestion and cloned into the same sites in pWB300 (Black, 2005) to generate pSR114. Using pSR114 as a template, a *mariner* IR was PCR amplified using primers M13F and 5_P_R1 (Table 2-2). This IR and the $P_{UV15tetO}$ fragments were joined by overlapping PCR because the 5' end of 5_P_R1 is complementary with P_R1. This IR- $P_{UV15tetO}$ fragment was further joined with R6Kγ which was PCR amplified using primers P_F2 and P_R2 (Table 2-2) with pSR114 as the template. The resulting fragment, which consists of the left half of *TnQuoVadis* (Fig. 2-2), was digested with *XbaI* and *SmaI* and cloned into the same sites in pBluescript KS to generate pSR120 (Table 2-1).

For the assembly of the Km-containing fragment (Fig. 2-2), a Km gene fragment was PCR amplified from pZerO-2 (Invitrogen) using primers pZErO_kanR_F and pZErO_kanR_R (Table 2-2). This fragment was cloned into pSE100 (Guo *et al*., 2007) as described in Material and Methods, and the resulting plasmid pSR118 contains the Km gene next to T4g32T. The Km-T4g32T fragment was then PCR amplified using primers pZErO_kanR_F and T4g32_kanRterm_R (Table 2-2). Next, a *mariner* IR was PCR amplified from pSR114 using primers M13F and 3_Ptet_F2 (Table 2-2). An IR- $P_{UV15tetO}$ fragment was obtained by overlapping PCR because 3_Ptet_F2 has complementarity with Ptet_R1. The Km-T4g32T and the IR-$P_{UV15tetO}$ fragments were then joined by overlapping PCR because part of T4g32_kanRterm_R is complementary to Ptet_F1. The resulting product, which forms the right half of the transposon (Fig. 2-2), was cloned into *SmaI* digested pBluescript KS to generate pSR119 (Fig. 2-3A).

Table 2-2. Oligonucleotides used in this study[a].

Primer name	Primer sequence 5'- 3'
P_F1	AATATTGGATCGTCGGCACC
P_R1	GGATCGTGCTCATTTCGGGC
M13F	CCCAGTCACGACGTTGTAAAACG
M13R	AGCGGATAACAATTTCACACAGG
5_P_R1	**GCCCGAAATGAGCACGATCC**TAATATGCATTTAATACTAG
P_F2	GGTGCCGACGATCCAATATTTATATAGTACCAACCTTCAA
P_R2	TCCTCGGTACCGGGCCCGGGATCCCTTAATTAACCCCGAAAA
pZErO_kanR_F	CCGGGCCCGGTACCGAGGACACGTAGAAAGCCAGTCCGCA
pZErO_kanR_R	GCT***AAGCTT***GGAACAACACTCAACCCTATCGC
T4g32_R	GGTGCCGACGATCCAATATTTAGCAATGCCTCCATGCGAT
3_P_F2	**GCCCGAAATGAGCACGATCC**TTCGACGCGTCAATTCGAGG
TPase_F	TCC***CCGCGG***CCGTCCAGTCTGGCAG
TPase_R	GC***TCTAGA***TTATTCAACATAGTTCCCT
Mariner_SP_R3	TGAAGGGAACTATGTTGAAT

[a] Complimentary regions in relation to P_F1 are underlined, complimentary regions in relation to P_R1 are bolded. The *HindIII* site in pZErO-KanR_R, the *SacII* site in TPase_F and the *XbaI* site in TPase_R are bolded and italicized.

The last step in the construction of *TnQuoVadis* is illustrated in Fig. 2-3A. The left fragment in pSR120 was excised by *XbaI* and *SmaI* digestion and cloned into the same sites in pSR119. This gave rise to pSR121 which now contains the full length *TnQuoVadis* (Figs. 2-2 & 2-3A).

Construction of the *TnQuoVadis* delivery vector

Although pSR121 contains the full-length *TnQuoVadis* transposon, it has no transposase (TPase) to catalyze transposition. The *mariner* TPase gene was PCR amplified from pMycoMar (Rubin *et al*., 1999) with primers TPase_F and TPase_R (Table 2-2). This fragment was digested with *SacII* and *XbaI* and cloned into pSR121. The resulting plasmid pSR123 (Fig. 2-3B) has both *TnQuoVadis* and the TPase gene. When we transformed pSR123 into *M. smegmatis* strain mc^2155, we obtained only low numbers of transposon mutants by direct selection of Km resistance carried by the transposon. For more efficient transposon mutagenesis, we moved the *TnQuoVadis*-TPase cassette in pSR123 into a vector with a mycobacterial thermosensitive origin of replication (myco ts-ori) (Guilhot *et al*., 1992). To accomplish this, the transposon and TPase from pMycoMar were replaced by the *TnQuoVadis*-TPase cassette from pSR123 as described in Materials and Methods (Fig. 2-3B). The resulting delivery plasmid pSR129 contains *TnQuoVadis,* the transposase gene, the myco ts-ori and the gene for gentamycin (Gm) resistance (Fig. 2-3B). Transformants with pSR129 can be selected by Gm resistance and propagated at 30°C, the permissive temperature for the thermosensitive origin.

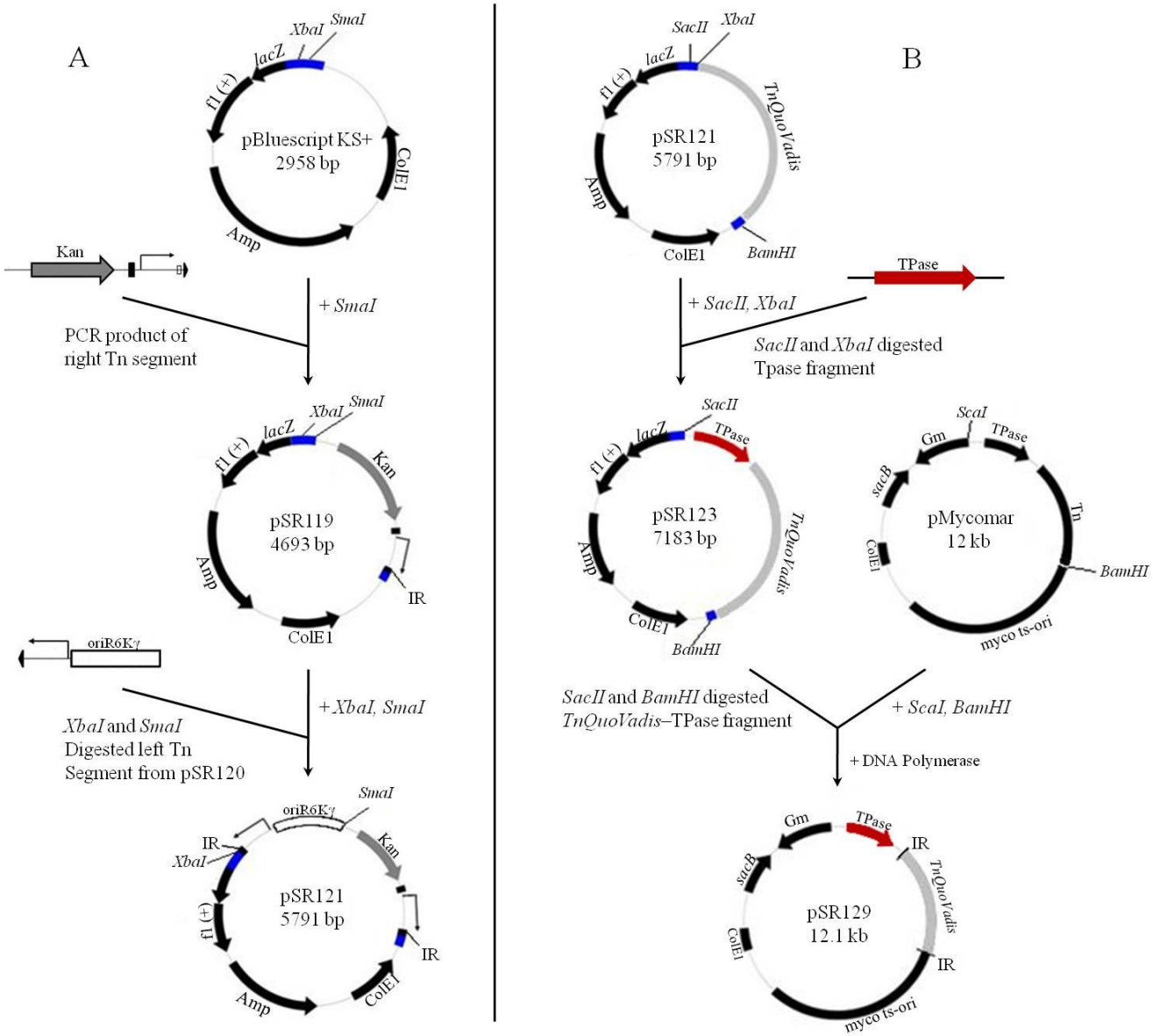

Figure 2-3. Formation of *TnQuoVadis* and construction of pSR129. (A) The right transposon segment was blunt cloned into *SmaI* digested pBluescript KS+ generating pSR119. The left transposon segment was digested with *XbaI* and *SmaI* and cloned into the same sites in pSR119 generating pSR121 and forming full length *TnQuoVadis.* The blue region indicates the MCS in *lacZ.* (B) The PCR product containing the transposase was digested with *SacII* and *XbaI* and cloned into the same sites in pSR121 at the left end of *TnQuoVadis* forming the *TnQuoVadis*-TPase construct. The *SacI* and *BamHI* site on pMycoMar and the *SacII* and *BamHI* site in pSR121 containing the *TnQuoVadis*-TPase construct were both digested then blunted with T4 DNA Polymerase and ligated to generate pSR129. pSR129 is a mycobacteria thermosensitive (ts) delivery vector which replicates in *M. smegmatis* at 30°C but not at 39°C. Gentamycin (Gm) is used for selection and screening purposes when necessary.

Construction of a TetR-expressing *M. smegmatis* strain for *TnQuoVadis* transposon mutagenesis

In order to use *TnQuoVadis* to mutagenize essential genes, it was necessary to construct a mycobacterial strain that constitutively expresses TetR to regulate the $P_{UV15tetO}$ promoters in the transposon. Plasmid pMC1s (Guo *et al*., 2007), which integrates at the phage L5 attachment site, was previously constructed to express TetR in mycobacteria from a constitutive promoter. This plasmid, however, contains the same Km selection marker as the transposon. We replaced the Km resistance gene on pMC1s with a hygromycin (Hyg) resistance gene as described in Materials and Methods. The resulting plasmid, pSR125, was introduced by electroporation into the *M. smegmatis* strain mc^2155 (Snapper *et al*., 1990) to construct the TetR-expressing strain VTY123.

Screening for mutants with ATc-dependent growth

The general screening procedures for mutants with transposons regulating potentially essential genes are shown in Fig. 2-4. We first prepared stocks of VTY123 transformants with pSR129 as described in Materials and Methods. These transformants were selected for Gm resistance at 30°C. After 4 days of incubation, an estimated 88,000 transformants were harvested from the plates and resuspended in LB with 0.05% Tween 80 at approximately $1.2x10^7$ cells per ml. This cell suspension was aliquoted and stored at -80°C for later use.

To identify transposon mutants in or near essential genes, approximately $1.2x10^4$ cells from the stock were spread on individual agar plates (100mm × 15mm standard petri dish) with LB containing Km and ATc. After 3 days of incubation at 39°C, the non-permissive temperature

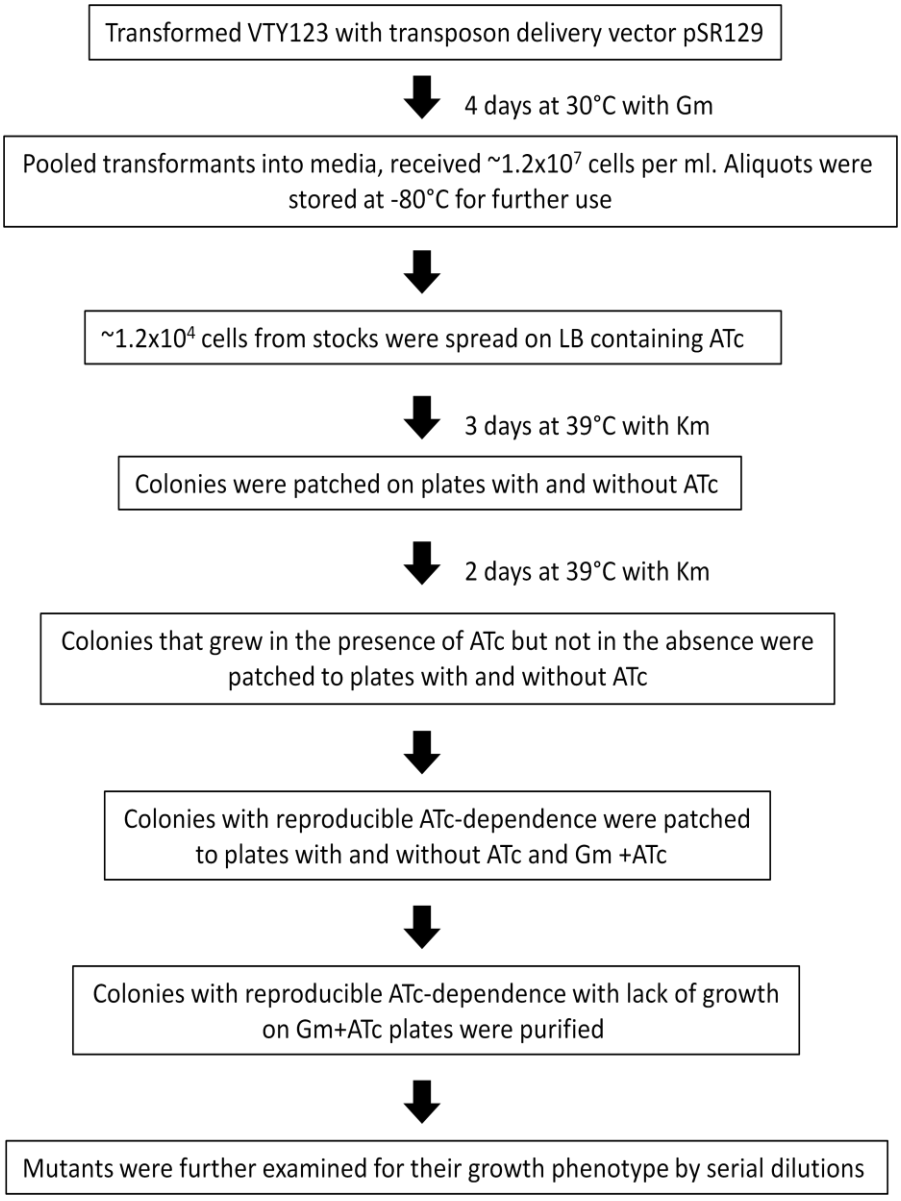

Figure 2-4. Mutant screening protocol. Depiction displaying the protocol for preparing and screening transposon insertion mutations.

for myco ts-ori (Guilhot *et al*., 1992), these colonies were patched onto plates with and without ATc. Approximately 150-300 colonies typically appeared per plate. We selected those that displayed relatively normal growth with ATc but no or diminished growth without ATc for further examination. Colonies selected were found to be sensitive to Gm, indicating the loss of the delivery plasmid and the chromosomal transposition of *TnQuoVadis*. In our first attempt with approximately 5,200 colonies, 13 transposon insertion mutants were found to have reproducible ATc-dependent growth during subsequent patching. The growth of these mutants was further examined by serial dilutions using the spot assay with and without ATc in comparison with WT (Fig. 2-5) as described in Material and Methods. As shown in Fig. 2-5, VTY201 and VTY204 have a moderate ATc-dependency while VTY205 displayed a weak growth deficiency without ATc. In contrast, the other 10 strains have a strong ATc-dependency; they grew very little in the absence of ATc with relatively normal growth when ATc was present in the growth media.

Over 200,000 transposition mutants have now been screened which yielded a total of 121 conditional ATc-dependent mutants (data not shown). Of these mutants, 20 displayed weak ATc-dependence, 24 displayed moderate ATc-dependence, and 77 displayed a strong dependence upon ATc for growth on LB agar.

Identification of *TnQuoVadis* insertion locations in mutants

The transposons with flanking genomic regions were cloned from the first 13 mutants and sequenced (Materials and Methods). The sequences of the genomic regions were used for BLAST searches against the *M. smegmatis* mc²155 reference genome sequence (Fleischmann *et al*., 2006) to identify the transposon insertion sites. Nine distinct genetic loci, I through IX (Fig. 2-6), were identified by these insertions. All loci except VI, VIII and IX were identified by single

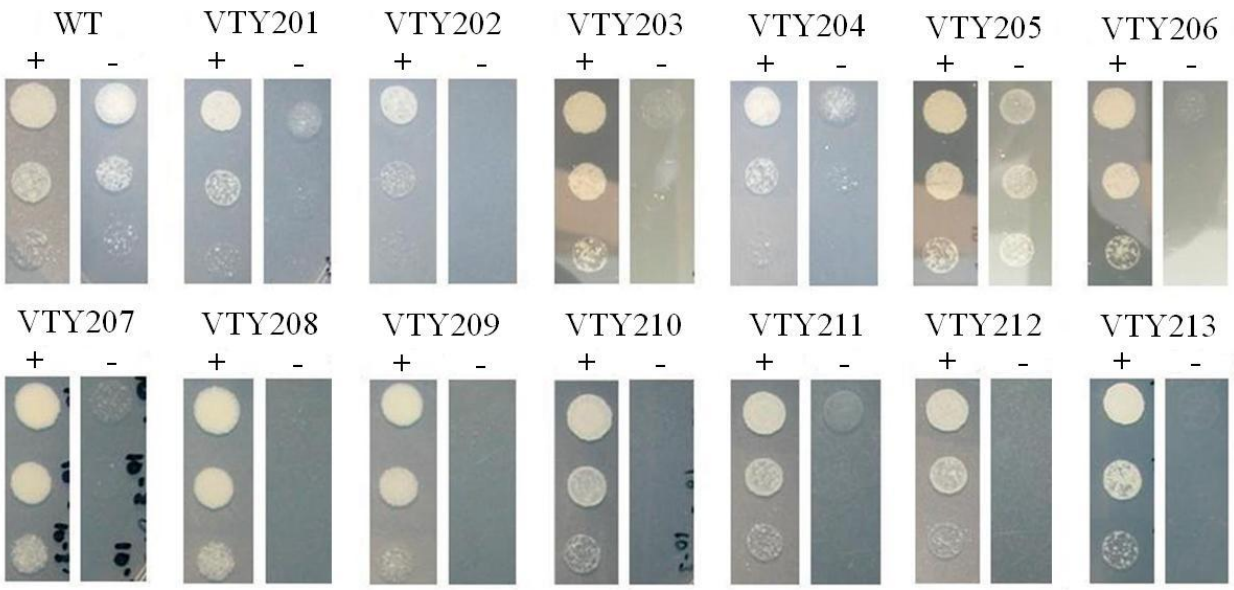

Figure 2-5. ATc-dependent growth of mutants. The varying degrees of growth of the 13 mutants from the first round of screening are shown. For each strain, 10 µl of cell suspension at OD = 0.1, 0.01 and 0.001 were spotted on plates from top to bottom and incubated for 2 days before documentation. "+" indicates presence of ATc and "–" indicates absence of ATc.

insertions. Locus VI was identified by two different insertions in VTY206 and VTY209, locus VIII by the exact same insertions in VTY208 and VTY213, and locus IX by three different insertions in VTY202, VTY210, and VTY212 (Fig. 2-6).

The position of transposon insertions and the genomic context of the nine genetic loci (Fig. 2-6 and Table 2-3) were analyzed for candidates of essential genes identified by these *M. smegmatis* mutants. MSMEG_0028 is the only candidate from locus I (Fig. 2-6). It is annotated as PknB based on its homology to the known essential gene with the same name in *M. tuberculosis* (Fernandez *et al*., 2006). The candidate from locus II is MSMEG_2580 (Fig. 2-6); it is annotated as IspG which is involved in the synthesis of isoprenoid, a key precursor for mycolic acid (Brown *et al*., 2010). MSMEG_2581 is likely not in the same operon with *ispG* because it is about 100 basepairs (bps) downstream. For locus III, MSMEG_4495-4493 are likely in the same operon with MSMEG_4496; among these, MSMEG_4493 encodes the 16S rRNA-binding GTPase Era which has essential functions in *E. coli* (Zhang & Inouye, 2002). Era is involved in many cellular processes and is known to be universally found in all bacteria except some eubacteria (Caldon and March, 2003). Locus IV identified a gene cluster important for peptidoglycan synthesis: MSMEG_4902 encodes a putative β-lactamase, for example, and MSMEG_4903 encodes the MurI racemase which converts L-glutamate to D-glutamate, an essential component of peptidoglycan (Sengupta and Nagaraja., 2008). The insertion at locus V occurred upstream of a potential five-gene operon (Fig. 2-6); *prfA* (MSMEG_4950) and *hemK* (MSMEG_4949), which are suspected to be essential in *M. tuberculosis* (Sassetti *et al*., 2003), are among these genes. The insertions at locus VI were in an intergenic region and the potential essential genes are among MSMEG_5220-5222 in one direction and MSMEG_5223 in the other; interestingly, while MSMEG_5223 is a predicted gene of unknown function, the protein product

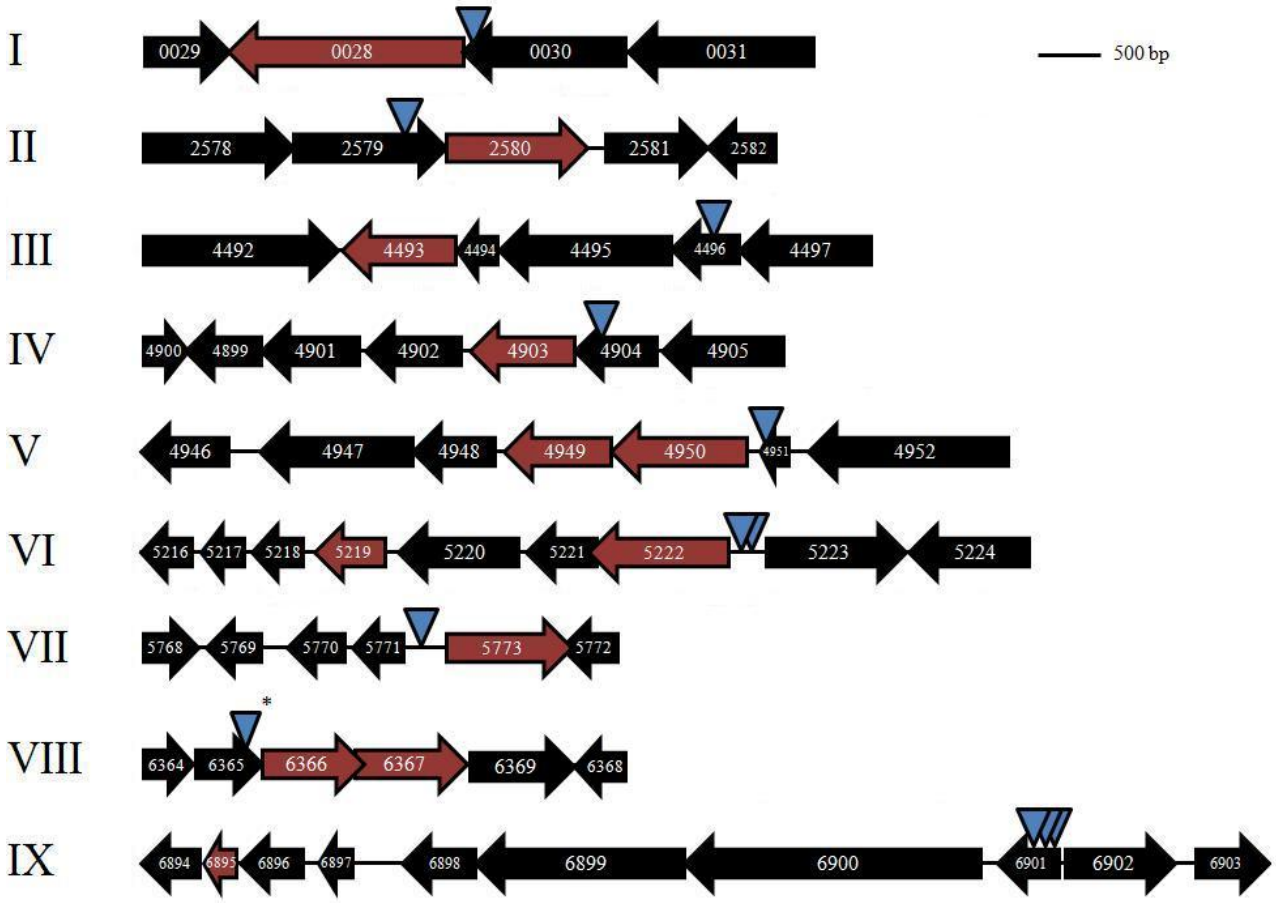

Figure 2-6. *TnQuoVadis* insertion sites. Depiction of the identified genetic loci. Blue triangles indicate site of transposon insertions. Red arrows indicate potential essential genes. Each gene is numbered according to the mc^{2}155 reference genome. * next to a blue triangle indicates two identical transposon insertions. (I) Mutant strain VTY211, (II) Mutant strain VTY204, (III) Mutant strain VTY201, (IV) Mutant strain VTY205, (V) Mutant strain VTY203, (VI) Mutant strains VTY206 and VTY209, (VII) Mutant strain VTY207, (VIII) Mutant strains VTY208 and VTY213, (IX) Mutant strains VTY202, VTY210, and VTY212.

of MSMEG_5222 is homologous to Era at locus III. It is also possible that MSMEG_5219, which is possibly essential in *M. tuberculosis* (Sassetti *et al*., 2003), is affected by the transposon insertions at locus VI. The insertion at locus VII was also in an intergenic region and it identified MSMEG_5769-5771 which encode genes of unknown function, and MSMEG_5773 which encodes a fatty acid desaturase that is suspected to be essential in *M. tuberculosis* (Sassetti *et al*., 2003). The two identical insertions at locus VIII occurred upstream of a potential three-gene ABC transporter operon (Fig. 2-6) which included an ATP-binding protein (MSMEG_6366) and a glycosyl transferase (MSMEG_6367), both of which are suspected to be essential in *M. tuberculosis* (Sassetti *et al*., 2003). The three insertions at locus IX clustered in MSMEG_6901 and the potential essential genes are among MSMEG_6898, 6899 and 6902 along with MSMEG_6895 which encodes the 30S ribosomal protein S18 found to be essential in *E. coli* (Bubunenko *et al*., 2007). Interestingly, *M. smegmatis* mutants that contain knockouts in MSMEG_6900 which encodes a penicillin-binding protein and *E. coli* mutants with knockouts in the gene corresponding to MSMEG_6897, which encodes the 30S ribosomal protein S6, are both known to be viable however grow extremely slowly in liquid media (Billman-Jacobe *et al.*, 1999; Bubunenko *et al.*, 2007).

The nine *M. smegmatis* genetic loci (Fig. 2-6 and Table 2-3) identified by these mutants showed varying degrees of conservation in *M. tuberculosis*. The gene orders at loci I, III and V are strictly conserved in *M. tuberculosis.* For locus I, both the sequence and order for MSMEG_0028 (PknB) and MSMEG_0029 (PknA) are conserved and the conservation extended beyond these two genes in both directions (Wolfe *et al.*, 2010). This is also the case for locus III in which MSMEG_4493 (Era) and genes in both directions (MSMEG_4492 and MSMEG_4494-4497) are

Table 2-3. Annotation of *TnQuoVadis* insertions.

Locus[a]	Annotation[b]	TB H37Rv identity (locus)[c]	Reference[d]
I	Serine-threonine protein kinase	70% identity to PknB (Rv0014c)	Fernandez *et al*., 2006
II	*ispG* - isoprenoid synthesis	91% identity to IspG (Rv2868c)	Brown *et al*., 2010
III	GTP-binding protein Era	83% identity to Era (Rv2364c)	Sassetti *et al*., 2003
IV	*murI* - glutamate racemase	86% identity to MurI (Rv1338)	Sengupta and Nagaraja., 2008
V	*prfA* - peptide chain release factor 1;	84% identity to PrfA (Rv1299)	Sassetti *et al*., 2003
	modfication methylase, HemK family protein	64% identity to probable HemK (Rv1300)	Sassetti *et al*., 2003
VI	GTP-dependent nucleic acid-binding protein EngD;	86% identity to EngD (Rv1112)	Teplyakov *et al*., 2003
	riboflavin biosynthesis RibD domain-containing protein	26% identity to RibG (Rv1409)	Sassetti *et al*., 2003
VII	fatty acid desaturase	82% identity to DesA1 (Rv0824c)	Sassetti *et al*., 2003
VIII	ABC transporter, O-antigen export system, ATP-binding protein	88% identity to RfbE (Rv3781)	Sassetti *et al*., 2003
	glycosyl transferase	77% identity to L-rhamnosyltransferase (Rv3782)	Sassetti *et al*., 2003
IX	*rpsR* - 30S ribosomal protein S18	55% identity to RpsR (Rv2055c)	Bubunenko *et al*., 2007

[a] Identified genetic loci in *M. smegmatis.*
[b] Name and function of potential essential *M. smegmatis* genes regulated by *TnQuoVadis* insertions.
[c] Protein identity BLAST hits of TB strain H37Rv compared to regulated genes of ATc-dependent *M. smegmatis* mutants.
[d] References for the potential and known essential TB genes.

highly conserved (Fleischmann *et al*., 2002), and for locus V for which the sequence and order for MSMEG_4947 - 4951 are highly conserved. The gene orders at loci II, IV, and VI were semi-conserved in *M. tuberculosis.* For locus II, the sequence and order for MSMEG_2578, 2579 and MSMEG_2580 (IspG) were highly similar but the surrounding genes were not (Wolfe *et al.*, 2010; Cole *et al.*, 1998). For locus IV, the sequence and order for MSMEG_4899, 4901, 4903 (MurI) and MSMEG_4905 were conserved in *M. tuberculosis* with the surrounding genes not conserved, as well as for locus VI in which the sequence and order for MSMEG_5222, 5223, and 5224 were conserved. The gene sequences of the remaining loci, VII, VIII and IX, were conserved in *M. tuberculosis* but the order and sequences of the surrounding genes were not conserved.

DISCUSSION

Here we describe the development and use of a novel transposon that allows for the identification of essential genes in *M. smegmatis.* This approach utilizes the TetR(B)2 mycobacterial expression system (Fig. 1-2). Two ATc-inducible promoters were placed facing outwards at either ends of a *mariner* transposon (Fig. 2-2). In order to mutagenize potential essential genes, we constructed a mycobacteria strain that constitutively expresses TetR to regulate the inducible promoters in the transposon. When the transposon is inserted upstream of an essential gene (Fig. 2-1), the mutants will only be viable in the presence of ATc, and in the absence of ATc, they will cease to grow or show significantly diminished growth. This allows for their easy differentiation from transposon mutants of non-essential genes. The identification of the regulated genes of these mutants is also straightforward because each mutation is marked by the transposon insertion. Due to the relatively low transformation of mycobacteria, we used a temperature-sensitive delivery strategy in which the engineered transposon was inserted into a thermosensitive plasmid which significantly increased our mutagenesis.

For proof-of-concept, we chose to mutagenize the model species *M. smegmatis* for ease of experimentation. The system performed well in the initial testing, producing 13 conditional ATc-dependent mutants with varying phenotypes (Fig. 2-5). These mutants were classified on three degrees of ATc-dependency, categorized as either having a weak, moderate or strong growth deficiency without ATc. The initial 13 screened mutants displayed 1 weak, 2 moderate and 10 strong ATc-dependent phenotypes when grown on LB agar. This skewed result towards strong ATc-dependent phenotypes could be due to several reasons. First, the screening technique had yet to be perfected; strong ATc-dependent phenotypes among a backdrop of thousands are

the most noticeable and thus the first to be isolated. Second, the nature of the screening protocol on LB media likely overlooks essential genes necessary for growth in other conditions, thus more numerous mutants that produce the mild or moderate ATc-dependency may only display those phenotypes in other growth conditions. Finally, experiments using a mycobacterial codon-adapted *tetR* gene may be useful for isolating a greater variety of mutants. The Tet system is tightly regulated but it still allows for low level basal transcription, making it leaky. Due to the high G-C content of the mycobacteria, the coding regions display an inherent codon bias. Using a codon-adapted *tetR* gene may lead to improved translation which will increase TetR protein in the cell and thus tighten repression. This strategy has recently been used successfully in mycobacterial systems (Klotzsche *et al.,* 2009).

From the first trial, we were able to isolate mutants to identify genes necessary for a variety of mycobacterial cellular biosynthetic pathways (Table 2-3), thus validating our approach. These *TnQuoVadis* insertions were upstream of genes that affected the overall viability of the cell. Many of the identified genes were either known essential or potentially essential in mycobacteria. For example, two genes were identified to be involved in cell wall biosynthesis: *ispG* and *murI.* The IspG protein, which is required for mycolic acid biosynthesis, is vital to mycobacteria cells, and MurI, a glutamate racemase, produces a component of peptidoglycan. Both of these genes are known essential in mycobacteria (Brown *et al*., 2010; Sengupta and Nagaraja, 2008). All of the mutants contained transposon insertions that were either directly upstream of an essential gene or within two genes distance (Fig. 2-6) with the exceptions of VTY202, VTY210 and VTY212 (locus IX). These mutants were peculiar due to the fact that the known essential gene (*rpsR*) was 5 ORFs away from the transposon insertion, and these genes were not likely in an operon. The more likely essential gene candidates,

MSMEG_6898, 6899 and 6902, are genes of unknown function. It is also possible that MSMEG_6900 and MSMEG_6897 are under regulation. MSMEG_6900 encodes a penicillin-binding protein, and although it is non-essential, *M. smegmatis* mutants defective in this region grow slowly in liquid media (Billman-Jacobe *et al.*, 1999). Similarly, MSMEG_6897 encodes the 30S ribosomal protein S6 which is non-essential, but knockouts of the corresponding gene in *E. coli* grow extremely slowly in liquid media (Bubunenko *et al.,* 2007). Thus, it is possible that the combination of both phenotypes led to the isolation of these mutants. The results also indicated that the screening method is prone to produce multiple mutations in the same regions. Out of the 13 mutants, there were 9 identified genetic loci with three of the genetic loci sites containing *TnQuoVadis* insertions in multiple strains (Fig. 2-6). The phenotype of these mutants all displayed a strong ATc-dependency, suggesting the apparent bias in the initial screening. Also, as more mutants are sequenced, identified genetic loci with multiple insertions will increase.

There is much work to be done. This project pioneered a new approach in identifying essential genes, leaving multiple avenues open for experimentation. Future work includes but is not limited to: continuing the identification of essential genes from the currently available mutants, screening for additional mutants, testing identified genes of unknown function for essentiality, and characterizing identified essential genes as potential drug targets. The utility of the transposon could be also explored by identifying other genes of interest (genes encoding biofilm formation, motility, etc) by selecting for the desired phenotype on agar plates or in liquid media with and without ATc.

TnQuoVadis and its delivery vector provide a powerful new tool for genome-wide analysis of mycobacteria. Identifying potential essential genes using this tool is quick and

inexpensive compared to established methods. Although this system was designed for use in *M. smegmatis*, it has the potential to be adapted for other organisms that are permeable to ATc and can modulate transcription from a modified promoter containing *tetO* sites. It is understood that the low basal activity of expression from the *tetO* promoter may limit the essential genes that can be identified by *TnQuoVadis*, thus complete saturation of the *M. smegmatis* genome may be difficult using this method. However, adjusting the growth conditions, using a variety of media, and utilizing a codon-adapted *tetR* gene may overcome some of these limitations.

LITERATURE CITED

Ade´kambi, T., and Drancourt, M. (2004) Dissection of phylogenetic relationships among 19 rapidly growing *Mycobacterium* species by 16S rRNA, hsp65, sodA, recA and rpoB gene sequencing. *International Journal of Systematic and Evolutionary Microbiology*, **54**, 2095–2105

Alahari, A., Trivelli, X., Guerardel, Y., Dover, L.G., Besra, G.S., Sacchettini, J.C., Reynolds, R.C., Coxon, G.D. and Kremer, L. (2007) Thiacetazone, an Antitubercular Drug that Inhibits Cyclopropanation of Cell Wall Mycolic Acids in Mycobacteria. *Plos One*, **2**.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. **25**,3389-3402.

Argast, M. and Beck, C.F. (1984) Tetracycline diffusion through phospholipid bilayers and binding to phospholipids. *Antimicrobial agents and chemotherapy*, **26**, 263-265.

Bicmen, C., Coskun, M., Gunduz, A.T., Senol, G., Cirak, A.K. and Tibet, G. (2010) Nontuberculous mycobacteria isolated from pulmonary specimens between 2004 and 2009: causative agent or not? *New Microbiol*, **33**, 399-403.

Billman-Jacobe, H., Haites, R.E. and Coppel, R.L. (1999) Characterization of a *Mycobacterium smegmatis* mutant lacking penicillin binding protein 1. *Antimicrobial agents and chemotherapy*, **43**, 3011-3013.

Black, W.P. (2005) Regulation of exopolysaccharide production in *Myxococcus xanthus*. Blacksburg, Va.: University Libraries, Virginia Polytechnic Institute and State University. etd-12232005-042454.

Blokpoel, M.C., Murphy, H.N., O'Toole, R., Wiles, S., Runn, E.S., Stewart, G.R., Young, D.B. and Robertson, B.D. (2005) Tetracycline-inducible gene regulation in mycobacteria. *Nucleic acids research*, **33**, e22.

Brown, A.C., Eberl, M., Crick, D.C., Jomaa, H. and Parish, T. (2010) The nonmevalonate pathway of isoprenoid biosynthesis in *Mycobacterium tuberculosis* is essential and transcriptionally regulated by Dxs. *J Bacteriol*, **192**, 2424-2433.

Bubunenko, M., Baker, T. and Court, D.L. (2007) Essentiality of ribosomal and transcription antitermination proteins analyzed by systematic gene replacement in *Escherichia coli*. *J Bacteriol*, **189**, 2844-2853.

Caldon, C.E. and March, P.E. (2003) Function of the universally conserved bacterial GTPases. *Curr Opin Microbiol*, **6**, 135-139.

Carroll, P., Muttucumaru, D.G.N. and Parish, T. (2005) Use of a tetracycline-inducible system for conditional expression in *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*. *Appl Environ Microb*, **71**, 3077-3084.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd *et al.* (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, **393**, 537-544.

Collins, C. H., J. M. Grange, and M. D. Yates. (1984). Mycobacteria in water. J. Appl. Bacteriol. **57,**193–211.

Ehrt, S. and Schnappinger, D. (2006) Controlling gene expression in mycobacteria. *Future Microbiol*, **1**, 177-184.

Ehrt, S., Guo, X.V., Hickey, C.M., Ryou, M., Monteleone, M., Riley, L.W. and Schnappinger, D. (2005) Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Res*, **33**, e21.

Falkinham, J.O., Parker, B.C. and Gruft, H. (1980) Epidemiology of Infection by Nontuberculous Mycobacteria .1. Geographic-Distribution in the Eastern-United-States. *Am Rev Respir Dis*, **121**, 931-937.

Fernandez, P., Saint-Joanis, B., Barilone, N., Jackson, M., Gicquel, B., Cole, S.T. and Alzari, P.M. (2006) The Ser/Thr protein kinase PknB is essential for sustaining mycobacterial growth. *J Bacteriol*, **188**, 7778-7784.

Fine, P.E.M., Carneiro, A.M., Milstein, J.B., Clements, J.C. (1999) Issues relating to the use of BCG in immunization programmes. A discussion document. Geneva: WHO.

Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D. *et al.* (2002) Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol*, **184**, 5479-5490.

Fleischmann,R.D., Dodson,R.J., Haft,D.H., Merkel,J.S., Nelson,W.C. and Fraser,C.M. (2006) Direct Submission. The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA.

Frieden,T.R., Sterling,T.R., Munsiff,S.S., Watt,C.J. and Dye,C. (2003) Tuberculosis. Lancet, **362**, 887–899.

Gangadharam, P.R.J. and Jenkins, P.A. (1998) *Mycobacteria*. Chapman & Hall, New York.

Guilhot, C., Gicquel, B. and Martin, C. (1992) Temperature-sensitive mutants of the *Mycobacterium* plasmid pAL5000. *FEMS Microbiol Lett*, **77**, 181-186.

Guo, X.V., Monteleone, M., Klotzsche, M., Kamionka, A., Hillen, W., Braunstein, M., Ehrt, S. and Schnappinger, D. (2007) Silencing essential protein secretion in *Mycobacterium smegmatis* by using tetracycline repressors. *J Bacteriol*, **189**, 4614-4623.

Hatfull, G.F. and Jacobs, W.R. (2000) *Molecular genetics of mycobacteria*. ASM Press, Washington, D.C.

Hutchison, C.A., Peterson, S.N., Gill, S.R., Cline, R.T., White, O., Fraser, C.M., Smith, H.O. and Venter, J.C. (1999) Global transposon mutagenesis and a minimal Mycoplasma genome. *Science*, **286**, 2165-2169.

Kaps, I., Ehrt, S., Seeber, S., Schnappinger, D., Martin, C., Riley, L.W. and Niederweis, M. (2001) Energy transfer between fluorescent proteins using a co-expression system in *Mycobacterium smegmatis*. *Gene*, **278**, 115-124.

Klotzsche, M., Ehrt, S., and Schnappinger D., (2009) Improved tetracycline repressors for gene silencing in mycobacteria. *Nucleic Acids Research*, 2009. **37**(6), 1778-1788.

Lampe, D.J., Churchill, M.E. and Robertson, H.M. (1996) A purified mariner transposase is sufficient to mediate transposition in vitro. *EMBO J*, **15**, 5470-5479.

Liu, J. and Nikaido, H. (1999) A mutant of *Mycobacterium smegmatis* defective in the biosynthesis of mycolic acids accumulates meromycolates. *Proc Natl Acad Sci USA*, **96**, 4011-4016.

Miller, J.H. (1972) *Experiments in molecular genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Mushegian, A.R. and Koonin, E.V. (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci USA*, **93**, 10268-10273.

Parish, T. and Brown, A.C. (2008) *Mycobacteria protocols*. 2nd ed. Methods in molecular biology. 465. Humana Press, New York, NY.

Parish, T., Turner, J. and Stoker, N.G. (2001) *amiA* is a negative regulator of acetamidase expression in *Mycobacterium smegmatis*. *BMC Microbiol*, **1**, 19.

Pfyffer, G. E., Brown-Elliot, B. A. & Wallace, R. J., Jr (2003). *Mycobacterium*: general characteristics, isolation, and staining procedures. *Manual of Clinical Microbiology*, 8th edn, pp. 532–559. Washington, DC: American Society for Microbiology.

Ranes, M.G., Rauzier, J., Lagranderie, M., Gheorghiu, M. and Gicquel, B. (1990) Functional analysis of pAL5000, a plasmid from *Mycobacterium fortuitum*: construction of a "mini" Mycobacterium-Escherichia coli shuttle vector. *J Bacteriol*, **172**, 2793-2797.

Roberts, G., Muttucumaru, D.G. and Parish, T. (2003) Control of the acetamidase gene of *Mycobacterium smegmatis* by multiple regulators. *FEMS Microbiol Lett*, **221**, 131-136.

Rubin, E.J., Akerley, B.J., Novik, V.N., Lampe, D.J., Husson, R.N. and Mekalanos, J.J. (1999) In vivo transposition of mariner-based elements in enteric bacteria and mycobacteria. *Proc Natl Acad Sci USA*, **96**, 1645-1650.

Sambrook, J. and Russell, D.W. (2001) *Molecular cloning : a laboratory manual*. 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Sassetti, C.M., Boyd, D.H. and Rubin, E.J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol*, **48**, 77-84.

Sengupta, S. and Nagaraja, V. (2008) Inhibition of DNA gyrase activity by *Mycobacterium smegmatis* MurI. *FEMS Microbiol Lett*, **279**, 40-47.

Snapper, S.B., Melton, R.E., Mustafa, S., Kieser, T. and Jacobs, W.R., Jr. (1990) Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol Microbiol*, **4**, 1911-1919.

Snyder, L. and Champness, W. (2007) *Molecular genetics of bacteria*. 3rd ed. ASM Press, Washington, D.C.

Teplyakov, A., Obmolova, G., Chu, S.Y., Toedt, J., Eisenstein, E., Howard, A.J. and Gilliland, G.L. (2003) Crystal structure of the YchF protein reveals binding sites for GTP and nucleic acid. *J Bacteriol*, **185**, 4031-4037.

Tortoli, E. (2006) The new mycobacteria: an update. *FEMS Immunol Med Microbiol*, **48**, 159-178.

Triccas, J.A., Parish, T., Britton, W.J. and Gicquel, B. (1998) An inducible expression system permitting the efficient purification of a recombinant antigen from *Mycobacterium smegmatis*. *FEMS Microbiol Lett*, **167**, 151-156.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G.F., Chater, K.F. and van Sinderen, D. (2007) Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol Mol Biol Rev*, **71**, 495-548.

Wei, J.R. and Rubin, E.J. (2008) The many roads to essential genes. *Tuberculosis*, **88**, S19-S24.

White, D. (2007) *The physiology and biochemistry of prokaryotes*. 3rd ed. Oxford University Press, New York.

WHO global tuberculosis control report 2010. Summary. *Cent Eur J Public Health* 2010, **18**(4), 237.

Wolfe, L.M., Mahaffey, S.B., Kruh, N.A. and Dobos, K.M. (2010) Proteomic definition of the cell wall of *Mycobacterium tuberculosis*. *J Proteome Res*, **9**, 5816-5826.

Wolinsky, E., and Rynearson, T.K. (1968) Mycobacteria in soil and their relation to disease-associated strains. Am. Rev. Respir. Dis. **97,**1032–1037.

Zhang, J. and Inouye, M. (2002) MazG, a nucleoside triphosphate pyrophosphohydrolase, interacts with Era, an essential GTPase in Escherichia coli. *J Bacteriol*, **184**, 5323-5329.