Investigating the Rates and Drivers of Drug Resistance in Mycobacterium Tuberculosis

Investigating the Rates and Drivers of Drug Resistance in *Mycobacterium tuberculosis*

A dissertation presented by

Richa Gawande

to

The Committee on Biological Sciences in Public Health

In partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in the subject of
Biological Sciences in Public Health

Harvard University
Cambridge, Massachusetts

April 2015

Professor Sarah M. Fortune Richa Gawande

Investigating the Rates and Drivers of Drug Resistance in *Mycobacterium tuberculosis*

ABSTRACT

The emergence and transmission of drug resistant strains of *Mycobacterium tuberculosis* (Mtb) calls for urgency in understanding the barriers to effective treatment. Here we developed a chemostat cultivation system to measure the mycobacterial mutation rate during periods of slow growth versus fast growth. We found that mutants accumulated at the same rate per unit time during slow growth and during fast growth, suggesting that mutation rate does not vary in accordance with growth rate. By competing genetically barcoded mutant strains at fast and slow growth rates and using a mathematical model to estimate mutant fitness, we confirmed that there is not aggravated loss of mutants at fast growth rates, but that equivalence in mutation rate per unit time reflects time-based mutation rather than replication-based mutation.

To investigate a possible driver of replication-independent mutation, we developed a mass-spectrometry-based analytic method to detect DNA damage during metabolism of fatty acids, a major component of the bacterial diet *in vivo*. We discovered that a novel glyoxylate-dG adduct forms when cells use the glyoxylate shunt to metabolize fatty acids, and that the levels of this adduct increase in the absence of nucleotide excision repair. Using fluctuation analysis to measure the mutation rate in these conditions, we found that fatty acid metabolism is mutagenic in nucleotide repair deficient cells. Finally, cholesterol metabolism was mutagenic in both wild-type and repair-deficient backgrounds. These findings demonstrate that metabolic state can drive both mutation and DNA damage.

Many patients develop recurrent tuberculosis (TB) despite receiving adequate treatment for TB, yet the causes of recurrent disease are poorly understood. We used whole-genome

sequencing (WGS) and MIRU-VNTR typing to investigate the cause of recurrent infection in 13 HIV-infected individuals who had been successfully treated as part of their enrollment in a prospective cohort study in KwaZulu Natal, South Africa. By comparing the genetic relatedness in the presenting and recurrent strains, we found that 7 of the 13 recurrent episodes appeared to represent relapse of the primary infection, despite the force of TB transmission in KwaZulu Natal and the successful HIV and TB treatment history. Results from MIRU-VNTR typing were concordant with WGS for highly related and highly divergent strains, but only WGS analysis could resolve intermediate genetic distances. Patients presented with relapsed infection and reinfection up to 3 years after completion of treatment, and no difference was found between the timing of relapse and reinfection cases. Strains causing relapsed infection were more likely to harbor genetic polymorphisms associated with changes in INH susceptibility or acquisition of INH resistance despite being phenotypically drug sensitive, suggesting a possible role for low-level or undiagnosed drug resistance in tuberculosis relapse.

TABLE OF CONTENTS

DEDICATION

"No mud, no lotus." -Thich Nhat Hanh

This work is dedicated to both mud and lotus.

ACKNOWLEDGEMENTS

I am grateful for the sacrifices of all who have made this journey possible for me. I am extremely grateful to my advisor, Dr. Sarah Fortune, for her encouragement, enthusiasm, wisdom, scientific rigor, and genuine interest in my development as a scientist and as a person. This work would not have been possible without my friends and colleagues in the Fortune Lab, who have provided invaluable scientific and personal support. To my committee members – thank you for the crucial advice and care you have provided throughout this journey. I am grateful to the BPH Program and IID Department for phenomenal programmatic support. To my family– thank you for your encouragement, humor, tradition, and love. To my parents – thank you for giving your whole selves, for your infinite love and sacrifice. Ben – our love has carried me through and is an inspiration; thank you.

LIST OF FIGURES AND TABLES

Chapter Two

Chapter Three

Chapter Four

Appendix 1

CHAPTER ONE: INTRODUCTION

1.1 Scope of the Global Tuberculosis Burden

Tuberculosis is a devastating disease that kills 2 million people per year. The causative agent, *Mycobacterium tuberculosis* (Mtb), has persisted in the human population for tens of thousands of years and continues to evolve strategies to successfully infect and transmit between humans. The modern era of antibiotics is no exception. Mtb strains have evolved resistance to almost every antibiotic in our arsenal: 3.5% of all incident TB cases are multidrug-resistant (MDR), and 5.5% of MDR cases are likely to be extremely drug resistant[1]. Moreover, the co-morbidity of HIV and Mtb has only exacerbated the difficulties in Mtb treatment and necessitates urgency in development of new antibiotics and a better understanding of the barriers to treatment.

1.2 *Mycobacterium tuberculosis* Infection

TB infection occurs when aerosolized Mtb is inhaled and establishes residence in the lung's alveolar macrophages. Infection results in a variable course of disease. Approximately 10% of infected individuals develop active disease, in which rapid bacterial growth causes a high bacterial burden and stimulates recruitment of immune lymphocytes, leading to clinical symptoms. The remaining 90% of infected individuals develop latent infection, characterized by immune control of bacteria, reduced bacterial burden and potentially reduced growth and mutation[2]. There is growing appreciation for the continuum of bacterial growth states during latent infection[3,4,5]. Latently infected individuals are at a 10% risk for reactivation over their lifetime, a probability that increases significantly with HIV infection. Treatment for active infection is designed to suppress the emergence of drug resistance and intended to cause a

sterilizing cure. This regimen and its delivery have been codified over the last twenty years as Directly Observed Therapy – Short Course (DOTS), and involve a multi-drug regimen over 6-9 months, requiring coordination of diagnostic, pharmacological, and community resources. Treatment of drug-resistant strains involves additional regimens whose implementation requires increased resources, supervision and up to two years of treatment. While DOTS has resulted in a 75% increase in case notification rates, successful treatment of 56 million people and has saved 22 million lives, there are still major gaps in our understanding of how to achieve successful treatment[6]. Recurrent infection occurs in 5-20% of successfully treated cases, and incomplete treatment occurs in 10-20% of drug-sensitive cases and 50-80% of drug-resistant cases; thus, drug resistance is a major driver of treatment failure[7,8,9,10,11,12].

1.3 Drug resistance in *Mycobacterium tuberculosis*

In most other bacteria, drug-resistance occurs by both chromosomal SNPs and by horizontal gene transfer (HGT). In mycobacteria, drug resistance occurs only by chromosomal SNPs, due to a lack of molecular machinery necessary for HGT as well as a relatively microbe-poor niche[13]. Efforts to understand the mutation rate in Mtb have been, until recently, limited to *in vitro* studies, yielding estimates of approximately 2.4×10^{-10} per base pair per generation despite the fact that mycobacteria lack mismatch repair, which contributes to several orders of magnitude in genomic fidelity in most other bacteria[14,15,16]. The probability of one drug-resistant cell arising in a mycobacterial population is a product of the SNP rate specific for that drug, and the population size. Even across the wide range of feasible bacterial population sizes in different stages of infection, mathematical models place the rate of acquiring sequential resistance to multiple drugs in a single bacterial cell significantly below the *in vitro* SNP rate, which means that multi-drug resistance should be exceedingly rare[17]. The prevalence of MDR is high; while

most new cases of MDR are assumed to result from transmission to new patients[18], several cases of *de novo* emergence of multi-drug-resistance within individual patients have been documented[19,20,21,22], indicating that the rates and drivers of drug resistance are poorly understood.

Recently, there has been a growing appreciation for bacterial variability at both the genetic and phenotypic levels that can contribute to a cell's ability to survive antibiotic treatment. Since the advent of whole-genome sequencing technology, the genetic sources of variability are particularly ripe for investigation. Recent efforts to determine the degree of inter- and intra-patient diversity during infection suggest that the ability of Mtb to acquire genetic diversity *in vivo* is greater and more complex than previously appreciated.

In the last decade, whole-genome sequencing of clinical strains has revealed at a high-resolution the extent of global diversity across the 5 phylogeographically-defined lineages of Mtb[23,24]. Although there is restricted genetic diversity across the global population of Mtb compared with other human pathogens[25], there are genetic and phenotypic sources of variability that have been associated with considerable variability in drug resistance outcomes. In particular, many strains in the East Asian lineage have been associated with high rates of clinical drug resistance and more severe disease[26,27,28]. Whole genome sequencing has made it possible to study the genetic basis of complex phenotypic traits such as drug resistance. Strains from this lineage were recently shown to have higher numbers of drug-resistance-associated genetic polymorphisms and have been experimentally associated with higher *in vitro* mutation rates, suggesting that these strains may be biologically predisposed to drug resistance in part because of higher mutation rates[14].

1.4 Genetic diversity *in vivo*

It is largely unknown during which stages of infection, and at what rate, drug resistance mutations are acquired *in vivo*. A macaque model of Mtb was developed to mimic the variable course of human Mtb infection[29]. In this model, whole-genome sequencing was applied to estimate the SNP rate over the course of infection. Mutation rate was expressed as a function of time rather than generation, because it was unknown how many generations had elapsed during infection. The number of SNPs that accumulated in individuals with latent, active, and reactivated disease over the same amount of time was similar, suggesting a similar mutation rate per day in all three disease states[30]. Groups have taken a similar approach to measure the accumulation of intra- and inter-patient diversity in human transmission chains. Gardy *et al* used whole genome sequencing to identify two simultaneous Mtb outbreaks within the same community in British Columbia[31]. Further analysis by Ford *et al* revealed that rate of accumulation of genetic diversity in outbreak strains compared with historical isolates from the same area was similar to the estimates in the previous *in vitro* and macaque models[14]. In larger analyses Walker *et al.* and Bryant *et al* showed a similar rate of mutation in within-patient isolate pairs and isolate pairs from epidemiologically linked clusters[32,33]. These studies introduced the idea that mutations accumulate at similar rates *in vitro* as during different disease states *in vivo*. Bryant *et al* further showed that the inclusion of drug-resistance mutations obscures the ability to estimate the mutation rate, due to potentially different selective pressures operating in antibiotic resistant and sensitive strains, suggesting that issues of fitness and sampling need to be further investigated to achieve more sensitive estimates of *in vivo* mutation rates.

Mutations could accumulate at similar rates in different disease states either because the Mtb replication time during latency is similar to or faster than the *in vitro* replication rate, or

actively replicating populations of bacteria are being selectively sampled. As it is unknown to what extent bacteria in sputum capture the entire "cloud" of genetic diversity within the individual, it is possible that there are populations of bacteria that are not represented within the sputum[34, 4]. However, all studies attempting to estimate replication time during latency support a model of slower growth and metabolism[35,36,37,38]. The idea that slow growing bacteria could achieve the same amount of genetic diversity as rapidly growing bacteria runs counter to the prevailing replication-centric understanding of mutation: replication errors are the primary source of mutation, and therefore only actively replicating bacteria are capable of developing the mutations that cause drug resistance[15]. An alternative explanation is that accumulation of mutations may be driven by mechanisms other than errors in replication. The spectrum of mutations for both the macaque and human analyses shows enrichment for products of oxidative damage, suggesting that DNA damage may be a source of mutations *in vivo*[30]. However, the contribution of DNA damage to mutations is largely unexplored in mycobacteria. Here, we are interested in uncovering mechanisms of mutation that are independent of DNA replication. Given the importance of latency in the *Mtb* life cycle, acquiring genetic diversity during periods of slow growth may be an important evolutionary adaptation.

1.5 DNA Damage and Repair During Infection

The main sources of chemically reactive species during Mtb infection are inducible nitric oxide (iNOS) and superoxide radicals produced by activated macrophages[39,40,41]. iNOS-/- mice are severely attenuated during Mtb infection, suggesting that oxidative and nitrosative damage is a critical defense against Mtb[42]. Reactive species can damage all intracellular macromolecules, including DNA, lipids, and proteins. Studies on the transcriptional response during macrophage and mouse infections show induction of a wide range of bacterial defenses against oxidative

stress including antioxidants and scavenging enzymes, as well as base excision repair (BER) and nucleotide excision repair (NER), the main pathways for DNA damage repair in mycobacteria[43,44,45,46,47].

Base excision repair (BER) plays a central role in repair of lesions created by reactive species[48]. Work in *Mycobacerium smegmatis* (Msm), a non-pathogenic soil bacterium whose BER machinery is homologous to the BER machinery found in Mtb, revealed that BER enzymes repair mutagenic DNA damage even in the absence of the immune pressure. Individual deletions of key BER glycosylases resulted in defective repair of *in vitro* DNA damage substrates, an increase in mutation frequency during basal growth conditions, and sensitivity to oxidative stress. Mycobacterial BER is unique compared with the E. coli counterpart in that there are additional DNA glycosylases, differing substrate activities of BER enzymes, four separate genes encoding *mutT* oxidized nucleotide cleansing enzymes (8-oxodGTPase: compared to one in *E. coli*)[49,50,51,52] and a high level of redundancy among BER components[53], suggesting that this pathway may play a unique role in mycobacteria. While mismatch repair is a primary component of damage-induced mutation repair in other organisms, Mtb may use NER to achieve this task[54,55]. While NER is best known for its role in repair of UV damage, it also plays a role in protecting against a variety of insults. NER is achieved by the coordinated actions of UvrA, B, C, D1, and D2. It is largely unclear which damage products this system is responsible for repairing during infection[56,57,58].

If DNA damage is a considerable source of genotoxicity during infection, strains deficient in DNA damage repair should be attenuated during infection. Genetic essentiality screens have revealed that components of both NER and BER are essential for growth during different stages of infection. Specifically, UvrB was required for growth in the mouse and

macaque model of infection, and during growth on acidified nitrite, a condition designed to recapitulate the intracellular environment of the macrophage[59]. UvrD1 was required for growth in both the mouse and macrophage models of infection[60]. Components of BER such as the uracil glycosylase, Ung, are required for growth in *in vitro* models designed to recapitulate certain features of the *in vivo* environment, such as exposure to hypoxia, cholesterol, and starvation-induced slow growth[61,62,63,64]. Together with results from studies examining the role of other replication and repair enzymes such as recA and the stress-induced error prone polymerase, DnaE2, these findings revealed that there are differential requirements for replication and repair machinery in different environments, suggesting that the selections imposed, and/or the bacterium's capacity for lesion repair may not be equal in each environment. Thus, there may be differential damage and mutational spectra for each of these host environments. In particular, the attenuated survival of the Ung transposon mutant during growth in cholesterol suggests that DNA damage may be a bigger stress when carbon is limited[65], yet nothing is known about the direct genotoxic effects of Mtb metabolism.

Damaged nucleotides can result in mutagenic pairing or stalled replication and transcription, causing stress-induced error-prone replication[66,67,68]. The high genomic GC content makes Mtb disproportionately vulnerable to guanine damage, as guanine is the most susceptible base to oxidation because it has the lowest redox potential of the four bases. In the presence of oxidative stress, hydroxyl radical can react with the guanine base, forming the oxidized adduct 8-hydroxy-deoxyguanosine (8-oxoG) or reduced to FAPy-guanine, depending on the redox potential of the environment. 8-oxoG is one of the most common DNA damage species[67]. The mutation spectrum of the macaque SNP data suggests that 8-oxoG formation and cytosine deamination into uracil may be specific drivers of mutation in *in vivo* environments[30].

Yet, whether a mutation arises from a DNA lesion depends on the integrated actions of the oxidative environment and the bacterium's capacity to repair the lesion. Little is known about the relationship between damage, repair, and mutation, though given the rich oxidative environment of the TB granuloma and the importance of DNA damage repair enzymes during infection, DNA damage is a primary candidate for sources of replication-independent mutation.

1.6 Summary of Aims

In this dissertation I have investigated the factors driving the evolution of drug resistance in Mtb. Application of whole-genome sequencing technology combined with the power of prospective cohorts has allowed us to ask more sophisticated questions about the evolution of drug resistance in *in vivo* and in *in vitro* models designed to mimic aspects of the *in vivo* environment.

In **Chapter 2,** we tested the hypothesis that mutations can accumulate independent of growth rate using chemostat cultivation to precisely regulate growth rate *in vitro*. The finding that mutations accumulate at similar rates during stages of infection in which growth rate was thought to be very different[30,31,32,33] suggests that mutation rate and growth rate may be uncoupled *in vivo*. Using the mutant accumulation assay in *Mycobacterium smegmatis* as a model for mutant accumulation in Mtb, mutants were found to accumulate at similar rates per day across fast and slow growth rates. The chemostat model combined with genetic barcoding allowed us to explore the effects of selection on accumulation of genetic diversity, a feature that has challenged previous studies investigating *in vivo* mutation rates. We could not attribute the above finding to a significantly different selective pressure in fast and slow growth rates. The ability of Mtb to acquire mutations during slow growth in the chemostat and during periods of

presumed slow replication in the host suggests a model of replication-independent mutation (Ford *et al.,* 2011).

In **Chapter 3,** we tested the hypothesis that mutation accumulation can be driven by metabolic states that bacteria encounter in the host. We found that fatty acid metabolism is genotoxic, inducing a previously unknown DNA adduct. Adduct levels increased in the absence of nucleotide-excision repair, suggesting a role for this pathway in fatty acid metabolism. Consistent with the mutagenic effect of this adduct, fatty acid metabolism was found to be mutagenic in the absence of nucleotide-excision repair. This work demonstrates metabolically induced genotoxicity and mutagenesis, and highlights the importance of studying the evolution of drug resistance mutations in conditions that mimic the *in vivo* environment.

In **Chapter 4**, we studied the cause of recurrent infection in a prospective cohort of HIV-infected individuals living in a high TB-burden area of South Africa. The causes of recurrent infection may be multiple, including drug-resistance of the infecting strain, reinfection by a new strain, poor adherence, and/or inappropriate drug regimens. We found that over half of recurrent infections represent true relapses of an earlier infection despite completion of successful treatment. Despite being phenotypically drug-sensitive, the strains associated with relapse had an enrichment of drug-resistance associated polymorphisms, suggesting that undiagnosed low-level drug-resistance may biologically predispose these strains for survival.

Together, the work presented in this dissertation reveals previously unappreciated capacity for mutation during *in vivo* conditions and reveal undiagnosed drug-resistance as a potential driver of TB relapse.

CHAPTER TWO

Use of the Chemostat to Investigate Replication-Independent Mutation in Mycobacteria

Richa Gawande[1], Nicolai Panikov[1], Aditya Bandekar[2], Jeremy Rock[1], Constance Martin[1], Jeffrey Wagner[1], Chris Sassetti[2], Sarah Fortune[1]

[1]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA
[2]Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, Worcester, MA

Author contributions R.G. and N.P., and S.F. designed the experiments. **R.G. and N.P.** performed the experiments, mathematical modeling, and data analysis (**R.G.:** construction of barcoded strains, chemostat sampling and plating, qPCR experiments, adaptation of mathematical model to MATLAB, optimization of parameters and data fitting; **N.P.:** methodological development and maintenance of chemostat cultivation, development of chemostat media, empirical determination of growth kinetics, chemostat sampling and plating, development of mathematical model). **A.B and C.S.** performed preliminary experiments and data analysis and provided valuable discussion. **J.R.** provided strains and valuable discussion. **C.M.** designed and optimized the barcoding strategy and qPCR primers. **J.W.** adapted the mathematical model to MATLAB and provided valuable discussion. **R.G. and S.F.** drafted the manuscript.

2.1 Abstract

The rapid emergence of drug resistant strains of *Mycobacterium tuberculosis* requires a better understanding of how drug resistance evolves during infection. Recent findings suggest that mycobacteria can acquire mutations at the same rate per day during active and latent disease as during *in vitro* growth. Yet, the replication rate of Mtb is presumed to be faster in active disease than in latent disease, and still faster during *in vitro* growth, raising the question of whether mycobacteria have a similar capacity to generate genetic diversity across multiple growth rates. To test this hypothesis, we developed a chemostat cultivation system to control the bacterial growth rate and performed a mutant accumulation assay at fast and slow growth rates. Mutants were found to accumulate at similar rates per day at fast and slow growth rates, consistent with replication-independent mutation. This observation can also be explained by replication-dependent selection, where relative fitness of mutant strains compared to wild-type is growth-rate dependent. To distinguish between these two hypotheses, we competed genetically barcoded mutant strains with wild-type strains at fast and slow growth rates. When we applied a mathematical model to fit the resulting growth kinetics, mutant growth parameters were found to be similar relative to wild-type at fast versus slow growth rates. Together, our findings suggest that that the mutant accumulation rate is similar at fast versus slow growth rates, supporting a time-based model of mutation, consistent with the preceding *in vivo* studies. These findings challenge the current replication-dependent understanding of the evolution of drug resistance and provide a new framework for molecular mechanisms underlying genomic fidelity in mycobacteria.

2.2 Introduction

The rapid emergence and global spread of drug resistant TB calls for a better understanding of the rates and drivers of drug-resistance in conditions relevant to infection. Multi-drug resistance is emerging in up to 20% of previously treated TB cases[69]. The prevailing view is that the mutations driving TB drug resistance, which are chromosomally encoded in the Mycobacterium *tuberculosis* (Mtb) genome, arise because of spontaneous errors in DNA replication in populations of actively dividing mycobacteria[15]. From this, the capacity for mycobacteria to acquire drug resistance is thought to be reduced during periods of slow or no growth. Recent findings suggest that Mtb's mutational capacity over time may be similar in states where the replication rate is thought to be different, from studies using whole-genome sequencing in human transmission chains and in the macaque model of TB[30,31,32,33]. These findings point to the possibility that mycobacterial populations could acquire mutations as a result of processes other than errors made during chromosomal replication. This could account for the equivalent rate of accumulation of mutations during latent disease seen in macaques[30]. These studies are complicated by the fact that little is known about the growth rate during different disease states, though estimates from experimental models of infection vary from 45 to 135 hours per generation[36,70,71].

It is important to understand whether the mutational capacity is reduced during latency, as 90% of individuals are latently infected, this disease state underlies this organism's success. Moreover, it is clinical practice to treat individuals who have latent TB infection with INH monotherapy based on the limited evidence that latent bacteria in humans have a reduced capacity for growth and therefore mutation. Indeed, where it has been studied, individuals with latent TB infection treated with INH prophylaxis do not have higher drug-resistance rates[72,73].

However, it is unknown whether these patients harbor higher numbers of drug-resistant organisms, which could potentially be transmitted or selected for in the face of treatment.

There is a precedent in experimental models of slow-growing *E. coli* for relatively rapid mutation accumulation during slow growth, where chemostat cultures of *E. coli* maintained at different growth rates acquired phage-resistance at similar rates per day[74]. There is also precedent for higher numbers of drug-resistant bacteria accumulating in aging colonies, and this increased rate of accumulation has been attributed to either an increased mutation rate during starvation or an increased fitness of mutants during starvation. Stress-induced mutagenesis has been recognized as a source of mutations in *E. coli*, yeast, and human cells. In *E. coli*, recruitment of error-prone polymerases to repair double-strand breaks during a stress condition has been shown to be a mechanism of stress-induced mutagenesis[75]. In *P. putida*, stress-induced DNA damage is mutagenic in the absence of DNA damage repair during starvation, but not in growing cultures, again showing a condition-dependent induction of mutation[76]. While little is understood about these processes in mycobacteria, these findings point to a possible source of mutations that arise independently of errors in chromosomal replication, and could help explain a possible mechanism driving mutations in states where the bacteria are thought to be dividing at different rates.

We investigated the dynamics of mutant accumulation during periods of rapid and slow growth for the soil-dwelling mycobacterium, *Mycobacterium smegmatis* (MSm), whose DNA replication and repair pathways have a high degree of homology to those in Mtb[77]. To control growth rate, we developed a chemostat cultivation system, which allowed for maintenance of MSm cultures in a defined state for many generations at different growth rates. Mutants were tracked over time, and mutation rate was estimated by the mutant accumulation assay[78]. As

compared to other methods to measure mutation rate such as fluctuation analysis or whole-genome sequencing, chemostat cultivation allows maintenance of a large enough number of cells for enough generations such that the probability of spontaneous mutations arising is 1 for every generation even at very low basal mutation rates, and mutant accumulation can be tracked over a relatively short period of time. Growth rate can be altered by altering dilution rate. If mutants arise only due to replication errors, mutant accumulation per day should vary proportional to replication rate, resulting in a higher mutant accumulation per day with higher growth rates. We compared the rate of accumulation of antibiotic-resistant mutants at fast and slow growth rates to determine whether mutation is entirely growth-rate dependent. On the contrary, we found that the growth kinetics of mutants was similar across slow and fast growth rates, and this could not be explained by differential loss of mutant strains at fast growth rates, suggesting that the mutation rate per unit time was the same.

2.3 Results

Chemostat cultivation of M. *smegmatis*

We established a chemostat cultivation system for MSm, shown in Figure 2.1. A log-phase batch culture of MSm mc^2155 was used to inoculate the chemostat vessel, which was maintained at 37 degrees with constant stirring. A steady drip of fresh medium was pumped in at *d* the dilution rate. An overflow tube set the volume of the culture in the vessel, through which the culture was pumped out at the same rate that fresh medium was pumped in. Cultures were fed a chemically-defined minimal medium, limited by glycerol as the only carbon source. After a period of exponential growth during which the initial glycerol was consumed, the cultures reached steady state, characterized by a steady biomass and CO_2 output. At this point, glycerol

became limiting in the growth vessel and growth rate was approximately equal to *d*, the dilution rate.

MSm mc^{2}155 was grown under glycerol-limited conditions at four different growth rates (.015, .028, .11, .23, corresponding to doubling time = 46 h, 24 h, 6 h, and 3 h, respectively). Temperature, CO_2, pH, and optical density were monitored regularly, and cultures were sampled daily. We established the upper limit of growth of MSm mc^{2}155 to be 2.7 hours per generation and the lower limit to be greater than 60 hours per generation. Steady state was defined by steady CO_2 output, OD, and CFU.

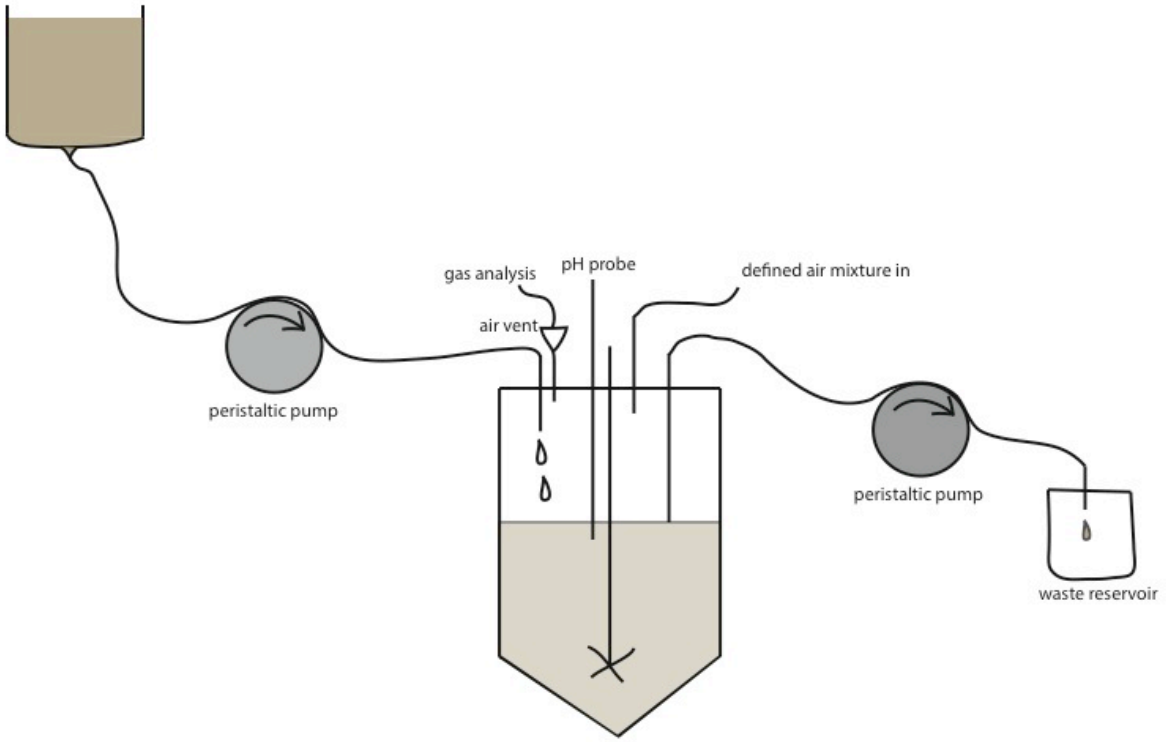


Figure 2.1 Schematic of chemostat cultivation in *M. smegmatis*. Glycerol-limited continuous cultures of *M. smegmatis* were grown in 1-liter bioreactors fitted with a sealed top plate. Inflow media was contained in a 10L reservoir and addition of nutrients to chemostat cultures was controlled by a peristaltic pump. Sample outflow was regulated using an overflow tube fitted to another pump. Culture was stirred constantly, temperature was regulated using a heating blanket set to 37 degrees, and temperature, pH, and dissolved O_2 measurements were automated. A defined air mixture was provided through a sterile filter in the top plate. Effluent gas was passed through a cooling condenser and CO_2 gas was analyzed continuously.

Tracking mutant accumulation in the chemostat

In order to determine the relationship between growth rate and mutation rate in MSm, we tracked concentration of antibiotic resistant mutants in steady-state chemostat cultures grown at four different growth rates. Chemostat samples were plated daily on 200ug/mL of rifampicin and 2ug/mL of ciprofloxacin once chemostats had reached steady state. Inherent in the mutant accumulation assay is the presumption that mutants will accumulate. We defined the period of time during which mutants accumulated and defined the rate of increase for each growth rate. The overall pattern of mutant accumulation in the three growth rates was similar for both Rif and Cip mutants, when expressed as maximum mutant frequency, distribution of mutant frequencies, or slope of a linear regression fit to each curve (Figure 2.2). When these data are plotted as a function of generations, the approximate per-generation rate of mutant accumulation varies with growth rate (Figure 2.3a). Plotting these data as a function of time shows an approximately equal accumulation rate, $\sim 1 \times 10^{-9}$ mutations per cell per day, for all three growth rates. This is despite the growth rates varying by 7-15 fold (Figure 2.3b), suggesting that the rate of MSm mutant accumulation can be uncoupled from its growth rate.

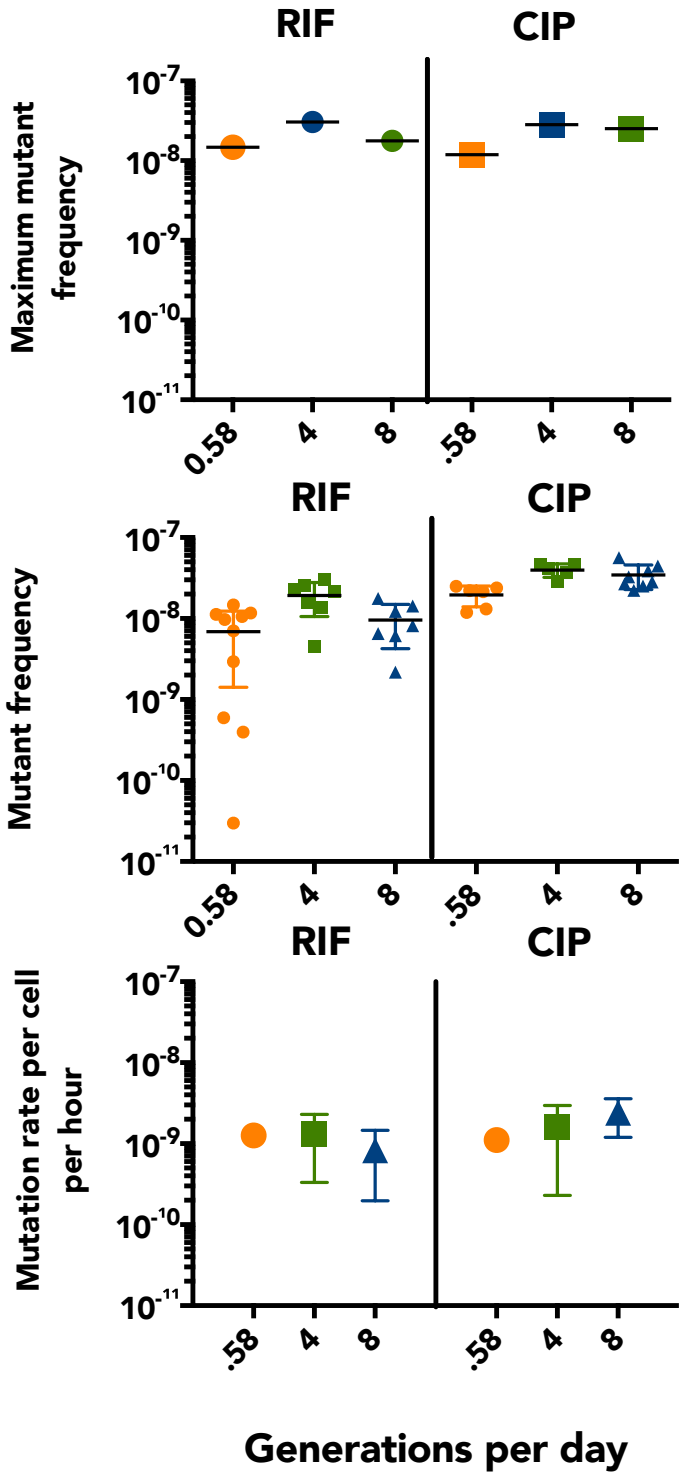


Figure 2.2 Mutants accumulate at similar rates per-day across multiple growth rates. Top panel: Maximum mutant frequency; Middle panel: distribution of mutant frequencies; Bottom panel: the slope of a linear regression fit to the mutant increase, for Rif and Cip mutants, for D = .015, .11, and .23, corresponding to *td* = 46 h, 6 h, and 3 h, respectively).

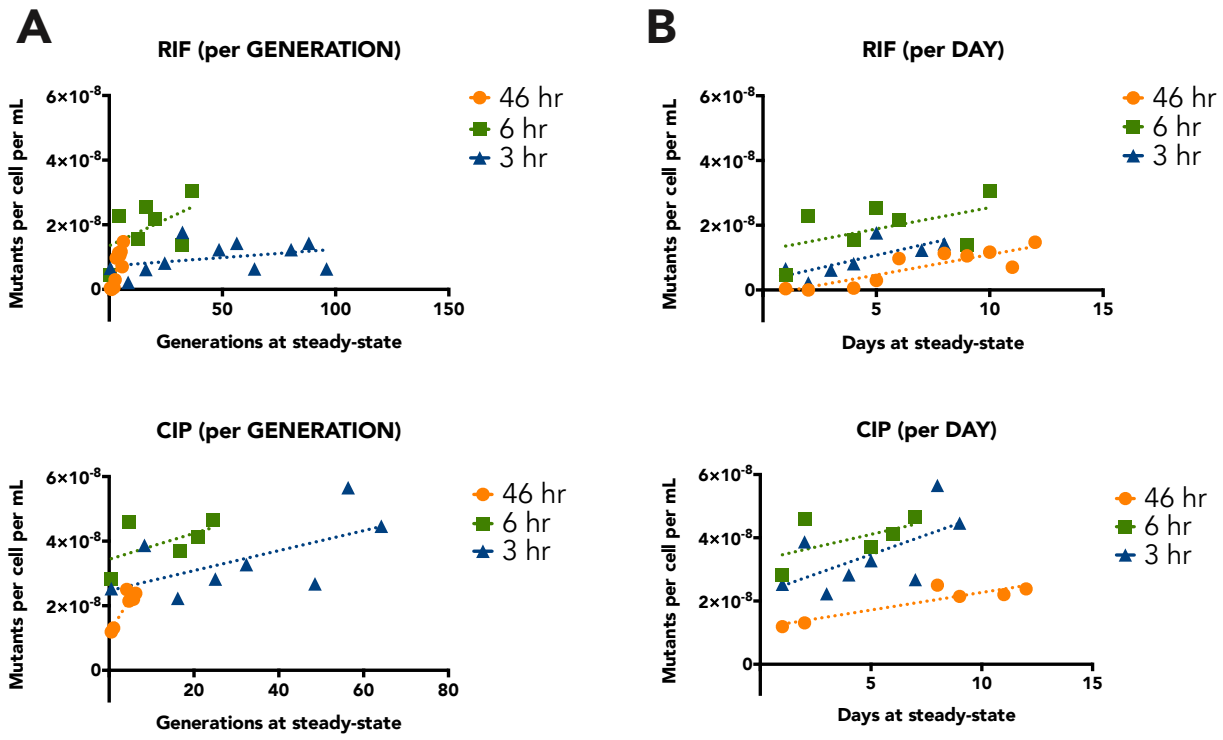


Figure 2.3 The mutant accumulation rate is time-dependent, not replication-dependent. Rate of mutant increase plotted as a function of a) generation and b) time for Rif mutants (top) and Cip mutants (bottom). Points represent actual data; dotted lines represent linear regression fits. Data are shown for mutant accumulation for chemostats run at doubling times of 46 hours, 6 hours, and 3 hours. Y-axis is mutant concentration, x-axis is number of days at steady state.

The role of selection on mutant growth kinetics

According to chemostat theory, if mutants have approximately the same growth rate as the parent, the density of the mutant strain is expected to increase linearly with growth[74,78]. When we extended the time series, we did not observe a continued increase in mutant frequency; instead the mutant population collapsed, highlighting the problem of competition within the chemostat (Figure 2.4). These dynamics raise the possibility that there are differential fitness effects of Rif and Cip resistance at different growth rates. If mutant fitness were growth-rate dependent, this would support an alternative interpretation of the above data: rather than growth-rate-independent mutant accumulation, both mutant accumulation and mutant wash-out could be growth-rate dependent. This could result in apparent growth-rate-independent mutant accumulation due to a disproportionately higher rate of both mutation and mutant loss at fast growth rates. While *rpoB* (target of rifampicin) mutants have been shown to have fitness defects in batch culture experiments, little is known about the fitness of *rpoB* mutants in different growth conditions. We thus investigated whether mutant fitness is growth-rate dependent in the chemostat.

One feature of decreased fitness may be a decrease in the mutant population's genetic diversity due to reduced survival of less fit mutants[79,83]. To determine whether the diversity of mutants changed with growth rate, we focused on Rif-resistant mutants and sequenced the *rpoB* Rif-resistance determining region in ~50 Rif-resistant mutants isolated across time from chemostats run at slow (t_g = 46 hr) and fast (t_g = 3 hr) growth rates (Figure 2.5). The overall *rpoB* target size was similar at the two growth rates, 12 nucleotides for the slow growth rate and 10 nucleotides for the fast growth rate, consistent with previous *in vitro* estimates of rpoB target

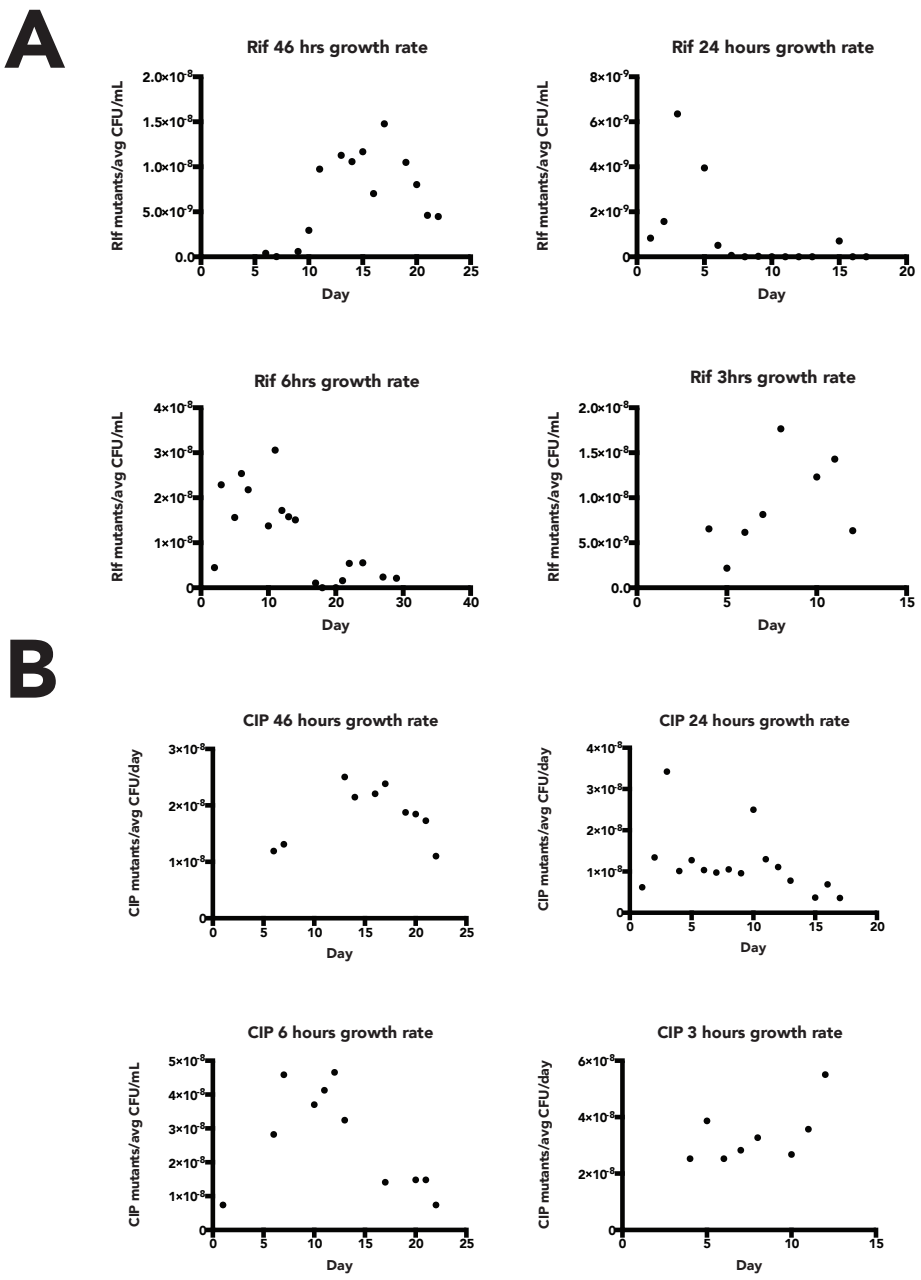


Figure 2.4 Mutant populations collapse after an initial increase. Accumulation of antibiotic-resistant mutants at steady state for a) Rif and b) Cip, for four growth rates (D = .015, .028, .11, .23, corresponding to *td* = 46 h, 24 h, 6 h, and 3 h, respectively).

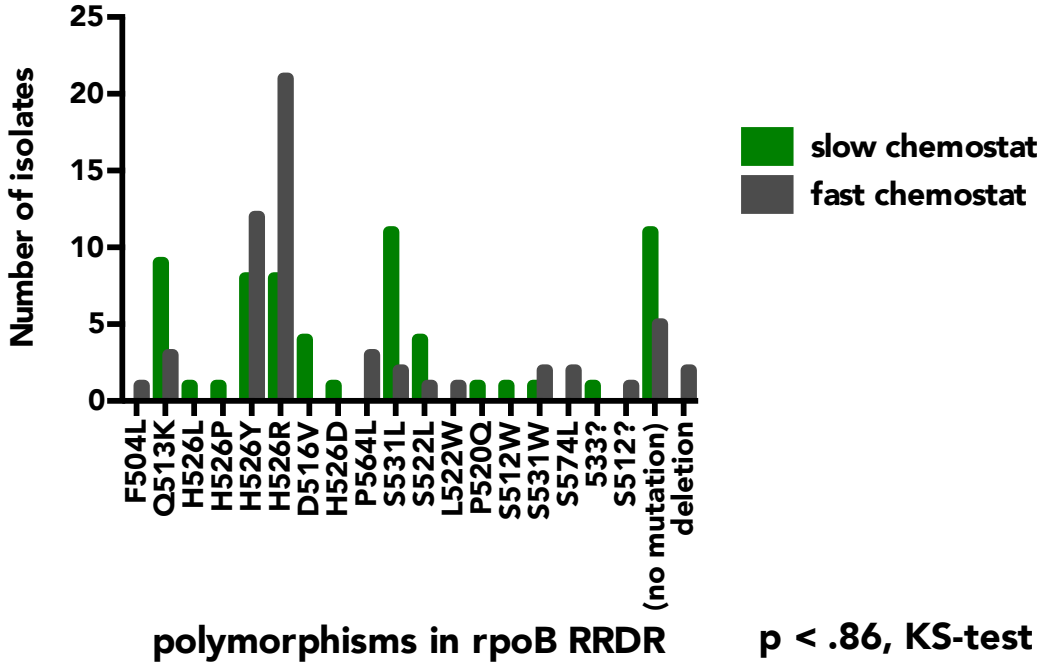


Figure 2.5 rpoB target size is not significantly different for RIF-resistant isolates in fast versus slow chemostats. The Rif-resistance determining region (RRDR) was PCR-amplified and Sanger-sequenced for approximately 60 isolates over time for the fast and slow chemostats. There were also no significant differences in the temporal distribution of these isolates between fast and slow chemostats (not shown).

size[14], suggesting that selective pressure in the fast chemostat does not significantly reduce the genetic diversity compared to the slow chemostat. There were also no major differences in the temporal distribution of these isolates across growth rates, though the sample size was small for each mutant type (data not shown). There were differences in the representation of each mutant at the two growth rates, including in the group of mutants that did not harbor genetic mutations in the rpoB RRDR. These mutants may harbor mutations elsewhere in rpoB, or in another gene.

While *rpoB* sequencing allows an estimate of the breadth of genetic diversity at the two growth rates, it is difficult to gain temporal resolution with this sample size. To gain higher temporal resolution into the rate of mutant wash-out at the two growth rates, we performed a competition experiment in which we competed abundant and rare *rpoB* RRDR mutants and wild-type strains isolated from the fast growth rate (d = 3hr) and slow growth rate (d = 46hr) chemostats. Mutants were competed against wild-type MSm at fast and slow growth rate to determine whether selection against mutants increased with growth rate. Each strain was tagged with a unique genetic barcode and a common constant region, allowing for strain tracking, enumeration of total genome count, and normalization across samples by quantitative PCR. Data are plotted in Figure 2.6. The top panel shows competition among 7 rpoB mutants and three wild-type MSm strains at fast and slow growth rates. All rpoB mutants showed a competitive disadvantage compared to wild-type MSm at both growth rates, though not all mutant strains competed equally. At both growth rates, mutants harboring the H526Y and S522L mutations had a growth advantage compared to the other mutants, and additionally, strain S512W appeared to have a growth advantage at the slow growth rate but not the fast growth rate. Not all wild-type strains competed equally. At the slow growth rate, two wild-type strains out-competed the rest, while at fast growth rate, all three wild-type strains persisted, with one dominant strain and

two intermediate strains. When the same seven mutants plus three additional mutants were competed in the absence of wild-type strains (bottom panel), the same three dominant mutants (carrying the H526Y, S512W, and S522L mutations) plus the mutant with the Q513K mutation out-competed the rest at both the slow and fast growth rates. Three of these four strains had been previously adapted to a slow growth rate chemostat, while the Q513K mutant had been previously adapted to a fast growth rate chemostat.

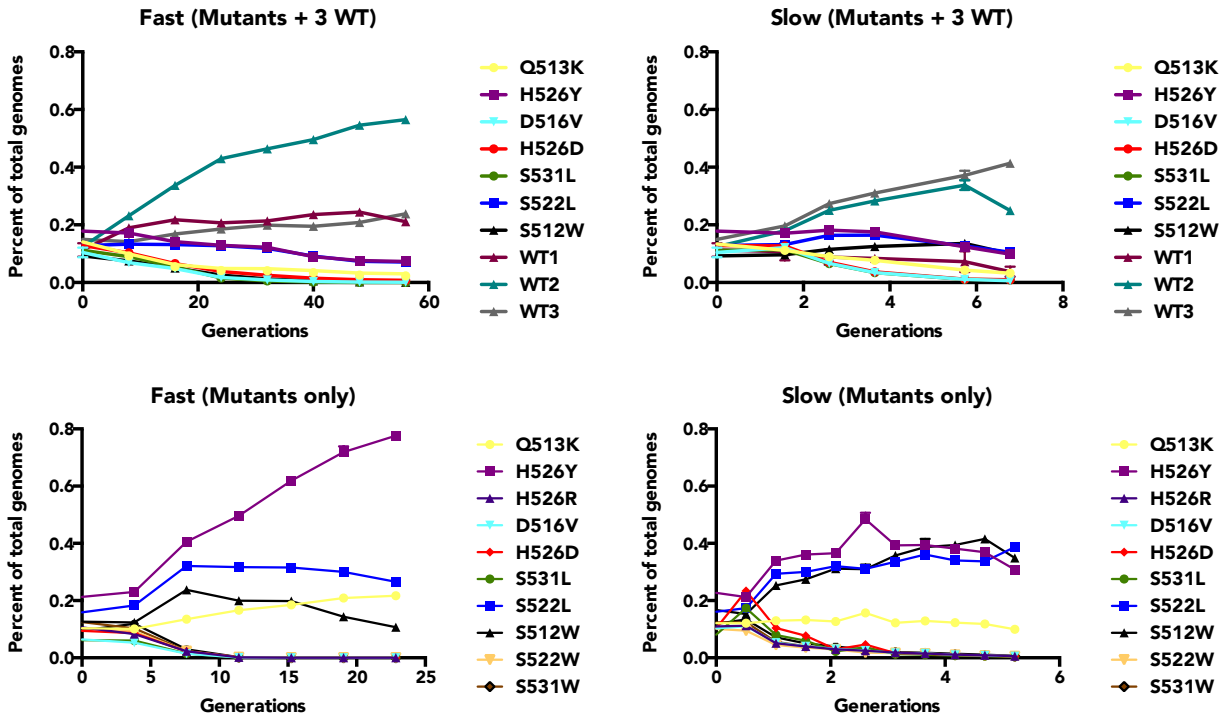


Figure 2.6 Competition among rpoB mutants and wild-type strains is similar at slow and fast growth rates. Top panel: 7 rpoB mutants and 3 wild-type strains were competed at fast a) and slow b) growth rates. 10 rpoB mutants were competed at fast c) and slow d) growth rates. Data are plotted as a percent of the total genomes present at that time-point. Y-axis is biomass in mg of the mutant strain.

The fitness of *rpoB* mutants appeared to be qualitatively similar between the fast and slow growth rates. In order to make a quantitative estimate of the fitness parameters for each strain, we simulated these competition dynamics using the Monod model of bacterial growth in the chemostat. In the Monod model, growth of each strain depends on the dilution rate, d, glycerol concentration, S, glycerol consumption K_s (the Monod constant), and μ_{max}, the maximum growth rate (Figure 2.7)[74,80,81]. Using the concentration of each strain in the inoculum as the initial conditions for the model, we approximated K_s and μ_{max} for each strain, at low and high dilution rates. The simulated competition experiment is plotted along with the actual experimental data in Figure 2.8a and b. Raw best-fit parameters are listed in the Methods, and data normalized to the most-fit strain are plotted in 2.8c and d. Consistent with the visual analysis, the competitive advantage of mutants appeared to be overall similar between the two growth rates. The mutant carrying the S512W mutation and the WT3 strain were both slightly more fit relative to the other strains at the slow growth rate compared to the fast growth rate. All other strains were more fit at the fast growth rate than the slow growth rate. In fact, the baseline fitness was higher for almost all strains at the fast growth rate than the slow growth rate. We concluded from these experiments that the fast growth rate does not impose a significantly more severe selection on mutants than the slow growth rate.

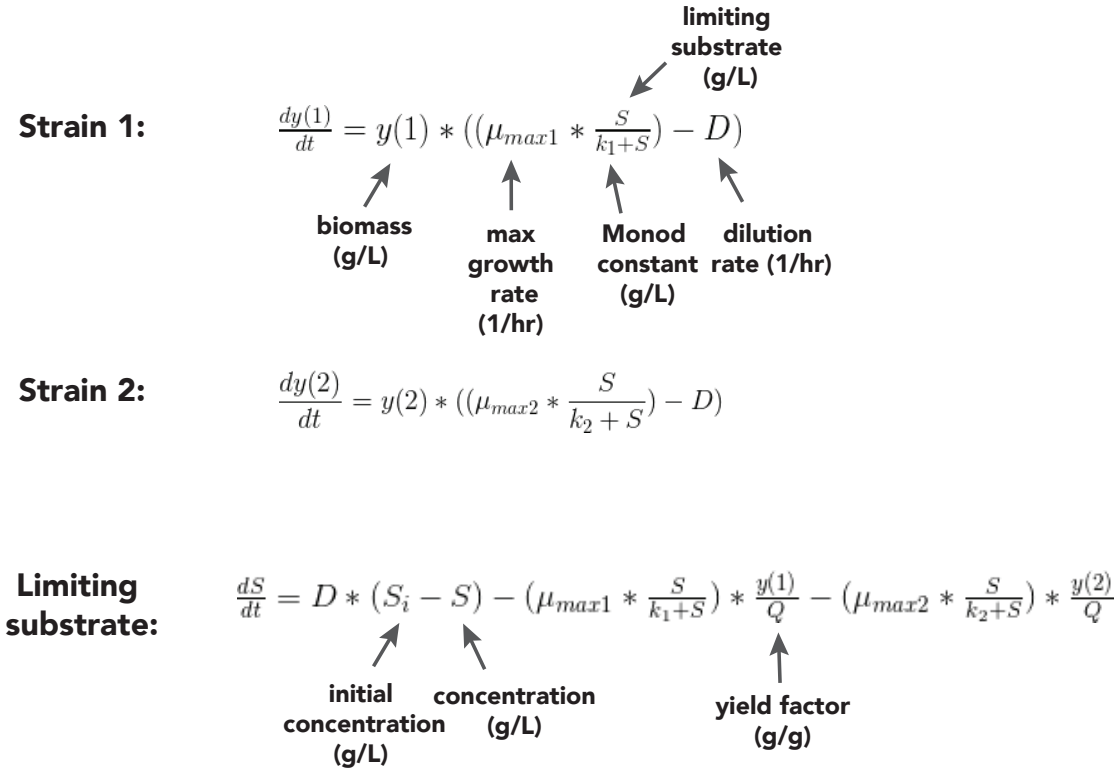


Figure 2.7 The Monod model of bacterial growth in the chemostat. The above two-strain model was adapted for 11 strains to simulate inter-strain competition in the chemostat. Parameters for S, S_i, Q, and initial biomass conditions were chosen based on empirical measurements; a parameter range was chosen for μ_{max}, k_s, and then optimized to fit the empirical data.

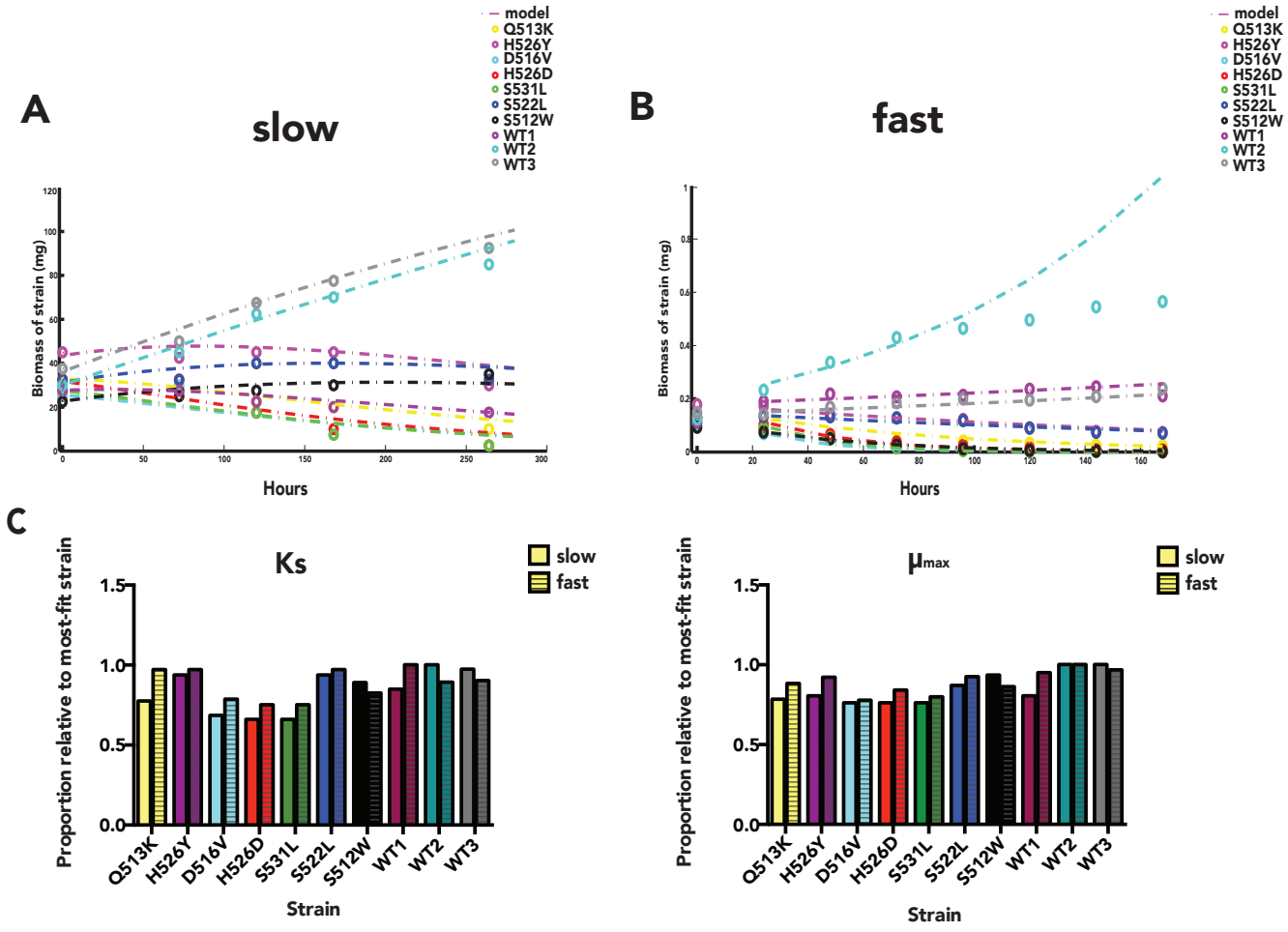


Figure 2.8 Growth parameters of mutant and wild-type strains as simulated in the Monod model are similar at fast and slow growth rates. Top panel: simulation of 7 rpoB mutants and 3 wild-type strains at a) slow and b) fast growth rate. Bottom panel: Best-fit ks and mu-max for each strain at each growth rate. Y-axis is biomass of each strain in mg.

Together, our findings show that the mutant accumulation and wash-out rates are similar across growth rates. These data are in support of a replication-independent relationship between mutation rate and growth rate.

2.4 Discussion

Here we used chemostat cultivation to study mutant accumulation in MSm at multiple growth-rates. We found that mutants accumulated at a similar rate over time in both fast and slow growth rates. These findings bring up two main considerations.

Balance between mutation and selection in the chemostat: We were interested in comparing the rate of evolution of antibiotic resistant mutants at multiple growth rates. The relative fitness of these mutants is continually changing over time based on the rate of "background" evolution in the chemostat. The chemostat population is under constant selection to consume the limiting substrate at a rate fast enough to keep up with the dilution rate. Therefore, affinity for the limiting substrate is a major driver of evolution the chemostat, and may explain why we observed a different competitive advantage among the three wild-type strains in Figure 2.7. Previous work in E. coli chemostat cultures where glucose was the limiting substrate, showed that the overall affinity of the population for glucose increased significantly over time, and this was accompanied by selective sweeps of strains that had faster and faster growth rates in this medium[82]. We found that the rate of mutant accumulation, reflecting a balance of mutation and selection, was similar across growth rates. Using Rif-resistant mutants to explore this balance further, we found that mutants had decreased fitness compared to a number of wild-type strains, consistent with previous findings[83,84]. However, we did not find a significant decrease in their relative fitness at fast growth compared with slow growth,

suggesting that the similar rates of accumulation may reflect similar rates of mutation. In order to better track the accumulation and loss of mutants, the mutant accumulation assay could be performed with a library of uniquely barcoded strains. A library with diversity of 150,000 unique barcodes could be expanded by 10X to produce enough inoculum for a small-volume chemostat. This library could be plated on Rif, followed by Sanger sequencing of the barcode to identify the starting mutant set. Though the probability of drawing the same barcode twice is high with this much expansion, the probability of getting the same Rif mutation in two identical barcodes is exceedingly low. Use of a counter-selectable marker would purge pre-existing mutants. Once chemostat cultivation begins, samples could be regularly plated on Rif and target sequencing of resistant mutants could be performed to track appearance of new mutants versus expansion of existing mutants.

Contribution of replication-dependent and replication-independent mutation: Our mutant accumulation data could be explained by a combination of both replication-dependent and replication-independent mutational processes. Investigations into the mediators of genomic fidelity in mycobacteria suggest that both replication-dependent processes such as DNA proofreading, and replication-independent processes such as DNA damage repair and error-prone polymerases, contribute to the mutation rate[85]. Little is known about the relative contribution of these processes during infection, though genetic essentiality studies in experimental models suggest a role for DNA damage repair, and whole-genome sequencing of clinical isolates shows polymorphisms in these areas (unpublished work, our lab). In particular, oxidative damage may be one primary driver of DNA damage during infection[39,40,41,42,43,44]. The spectrum of mutations in Mtb populations isolated from macaque lesions points to an oxidative damage signature, consistent with the observations that Mtb is exposed to oxidative radicals during infection[30].

There is also a precedent in the literature for stress-induced mutagenesis, though little is known about this in mycobacteria[75]. Studies of aging colonies in E. coli suggest that the proportion of Rif-resistant mutants increases as a colony ages, but like mutant accumulation in the chemostat, this rate reflects a balance of mutation and selection – studies conflict as to the degree of contribution from an increased mutation rate versus an increase fitness compared to wild-type[86,87]. Novick and Slizard observed that the growth-rate dependence of mutation differed depending on which growth medium they used[74]. Because of the potential for different growth media to exert different mutational and selective pressures, each growth medium may be associated with its own balance of mutation and selection.

In our mutant accumulation data, it is possible that the relative contribution of each mutational process is not constant across growth rates. From work in glucose-limited E. coli chemostats, and in metabolic studies of *Mycobacterium bovis* BCG, lower concentrations of residual glycerol in chemostats run at low dilution rates suggests that there may be stronger selection for glycerol affinity at low growth rates compared to high growth rates[65,82]. Our measurements of residual glycerol at fast and slow growth rates are consistent with this (data not shown). More limited nutrient concentration could lead to more stress-induced mutagenesis. Supporting this, a study of genetic requirements for survival during fast and slow growth in BCG showed differential survival for the Ung transposon mutant, deficient in DNA damage repair[65]. In order to begin to parse the differential contribution of replication errors, DNA damage, and stress-induced mutagenesis, the mutant accumulation assay could be performed in the genetic background of mutants deficient for these processes – a differential increases in mutant accumulation at fast and slow growth rates would point to differential contribution of each mutational process.

A working model: Our findings point to a number of important factors that should be included when considering the inter-dependence of growth rate, fitness, and mutation rate. While more complex mathematical modeling is required to capture the dynamics that we observed in the chemostat, a schematic for how these parameters may interact is illustrated in Figure 2.9. Mutant accumulation can be driven by time-dependent mutation (including mutation arising from DNA damage or stress-induced mutagenesis) or replication-dependent mutation. Mutant wash-out is driven by the dilution rate, and the relative fitness cost compared to the current most-fit strain. The wild-type population is also continually evolving, there are periodic genetic sweeps accompanied by an overall increase in population affinity for residual glycerol. The mutant population may collapse due to the accompanying loss in relative fitness, or the mutant population may collapse because it has exceeded the threshold for the given combination of fitness cost and growth rate for this chemostat. Indeed, we adapted the Monod model used in Figure 2.8 to simulate simple mutant accumulation assay in which mutants arise as a function of growth rate. We varied the parameters d, K_s and μ_{max} for the mutant strain across a physiological range, and we found that with a high enough starting mutant concentration, the mutant population collapses immediately (Supplementary Figure 2.1a). We also found that the K_s and μ_{max} parameters are interdependent and together with growth rate, determine the timing and shape of mutant growth kinetics (Supplementary Figure 2.1b-d).

Further work investigating the molecular drivers of replication-independent mutation will elucidate possible links between slow growth, starvation, stress, and mutation. Molecular drivers may include components of the SOS response and DosR regulon, the error-prone and translesion polymerases DnaE2 and DinB, and components of DNA damage repair. Important in future experiments will be comparisons of transcript levels and assessing changes in mutant

accumulation rates in fast and slow growth for strains deficient in these components. Because a hallmark of *Mtb*'s life cycle is the ability to enter a latent state, it will be important to investigate further the drivers of replication-independent mutation, potentially an important mechanism for continued adaptive evolution during latency.

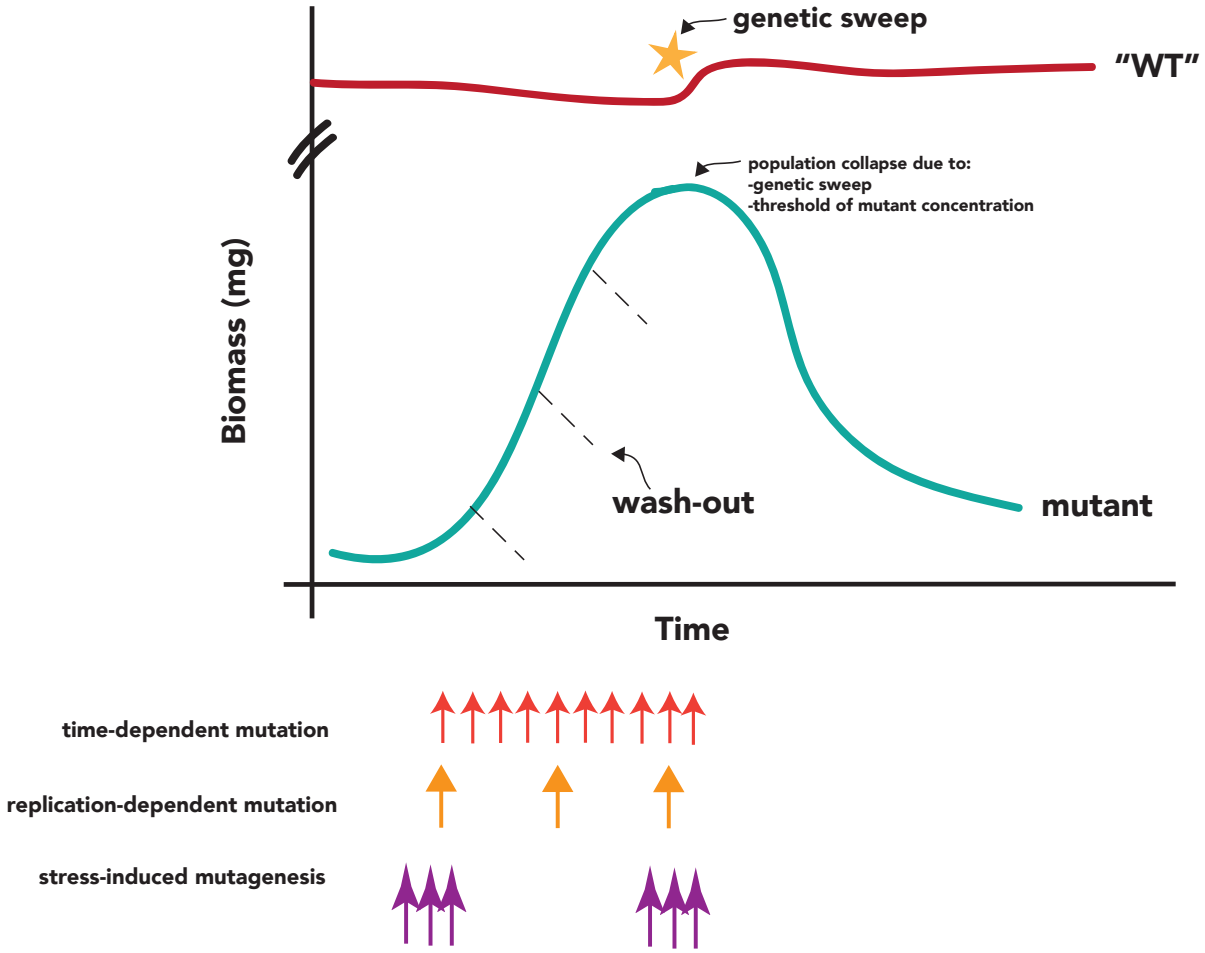


Figure 2.9 Possible contributors to time-dependent and replication-dependent mutation in the chemostat. Mutant accumulation can be driven by replication errors or time-dependent mutation. Replication errors (red arrows) may be constant per replication cycle. Time-dependent mutation may arise as a constant "tick" of DNA damage (orange arrows) due to metabolic byproducts or as a more punctuated event (purple arrows) such as mutation arising from stress-induced mutagenesis. Stress-induced mutagenesis may occur during periods of transition that result in temporary severe nutrient limitation, such as start of chemostat pumps, or emergence of a more fit "wild-type" strain resulting in increased competition for the limiting substrate. Here mutant wash-out is driven by the dilution rate and the relative fitness cost compared to the current most-fit strain. The wild-type population is also continually evolving; there may be periodic genetic sweeps accompanied by an overall increase in population affinity for residual glycerol (star). The mutant population may then collapse due to the accompanying loss in relative fitness, or the mutant population may collapse because it has exceeded the threshold for the given combination of fitness cost and growth rate for this chemostat.

2.5 Materials and Methods

Bacterial strains and media

M. smegmatis mc^2155 was used for all experiments. Plating was performed on 7H10 plates (BD Biosciences) supplemented with albumin (Sigma), dextrose, and catalase (VWR) and 20% Tween 80 (Sigma). Antibiotic concentrations were as follows: rifampicin (Rif) (200ug/mL); ciprofloxacin (Cipro) (2g/mL), kanamycin (20ug/mL) (Sigma). Some culturing was performed in 7H9 media (BD Biosciences), supplemented with albumin/dextrose/catalase and 20% Tween-80. Chemostat media was defined as follows: For 1L of media: 1mL of 1000X trace elements 1 ($CoCl_2$x6H_2O (0.5g/L); CuSO4×5H_2O (0.5g/L); FeSO4x6H2O (7g/L); KI (0.2g/L); MnSO4x4H2O (0.4g/L); Na2MoO4×2H2O (0.2g/L); NiCl2 anhydrous (0.2g/L); ZnCl2 (1g/L), 1mL of 1000X trace elements 2 (H3BO3 (0.2g/L); Na-EDTA (15g/L)), 2mL of 10% tyloxapol, 2mL of 50% glycerol. 1g of $(NH_4)_2SO_4$, 0.2g of $MgSO_4$×7H_2O, 0.04g of $CaCl_2$×2H_2O. pH was adjusted to 6.5-7. Volume was brought to 1L with sterile, de-ionized, distilled H_2O and the 1L was autoclaved. Autoclaved separately in 50mL of de-ionized, distilled H_2O was 0.8g of KH_2PO_4, 1.2g of Na_2HPO_4, and after cooling, was combined with the 1L of media.

Continuous culture of *M. smegmatis*

Glycerol-limited continuous cultures of *M. smegmatis* were grown in 1-liter bioreactors fitted with a sealed top plate (Applikon Biotechnology). Inflow media was contained in a 10L reservoir and addition of nutrients to chemostat cultures was controlled by a peristaltic pump (Watson Marlow). Sample outflow was regulated using an overflow tube fitted to another pump. Culture was stirred constantly using an Applikon 1032 stirrer controller fitted with a tissue culture impeller set. Temperature was regulated using a heating blanket, set to 37 degrees. Temperature, pH, and dissolved O_2 measurements were automated through use of an Applikon

1010/1025 Bio-controller and console and accompanying probes. A defined air mixture was provided through a sterile filter fitted in the top plate. Effluent gas was passed through a cooling condenser and CO_2 gas was analyzed continuously.

Identical 100uL aliquots of mc^2155 freezer stocks were used to start all chemostat experiments. Starter cultures were inoculated with a freezer stock aliquot and grown in 50mL of chemostat media in 150mL PETG square bottles (Nalgene, Rochester NY). When the starter culture was turbid, it was transferred into the 1L vessel via syringe through a sterile rubber septum fitted into the top plate. Cultures were then grown in batch culture mode (without inflow or outflow pumps running) until nearly all residual glycerol was consumed, at which point pumps were started at dilution rates of 0.15 to .01 h^{-1}. Steady state was defined as equilibration of CO_2 output, biomass, and optical density.

Mutant accumulation assays

Chemostats were sampled daily and plated (20mL) on Rif plates and (5mL) on Cipro plates. Wherever possible depending on sample volume constrains, plating was performed in duplicate or triplicate. Daily microscopy of cultures and plating for CFU was also performed regularly on 7H10 plates containing no antibiotics to check for contamination or any changes in cell morphology. Rif plates were enumerated at 5 and 7 days, Cipro plates were enumerated at 3 days. The slope of the

Genetic barcoding and *rpoB* sequencing

For genetic barcoding experiments, chemostats run at doubling time = 46 hours and 3 hours were sampled and plated on Rif (200ug/mL) as part of mutant accumulation assays. Approximately 120 Rif-resistant colonies were picked and used as templates for colony PCR of the Rifampicin resistance-determining region of *rpoB*, and amplicons were sent for Sanger sequencing. A

subset of colonies was grown further in 10mL 7H9 cultures. Aliquots of these cultures were frozen for future experiments, and the remaining culture was prepared for electroporation by centrifugation followed by washing in 10% glycerol, repeated three times. Electrocompetent cells were transformed with one of 11 integrating plasmids containing a genetic "barcode," each containing a constant region (identical across barcodes) and a unique region (unique to each barcode). Barcodes have been described previously[88]. Barcode sequences were optimized such that TaqMan probes and primers could be used to amplify multiple barcodes in a single qPCR reaction. qPCR analysis was optimized further in the Fortune Laboratory (Martin *et al.,* unpublished). Primer pair efficiencies were confirmed for each pair separately for every run by calculating slope of crossing threshold values from 10-fold serial dilutions of plasmid template.

Competition experiments

To perform chemostat competition experiments, frozen stocks of barcoded strains were grown in 5mL of chemostat media to an optical density of approximately 1, and then expanded to 50mL cultures until turbid. 50mL cultures were normalized by optical density and combined in a laminar flow hood to produce the inoculum. Slow growth rate was achieved in a continuous culture chemostat, $d = .015h^{-1}$. Daily sampling was performed in triplicate.

Fast growth rate was achieved using a serial-fed batch culture model of chemostat growth, where 20mL of chemostat media was inoculated with 10mL of inoculum in a 50mL PETG square bottle. Cultures were grown overnight until the optical density ~ 1. 20uL of each culture was transferred to 20mL of fresh chemostat media under a laminar flow hood. Cultures were grown for 24 hours, and the 1:1000 dilution into fresh media was repeated daily. Serial batch cultures were performed in technical duplicates.

For each competition experiment, 5mL of chemostat culture was sampled in triplicate from the slow growth rate chemostat, and in duplicate from the serial-fed batch culture, daily. Samples were stored at 4 degrees Celsius until 1 week's worth of samples could be batch-analyzed together. DNA was then extracted from each sample using a phenol:chloroform:isoamyl alcohol purification with bead-beating, followed by isopropanol precipitation.

qPCR analysis was performed on the sample triplicates from the slow growth rate chemostat and on the sample duplicates from the fast growth rate serial-fed batch culture. Each qPCR amplification reaction was performed in technical triplicates (3 wells per primer pair). As the proportion of each competing strain was nearly identical across sample replicates (determined by overlapping mean and standard deviation), only one replicate is plotted for simplicity.

Mathematical modeling

The Monod model of bacterial growth was adapted from Monod *et al*. as shown in Figure 2.7 and expanded to 10 strains according to the following set of equations[74,81,89]:

(1) $dy(1) = y(1)*(\mu_1-D)$

(2) $dy(2) = y(2)*(\mu_2-D)$

(3) $dy(3) = y(3)*(\mu_3-D)$

(4) $dy(4) = y(4)*(\mu_4-D)$

(5) $dy(5) = y(5)*(\mu_5-D)$

(6) $dy(6) = y(6)*(\mu_6-D)$

(7) $dy(7) = y(7)*(\mu_7-D)$

(8) $dy(8) = y(8)*(\mu_8-D)$

(9) $dy(9) = y(9)*(\mu_9-D)$

(10) $dy(10) = y(10)*(\mu_{10}-D)$

(11) $dy(11) = (D)*(S_i-y(11)) - (\mu_1*y(1))/Q - (\mu_2*y(2))/Q - (\mu_3*y(3))/Q - (\mu_4*y(4))/Q - (\mu_5*y(5))/Q - (\mu_6*y(6))/Q - (\mu_7*y(7))/Q - (\mu_8*y(8))/Q - (\mu_9*y(9))/Q - (\mu_{10}*y(10))/Q$

Where $\mu 1 = \mu_{max1}*(y(11)/(k_1+y(11)))$ where k_1 = the Monod constant for strain 1

Simulations were performed in MATLAB. The equations were parameterized as follows.

Slow growth rate chemostat

S0 = 2 mg/L

S_i = 900 mg/L

Q = 0.4

D = .015 $^{h-1}$

Time = 0 to 280 hours

Inoculum for each strain: [0.134 0.178 0.104 0.129 0.112 0.128 0.093 0.113 0.124 0.149]*250mg

k_S and μ_{max}

k_s = [230 190 260 270 270 190 200 210 178 183] mg/L

μ_{max} = [.180 .185 .175 .175 .175 .20 .215 .185 .23 .23] h^{-1}

For the serial-fed batch cultures:

Serial-fed batch culture was modeled with doubling time = 0, and a 1:1000 dilution every 24 hours, starting on day 2. Initial dynamics during the first 24 hours were not included in the model.

S0 = 1000 mg/L

S_i = 1000 mg/L

Q = 0.5

Total time = 0 to 192 hours

Inoculum for each strain: [0.17 0.18 0.18 0.22 0.23 0.15 0.125 0.18 0.200 0.142]*.01mg

$\underline{k_s \text{ and } \mu_{max}}$

ks = [170 170 210 220 220 170 200 165 185 183] mg/L

μ_{max} = [.21 .219 .185 .2 .19 .22 .205 .226 .239 .23] h^{-1}

CHAPTER THREE

Metabolically-Induced DNA Damage and Mutagenesis in Mycobacteria

Richa Gawande[1], Watthanachai Jumpathong[2,3], Susovan Mohanthapa[2], Peter Dedon[2], Sarah Fortune[1]

[1]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA
[2]Department of Genetic Toxicology, Massachusetts Institute of Technology, Cambridge, MA
[3] Department of Chemistry, Chiang Mai University, Chiang Mai, Thailand

Author contributions R.G., **W. J., S.M., P. D., and S. F** designed the experiments and performed data analysis. **R.G., W. J., and S. M.** performed the experiments and data analysis (**R.G.**: constructed strains, *in vivo* growth experiments, fluctuation analyses; **W. J:** discovery of adduct, and methodological development to measure adduct and glyoxylate in cells; **S.M:** methodological development to measure adduct and glyoxylate in cells). **R.G., W.J., and S.F.** drafted the manuscript.

3.1 Abstract

Recent studies of the bacterial mutation rate during *Mycobacterium tuberculosis* infection suggest that replication-independent processes may govern mycobacterial genomic fidelity. Here we investigate DNA damage as a possible source of replication-independent mutation. Little is known about the types of DNA damage that mycobacteria sustain during infection, and whether this damage is associated with mutation. Here we focus on fatty acid metabolism, a major component of the bacterial diet *in vivo*, as an endogenous source of DNA damage and mutation. We found that a byproduct of fatty acid metabolism, glyoxylate, reacts with DNA to produce a previously undiscovered DNA adduct. Using a novel LC/MS-MS method to detect and quantify these adducts in *Mycobacterium smegmatis* (MSm) cells, we found that glyoxylate adduct levels increased during fatty acid metabolism, and that levels further increased in the absence of nucleotide repair (NER), suggesting a role for this pathway in glyoxylate adduct repair. Further, fatty acid metabolism was mutagenic specifically in the absence of nucleotide repair and not in the absence of other repair pathways, suggesting that NER repairs mutagenic DNA damage in this condition. Addition of exogenous glyoxylate increased the mutation rate of NER-deficient cells, consistent with glyoxylate as a potential driver of fatty acid-induced mutagenesis. Finally, we found that metabolism of cholesterol, another major carbon source during infection, is highly mutagenic to wild-type and repair-deficient cells. Our work is the first demonstration of metabolically-induced DNA damage and mutation in mycobacteria, and offers one potential mechanism for replication-independent mutation. Further understanding of the relationship between DNA damage and mutation will help elucidate the rates and drivers of drug resistance during infection.

3.2 Introduction

The rapid emergence of drug resistant Mycobacterium tuberculosis calls for a better understanding of the rates and drivers of drug-resistance mutations during infection. Results from the macaque model of Mtb from human transmission chains, and from *in vitro* measures of mutation in strains across different global lineages have contributed to a growing appreciation for the mutational capacity of mycobacteria. These studies have revealed that mutations accumulate at a similar rate per day across different *in vivo* infection stages as they do *in vitro*, and that rates can vary from lineage to lineage[14,30,31,32,33]. Furthermore, we showed in **Chapter 3** that mutants can accumulate in the chemostat at similar rates at slow and fast growth rates. The idea that slow growing bacteria could achieve the same amount of genetic diversity as rapidly growing bacteria goes against the understanding that replication errors are the primary source of mutation, and therefore only actively replicating bacteria are capable of developing the mutations that cause drug resistance. An alternative explanation is that accumulation of mutations may be driven by mechanisms other than errors in replication. An analysis of the spectrum of mutations that had accumulated during both the macaque and human infection showed enrichment for products of oxidative damage, suggesting that DNA damage may be a source of mutations *in vivo*[30].

Oxidative damage is one of the major defenses against mycobacteria during infection. The main sources of reactive species during Mtb infection are inducible nitric oxide (iNOS) and superoxide radicals produced by activated macrophages[39,40,41]. iNOS-/- mice are severely attenuated during Mtb infection, suggesting that oxidative and nitrosative damage are critical defenses against Mtb[42]. Transcriptional analyses of Mtb growth in the mouse model of infection showed induction of the main pathways for DNA damage repair in mycobacteria, base excision

repair (BER) and nucleotide excision repair (NER). Confirming the importance of these pathways during infection, genetic essentiality screens have revealed that components of both NER and BER are essential for growth in different stages of infection. Specifically, UvrB was required for growth in the mouse and macaque model of infection, and during growth on acidified nitrite, a condition that recapitulates the intracellular environment of the macrophage. The uracil glycosylase, Ung, is required for growth in *in vitro* models designed to recapitulate certain features of the *in vivo* environment, such as exposure to hypoxia, starvation-induced slow growth, and cholesterol[50,56,61,64,65]. The attenuated survival of the Ung transposon mutant during growth on cholesterol and during slow *in vitro* growth suggests that even in the absence of host superoxide and iNOS defenses, there are endogenous processes that can cause DNA damage. Finally, the damage-induced error-prone polymerase, DnaE2, was shown to be required for virulence in the mouse model of Mtb[66]. Yet almost nothing is known about whether endogenous processes are directly genotoxic to mycobacteria, and whether this genotoxicity lead to mutation.

In order to investigate whether there are replication-independent are sources of mutation in mycobacteria, we considered whether there could be endogenous metabolites that are mutagenic. Studies from other organisms suggest that intermediary metabolism can generate electrophilic species that have the potential to be genotoxic. These include acetyl coA from glycolysis and fatty acid oxidation, succinyl coA from the TCA cycle, and aldehyde-containing compounds, which are capable of reacting with nucleophilic sites in DNA and proteins[90,91,92]. While one study has shown that electrophilic species can react with DNA to produce a mutagenic adduct, it is unknown which electrophiles are produced during mycobacterial metabolism, and whether these products are mutagenic[93].

Mycobacteria rely heavily on metabolism of fatty acids during infection. Fatty acid metabolism redirects carbon away from the TCA cycle and through the glyoxylate shunt, allowing conservation of carbon through formation of glyoxylate instead of loss as carbon dioxide. Transcription of glyoxylate shunt enzymes is highly elevated during a dormancy model of infection. In addition, mutants that are unable to enter the glyoxylate shunt are attenuated for survival in the mouse model of infection, and cannot survive on media containing fatty acids as the only carbon source[94]. The glyoxylate shunt is similarly important for proper growth and development in plants during stress conditions[95,96]. We noted that glyoxylate is an aldehyde-containing compound that is likely to be electrophilic. We hypothesized that the aldehyde acts as a functional group that can modify nucleic acids and proteins in the same manner as formaldehyde and acetaldehyde[97,98]. While glyoxylate treatment was shown to be mutagenic in some Salmonella species, this finding has not been further investigated[99].

Jumpathong *et al* (2014) recently developed a reduction-based analytical method to show that glyoxylate is capable of reacting with 2-deoxyguanosine to form the unstable adduct, N2-carboxyhydroxymethyl 2'-deoxyguanosine (N2-CHMdG)[100]. Here we applied the method to test the hypothesis that this adduct accumulates during fatty acid metabolism, and explore the relationship between adduct levels, adduct repair, and mutation. We found that a shift from the TCA cycle to the glyoxylate shunt leads to elevated levels of glyoxylate adduct. We then tested the role of the major DNA damage repair pathways during growth in these conditions and found a specific role for NER in reducing the number of glyoxylate adducts. Further, consistent with the mutagenic effect of this adduct, we found that fatty acid metabolism is mutagenic in the absence of NER, but not in the absence of other repair pathways, and that addition of exogenous glyoxylate had a mutagenic effect in the absence of NER, further pointing to the specific role of

NER in repair of metabolically-induced adducts. Finally, we found a general mutagenic effect of other carbon sources that require the glyoxylate shunt, suggesting that the scope of metabolically-induced mutation extends beyond acetate.

3.3 Results

In order to determine whether use of the glyoxylate shunt causes glyoxylate-DNA adducts, we developed an analytical method to detect novel glyoxylate-DNA adducts, which we predicted would form as a result of using the glyoxylate shunt. This work has been detailed in Jumpathong (2014) and is briefly summarized here[100]. dG was reacted with ethyl glyoxylate and the only product formed was identified by chromatography-coupled mass spectrometry as N2-carboxyhydroxymethyl 2'-deoxyguanosine (N2-CHMdG). The structure was analyzed by high-resolution mass spectrometry with collision-induced dissociation. Characterization of N2-CHMdG revealed that it is unstable and degrades to release dG with a short half-life. Reduction of N2-CHMdG led to the formation of N2-carboxymethyl-2'-dexyguanosine (N2-CMdG), which was to be stable for weeks. With this stable adduct, an isotope-dilution, chromatography-coupled tandem mass spectrometric method was then developed to quantify N2- CHMdG as N2-CMdG in calf thymus DNA treated with glyoxylate. The method proved to be extremely sensitive and revealed a dose-response relationship for the formation of N2-CHMdG in glyoxylate-treated DNA.

We expected that we might be able to detect adducts *in vivo* in conditions in which metabolic flux is shifted to the glyoxylate shunt. We grew MSm in minimal media supplemented with acetate as the only carbon source. We focused on acetate because is a simple, two-carbon fatty acid that is metabolized entirely through the TCA cycle/glyoxylate shunt and

doesn't require additional pathways like the methylcitrate cycle, which is required for metabolism of propionate or cholesterol. We generated a knockout of a TCA cycle enzyme, isocitrate dehydrogenase, whose function is to bypass the glyoxylate shunt. In order to test whether adducts form in these conditions, we grew wild-type MSm or icdΔ in rich and acetate media. As the icdΔ strain does not grow in acetate media (we attribute this to a deficiency in alpha-ketoglutarate, data not shown), we measured glyoxylate-dG in this strain in rich media only. We isolated DNA using a method adapted to stabilize the glyoxylate adduct (see Methods), and then used LC-MS/MS to quantify adducts. Glyoxyl-dG adducts were present in wild-type cells grown in rich media, suggesting that there is basal use of the glyoxylate shunt. Adduct levels increased when wild-type cells were shifted to acetate media, but this increase was not significantly different. IcdΔ cells had a significantly higher level of adducts than wild-type cells grown on either rich or acetate media. (Figure 3.1).

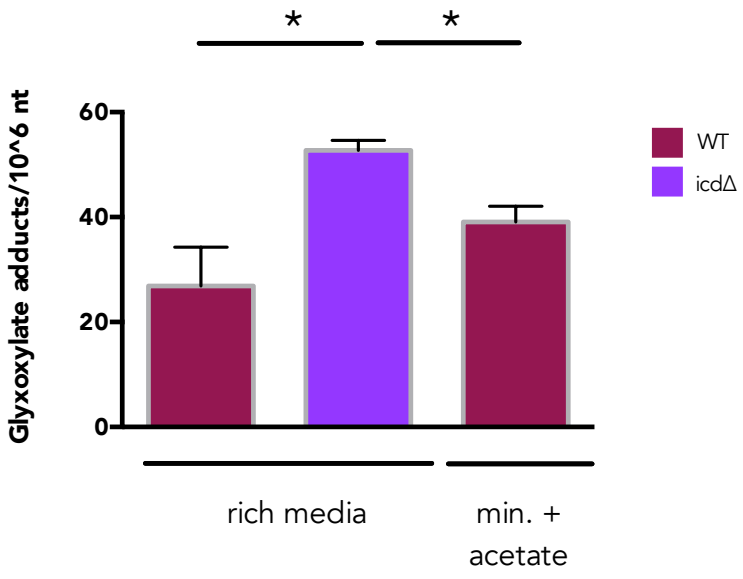


Figure 3.1 Forcing carbon flux through the glyoxylate shunt results in an increase in glyoxyl-dG adducts. Wild-type MSm mc^{2}155 and the icdΔ strain were grown to mid-log phase; DNA was extracted using acidified phenol and cyanoborohydrate to stabilize the glyoxylate adduct. Glyoxyl-dG adducts were quantified by LC/MS-MS. Data are representative of three independent experiments. * $p < .05$, unpaired t-test.

The increase in adducts led us to consider whether fatty acid metabolism induces an increase in the steady-state level of the cytoplasmic glyoxylate pool. An LC-MS/MS method to detect glyoxylate in MSm cell extracts using an aldehyde reactive probe (ARP) with an external standard was described in Jumpathong (2014)[100]. The levels of cytoplasmic glyoxylate were similar among strains and conditions (Figure 3.2).

As mutation reflects a balance of damage and repair, we considered whether the increase in adducts would be accompanied by a corresponding increase in mutation rate in these conditions. We considered three possibilities: one possibility is that the adduct is mutagenic in the presence of repair, the adduct is getting repaired and therefore is not mutagenic in the presence of repair, or the adduct is not mutagenic. We measured mutagenesis in the presence or absence of repair to distinguish between these possibilities. We first measured mutation rate by performing fluctuation analysis in the icdΔ strain or during growth on acetate. Both wild-type MSm grown on acetate and the icdΔ had equivalent mutation rates as compared to wild-type Msm grown in rich media (Figure 3.3). Growth on propionate did not significantly change the mutation rate.

We then considered whether glyoxylate adducts are mutagenic but are repaired sufficiently such that they don't cause an increase in mutation rate. To test the relationship between DNA repair, glyoxylate adducts and mutagenesis, we created DNA damage repair mutants in the two major DNA damage pathways, BER and NER. Specifically, we made mutants deficient in uracil repair (Ung), base-excision repair (Fpg and Fpg2), and nucleotide-excision repair (UvrC). We performed fluctuation analysis in these strains under rich and fatty acid growth conditions.

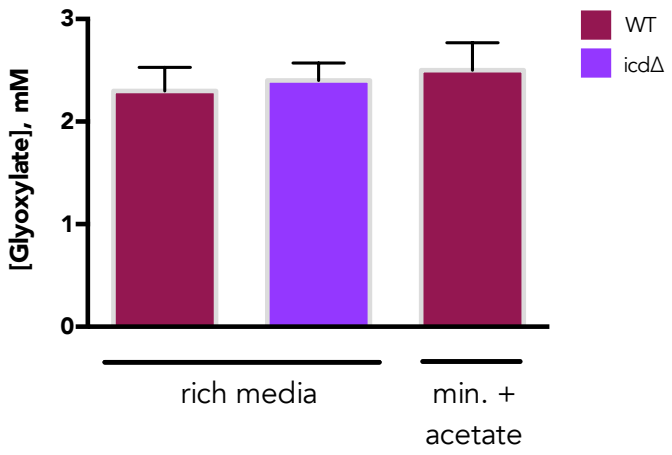


Figure 3.2. Cellular glyoxylate concentrations do not change upon shift to fatty acid metabolism. Glyoxylate pools were measured in cell extracts of wild-type mc^{2}155 and icdΔ cells using an aldehyde-reactive probe with external calibration. Data are representative of two independent experiments. Concentrations are normalized to the estimated number of cells per culture (3×10^8), with an estimate of cellular volume of 0.65 μm^3.

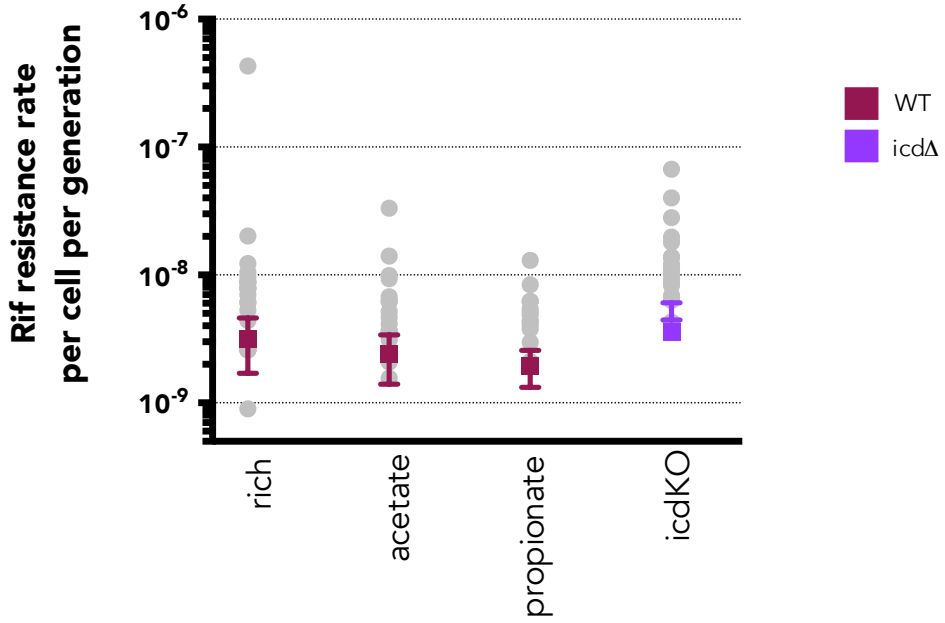


Figure 3.3 Mutation rate does not change upon shift to fatty acid metabolism. Fluctuation analysis was performed to determine the rifampicin resistance rate (200ug/mL) in wild-type mc^{2}155 cells grown in rich media, minimal media with either acetate or propionate supplemented as the only carbon source, and icdΔ cells grown in rich media. Grey circles represent the mutant frequency of an individual fluctuation analysis culture; boxes represent the Luria Delbruck estimation of the mutation rate; error bars represent the 95% confidence interval.

The UvrCΔ strain was the only strain to have a significantly higher mutation rate when shifted to fatty acid as compared to rich media, and also as compared to wild-type or the complemented strain grown on fatty acids. The UvrCΔ strain's basal mutation rate in rich media was also reproducibly higher than wild-type or the complemented strain, but this difference was not statistically significant. While the UngΔ strain had a higher basal mutation rate than the wild-type and complemented strains, this rate did not further increase upon a shift to fatty acid media. Finally, the FpgΔ strain did not have an increased basal mutation rate compared to wild-type or the complement in rich media, but there was a reproducible trend for an increased mutation rate upon the shift to fatty acid media, but this change was not statistically significant (Figure 3.4).

These data suggest that fatty acid metabolism is mutagenic specifically in the absence of NER. We further explored the carbon-source specificity of this mutagenesis. We first measured the mutation rate under minimal media conditions with glycerol as the only carbon source. We did not see a statistically significant increase in the mutation rate of UvrCΔ strain compared to the complement or wild-type. As the metabolism of fatty acids could cause accumulation of genotoxic metabolites, we explored whether the increase in mutation rate on fatty acid media in the absence of repair could be rescued by addition of glucose as an additional carbon source. When we supplemented acetate media with glucose, (Figure 3.5), the UvrCΔ strain mutation rate was significantly lower than its mutation rate in the absence of glucose, but still higher than compared to the complemented strain, suggesting that the absence of repair in the presence of fatty acid is sufficient to cause mutagenesis, though to a lesser extent when glucose is also present. We explored whether the mutation rate of the UvrCΔ strain would be differentially responsive to a shift to an odd-chained fatty acid, such as propionate. We found a similar increase in mutation rate for the UvrCΔ strain in propionate as compared to acetate, suggesting

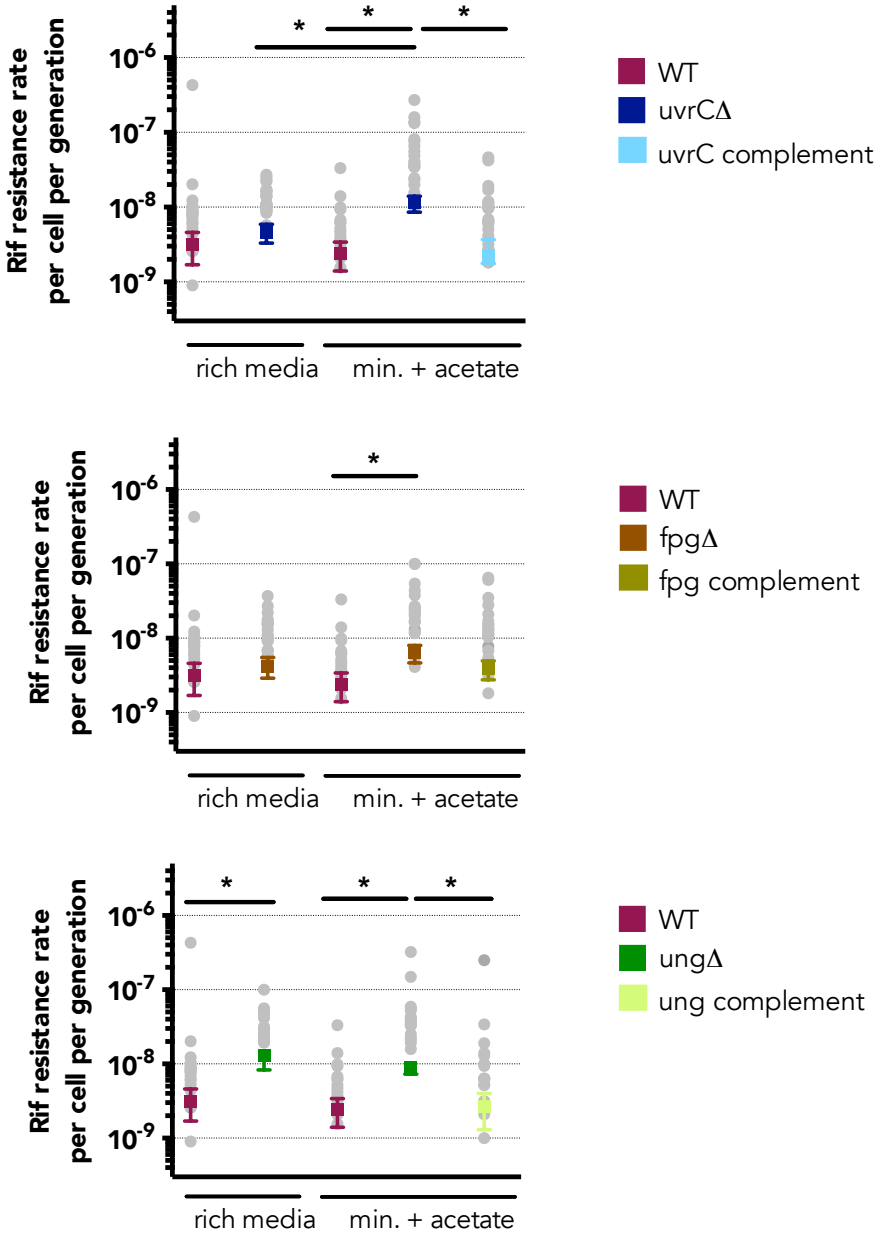


Figure 3.4 uvrCΔ cells have a higher mutation rate when grown on fatty acid compared with rich media; ungΔ cells have a higher mutation rate than wild-type mc^{2}155 when grown on rich media but this rate does not increase upon a shift to fatty acid media. Fluctuation analysis was performed to determine the rifampicin resistance rate (200ug/mL) in wild-type mc^{2}155 cells, ungΔ cells, fpgΔ cells, uvrCΔ cells, and the respective complemented strains grown in rich media or minimal media with acetate. Grey circles represent the mutant frequency of an individual fluctuation analysis culture; boxes represent the Luria Delbruck estimation of the mutation rate; error bars represent the 95% confidence interval. * represents significance at 5% by non-overlapping error bars. Data are representative of three experiments.

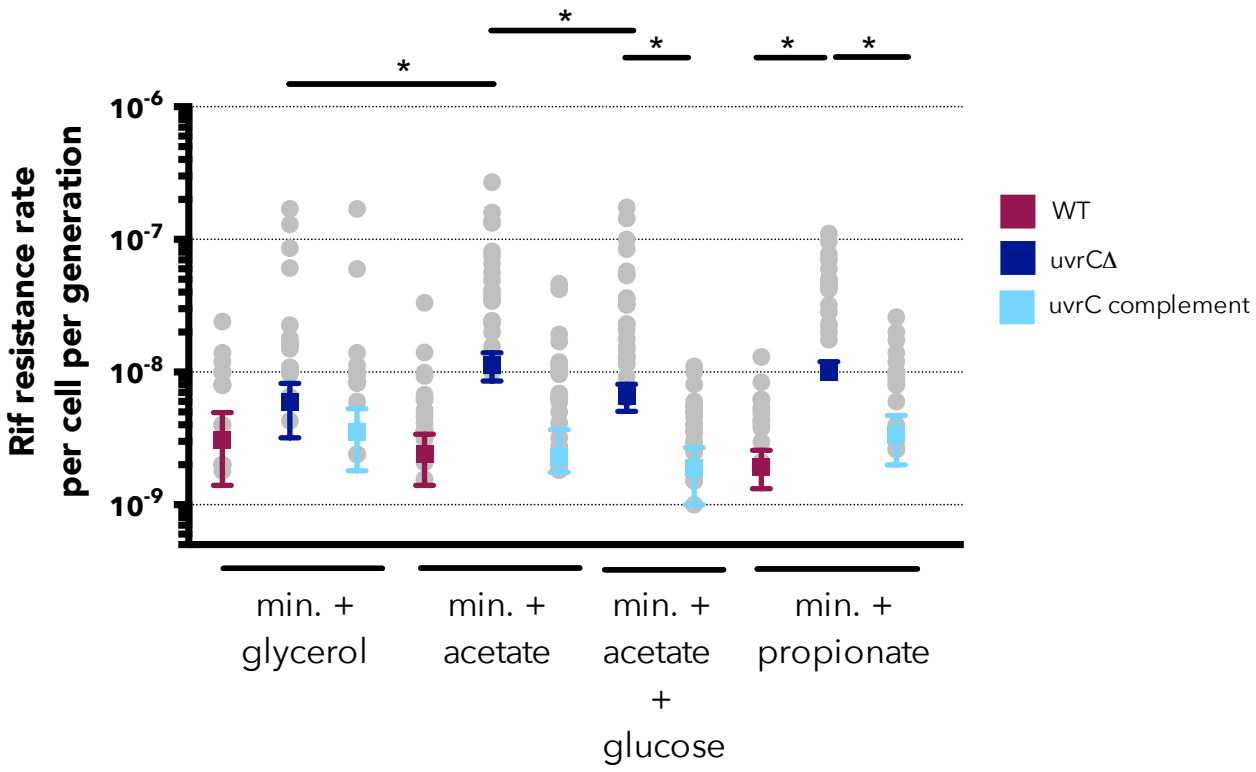


Figure 3.5 Addition of glucose partially rescues the increase in mutation rate for uvrCΔ cells that is observed upon shift to acetate alone; propionate media induces a similar increase in mutation rate as acetate. Fluctuation analysis was performed to determine the rifampicin resistance rate (200ug/mL) in wild-type mc^2155 cells, uvrCΔ cells, and uvrC complemented cells grown in rich media, minimal media supplemented with acetate and glucose, or propionate. Grey circles represent the mutant frequency of an individual fluctuation analysis culture; boxes represent the Luria Delbruck estimation of the mutation rate; error bars represent the 95% confidence interval. * represents significance at 5% by non-overlapping error bars. Data are representative of three experiments.

that the overall balance of damage and repair is not significantly different during propionate metabolism compared with acetate metabolism. The above data suggested that there may be a link between the glyoxyl-dG adducts, repair by UvrC, and mutagenesis during growth on fatty acid media in the UvrCΔ strain. While glyoxyl-dG adducts increased in wild-type MSm grown on acetate media compared to rich media, growth on acetate media was only mutagenic for the strain lacking NER. We hypothesized that if the UvrCΔ strain is responsible for repairing glyoxyl-dG adducts, there would be more adducts in the UvrCΔ strain grown on acetate than the complemented or wild-type strains in this condition, corresponding with the increase in mutation rate. We measured glyoxyl-dG adducts in wild-type, UvrCΔ, and complemented strains grown in acetate media. There was an increase in glyoxyl-dG adducts in the UvrCΔ strain compared to the complement and WT (Figure 3.6). These data suggest that UvrC is responsible for glyoxyl-dG adduct repair during acetate metabolism.

We hypothesized that lack of repair of glyoxyl-dG adducts is responsible for the increased mutation rate in the UvrCΔ strain grown on acetate. To elucidate whether glyoxylate is directly mutagenic to DNA, we performed a blue/white mutagenesis assay in glyoxylate-treated plasmid DNA containing the lacZ gene and a Kanamycin-resistance marker. When plasmid was exposed to increasing doses of either UV irradiation or glyoxylate and then transformed into UvrCΔ or UvrC complemented MSm, the transformation efficiency decreased for both strains, consistent with general toxicity of these treatments (data not shown). However, there was no significant increase in the white/blue ratio across any of the conditions for either strain, suggesting that this assay is not sensitive enough to reveal mutagenesis.

We instead treated cells directly with glyoxylate and looked for differential toxicity in the absence and presence of NER. The UvrCΔ and complemented strains were grown in glycerol-containing minimal media with or without glyoxylate and their growth rates were compared.

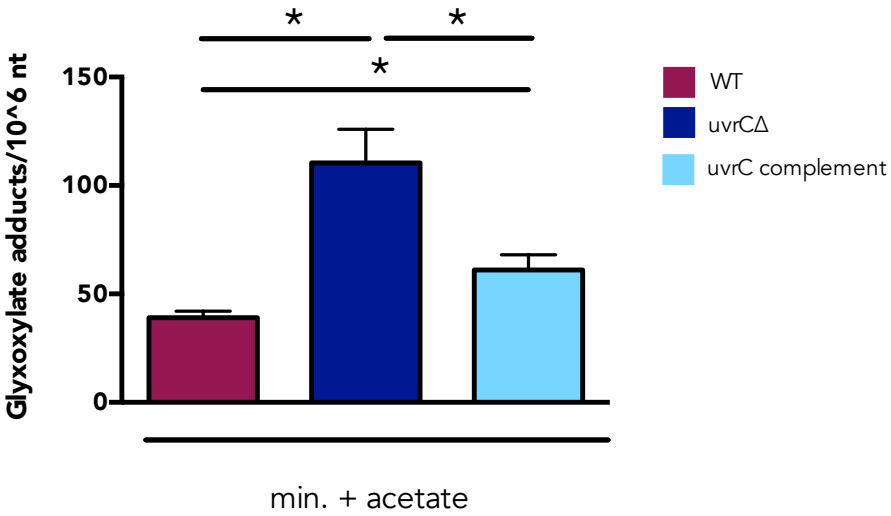


Figure 3.6 uvrCΔ cells have a higher concentration of glyoxyl-dG adducts than the wild-type mc^{2}155 or complemented strains. Wild-type MSm mc^{2}155, uvrCΔ, and uvrC complemented strains were grown to mid-log phase; DNA was extracted using acidified phenol and cyanoborohydrate to stabilize the glyoxylate adduct. Glyoxyl-dG adducts were quantified by LC/MS-MS. * indicates $p < .05$, unpaired t-test.

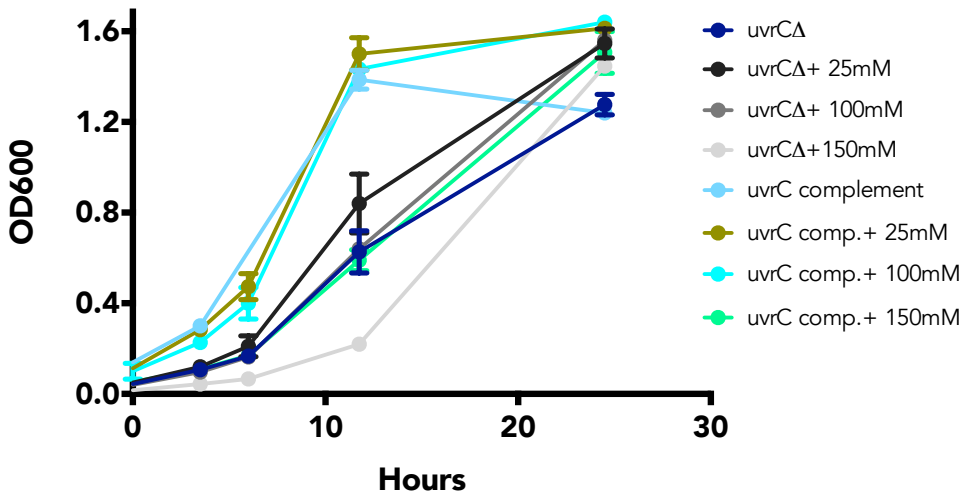


Figure 3.7 Addition of glyoxylate to minimal media supplemented with glycerol does not significantly change the relative growth rates of uvrCΔ or uvrC complemented strains. 50mL cultures grown in the indicated media with varying concentrations of glyoxylate (25mM, 100mM, or 150mM) were inoculated with frozen stocks of either uvrCΔ, uvrC complemented, or wild-type mc^2155 strains at a calculated OD600 of ~.001. OD600 measurements were taken at the indicated times. Error bars represent the standard deviation of three technical replicate cultures for each strain/media combination.

While the UvrCΔ strain showed a general growth defect compared to the complemented strain, this defect was not accentuated by the presence of glyoxylate. Both strains grew normally at low concentration of glyoxylate, but showed a growth defect on high concentrations of glyoxylate (> 100mM) (Figure 3.7).

We then measured the mutation rate in cells treated with a non-toxic concentration of glyoxylate (25mM). The UvrCΔ strain had an increased mutation rate compared with wild-type and complemented strains in the presence of glyoxylate, whereas in the no-glyoxylate condition, the mutation rate of all strains were equivalent (Figure 3.8). The UvrCΔ strain trended towards an increased mutation rate in the presence versus absence of glyoxylate, consistent with a mutagenic effect of glyoxylate, but this difference was not statistically significant. Further work will elucidate the dose response for glyoxylate mutagenesis and glyoxylate-dG adduct accumulation.

We then expanded our study to investigate the mutagenic effects of cholesterol metabolism, which has been suggested to involve production of toxic metabolites. Because cholesterol metabolism involves use of the glyoxylate shunt for beta-oxidation, we hypothesized that growth on cholesterol may have a mutagenic effect similar to growth on acetate. We measured the mutation rate of wild-type MSm and the DNA repair mutants in the presence or absence of cholesterol (Figure 3.9). The mutation rate increased by ~10-fold for all strains; there was no significant difference in mutation rate across repair mutants, suggesting that the relationship between DNA damage, repair, and mutation during cholesterol metabolism may be complex and involve multiple components.

Future experiments will be required to elucidate the dose response of the mutagenic effect, to develop assays to identify known and unknown DNA adducts induced in response to cholesterol metabolism, and further probe the role of repair pathways.

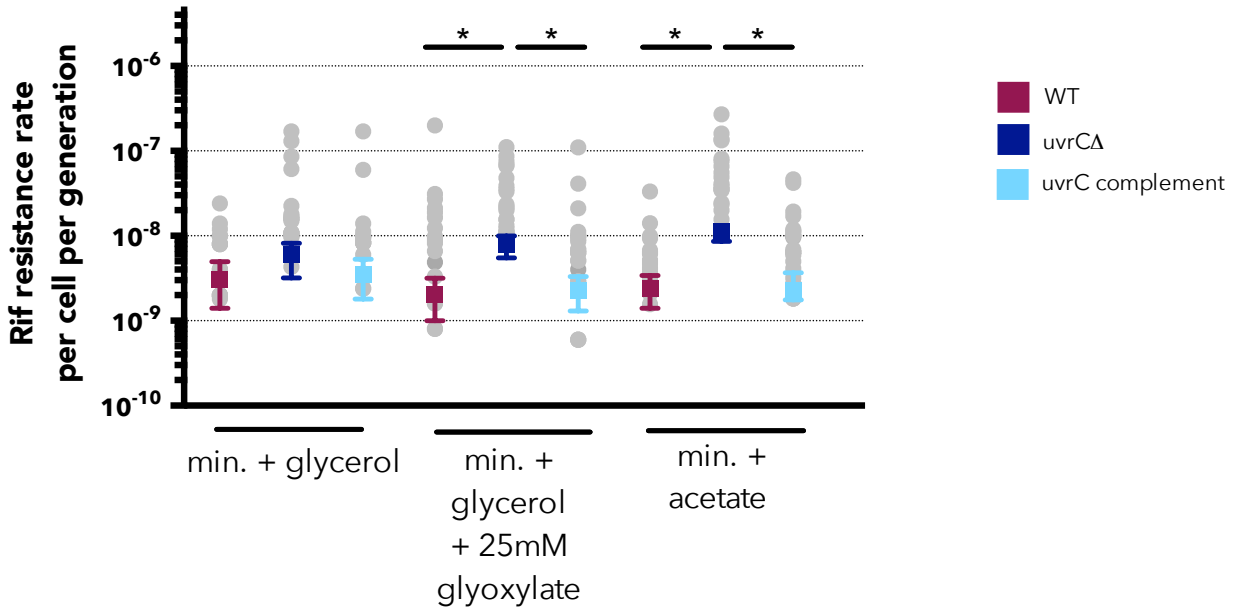


Figure 3.8 Addition of 25mM glyoxylate results in an increase in mutation rate in uvrCΔ cells compared with the mutation rate of uvrC complemented or wild-type mc^{2}155 cells in the presence of glyoxylate. Fluctuation analysis was performed to determine the rifampicin resistance rate (200ug/mL) in wild-type mc^{2}155 cells, uvrCΔ cells, and uvrC complemented cells grown in minimal media supplemented with glycerol, with or without the addition of 25mM glyoxylate. Grey circles represent the mutant frequency of an individual fluctuation analysis culture; boxes represent the Luria Delbruck estimation of the mutation rate; error bars represent the 95% confidence interval. * represents significance at 5% by non-overlapping error bars. Data are representative of three experiments.

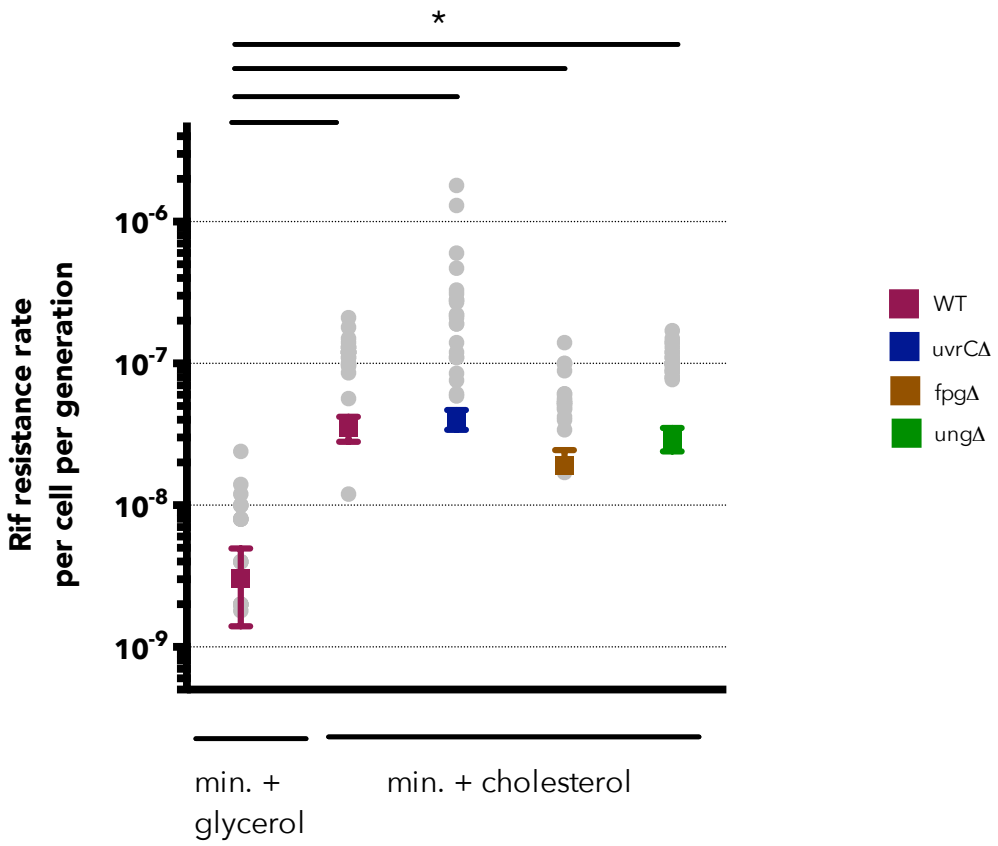


Figure 3.9 Cholesterol metabolism causes a uniform increase in mutation rate across wild-type mc^{2}155, uvrCΔ, fpgΔ, and ungΔ cells. Fluctuation analysis was performed to determine the rifampicin resistance rate (200ug/mL) in wild-type mc^{2}155, uvrCΔ, fpgΔ, and ungΔ cells grown in minimal media supplemented with cholesterol. Grey circles represent the mutant frequency of an individual fluctuation analysis culture; boxes represent the Luria Delbruck estimation of the mutation rate; error bars represent the 95% confidence interval. * represents significance at 5% by non-overlapping error bars. Data are representative of two experiments.

3.4 Discussion

In the current work, we discovered and characterized a DNA adduct formed in the reaction between DNA and glyoxylate, an electrophilic metabolite. We found that this adduct was induced by growth conditions that were predicted to require the glyoxylate shunt such as acetate metabolism and in the absence of the TCA cycle enzyme, isocitrate dehydrogenase. This is the first evidence of genotoxicity associated with use of the glyoxylate shunt, a pathway critical to the pathogenesis of Mtb *in vivo*. Further, we defined a role for nucleotide excision repair in modulating the levels of this adduct during growth on fatty acid.

We show that fatty acid metabolism is mutagenic in the absence of nucleotide excision repair, demonstrating the relationship between metabolism, DNA damage, DNA repair, and mutation. Our data point to glyoxylate as a potential cause of the increased mutagenesis during fatty acid metabolism. Previous findings have focused on byproducts of inflammation in mammalian cells. A few studies on glyoxylate in *Arabidopsis* suggests that enzymes that reduce glyoxylate are likely to play an important role in redox balance and stress response[101,102,103]. Our discovery and detection of glyoxylate-induced DNA adducts highlights a direct relationship between metabolic state and DNA damage that was previously uncharacterized in mycobacteria. Our finding of the NER in glyoxylate adduct repair highlights the role of DNA damage repair as a metabolically-driven guardian of genomic fidelity. It had been known that DNA damage repair is important to DNA fidelity, but little is understood about for what lesions and in what conditions. Ung-deficient cells were previously shown to have a higher mutant frequency during basal conditions, but formal measurements of mutation rate were not made[104,105]. In our study, fatty acid metabolism was mutagenic in the absence of NER, but the mutation rate didn't change

for Ung or Fpg deficient cells, suggesting an NER-specific mutagenesis. Future work will elucidate the role of other repair pathways, if at all, in glyoxylate adduct repair.

The glyoxylate adduct was detected when repair was intact, suggesting that there may be a low-level basal use of the glyoxylate shunt. It is possible that fatty acids present in rich media or catabolism of endogenous cell wall fatty acids are metabolized through the glyoxylate shunt. However, mutagenesis occurred only in the absence of NER during fatty acid growth conditions, suggesting that there may be a concentration threshold for glyoxylate adducts above which glyoxylate is mutagenic. Creation of an icdΔ/uvrCΔ double mutant could represent a genetic mimic of the uvrC/fatty acid condition, and could reveal a further increase of glyoxylate adduct levels and associated mutagenesis. In our study, free glyoxylate concentrations did not change significantly between rich media and fatty acid conditions. As glyoxylate is extremely reactive, it is possible that most of the cellular glyoxylate is in a bound form; cellular macromolecules such as DNA, lipids, and proteins may act as a glyoxylate "sink," obscuring our ability to detect increases in free glyoxylate. The relationship is between cytoplasmic glyoxylate pools, glyoxylate adducts, and mutagenesis remains to be understood. It is possible that in order for mutagenesis to occur, the glyoxylate adduct levels must reach a concentration above the threshold of the repair pathway, determined by both spatial and temporal features of adduct formation and repair enzyme availability. Treatment with higher doses of glyoxylate and measurement of adducts in these conditions will further elucidate the dose-response in glyoxylate-induced mutagenesis.

Metabolism of odd-chained fatty acids and sterols involves a more complex coordination of metabolic pathways than metabolism of acetate. We expanded our study to investigate the mutagenic effect of cholesterol metabolism. Cholesterol is a major source of carbon for Mtb

during late time-points in the mouse model of infection and involves use of the glyoxylate shunt for beta-oxidation, and the methylcitrate cycle and/or the methylmalonyl coA cycle for sterol degradation into acetyl-coA, propionyl-coA, and pyruvate[106,107]. In our study, cholesterol metabolism was mutagenic in repair-proficient cells and was equally mutagenic for UvrCΔ, UngΔ, and FpgΔ mutants. To our knowledge, this is the first demonstration of the mutagenic effect of cholesterol metabolism. Previous work on Mtb mutants that have defects in cholesterol metabolism showed that *in vitro* growth defects in these mutants cannot be rescued with the addition of another carbon source, but both *in vivo* and *in vitro* growth defects can be partially rescued in a background where cholesterol metabolism is completely avoided, either in a cholesterol-important mutant, or where the entire pathways are deleted. These results have been broadly interpreted as evidence that cholesterol metabolism involves toxic byproducts[108,109]. It is possible that cholesterol metabolism may be genotoxic, and the mutagenic effect that we observe could be due to unrepaired DNA damage or DNA damage-induced mutagenesis. Further work involving mass spectrometry analysis could elucidate whether cholesterol metabolism can be linked to known or novel types of DNA damage. Cholesterol metabolism may also produce byproducts that alter the function of DNA replication and repair pathways, leading to increased mutagenesis due to increased replication errors or decreased repair. We did not observe a differential mutagenic effect in the three DNA damage repair mutants that we tested. If multiple repair pathways are involved, single deletions may not be sufficient to reveal an effect. It is also possible that the cholesterol-induced mutations are repaired by pathways not involved in NER and BER.

Our data reveal that DNA is sensitive to metabolic state. Given the a paucity of information regarding the changes in metabolic state of mycobacteria during *in vivo* growth,

DNA damage could be a useful probe to understand what mycobacteria encounter in the *in vivo* environment. While the spectrum of mutations is not a sensitive read-out of metabolic state since many types of damage can lead to the same mutation, the spectrum of DNA damage can be highly specific to metabolic state. There are hundreds of known adducts, and many novel adducts that we know nothing about yet. It is possible that there are differences in the carbon metabolism of MSm (soil-dwelling) versus Mtb (human pathogen), as their environments may call for a different metabolic balance. For example, MSm has an intact methylcitrate cycle (MCC) to metabolize propionate, whereas the Mtb *icl1* is a bifunctional enzyme that plays a role both in the GC and MCC, though it is unclear whether these differences lead to functional consequences[110]. Further work in both MSm and Mtb will elucidate whether there are differences in the relationship between carbon metabolism, DNA damage, and mutation in these species. If there is metabolically-driven mutagenesis in Mtb, this may suggest that the mutation rate is higher *in vivo* than our *in vitro* estimates predict. Taken in light of the recent finding that some clinical strains have higher mutation rates than others[14], it is possible that differential DNA replication and repair, as well as differences in central metabolism and lipids to mop up toxic byproducts could all contribute to the mutational capacity of mycobacteria *in vivo*. Further work will continue to elucidate the complex relationship between metabolism, DNA damage, DNA repair, and mutation.

3.5 Materials and Methods

Isolation of glyoxylate adduct and quantitation of cytoplasmic glyoxylate using an aldehyde reactive probe

Please refer to Jumpathong (2014)

Bacterial strains and media

M. smegmatis mc^2155 was cultured in Middlebrook 7H9 salts (BD Biosciences) supplemented with albumin/dextrose/catalase (ADC), 5g/L glycerol, and 0.05% Tween-80 (Sigma). Selection was performed on 20ug/mL kanamycin, 150ug/mL hygromycin, 10ug/mL zeocin, or 50ug/mL nourseothricin (NAT) (Sigma) on 7H10 plates (BD Biosciences) supplemented with ADC and 5g/L glycerol. Minimal media was defined as follows: For 1L of media: 1mL of 1000X trace elements 1 ($CoCl_2$x6H_2O (0.5g/L); CuSO4×5H_2O (0.5g/L); FeSO4x6H2O (7g/L); KI (0.2g/L); MnSO4x4H2O (0.4g/L); Na2MoO4×2H2O (0.2g/L); NiCl2 anhydrous (0.2g/L); ZnCl2 (1g/L), 1mL of 1000X trace elements 2 (H3BO3 (0.2g/L); Na-EDTA (15g/L)), and 2mL of 10% tyloxapol. 1g of $(NH_4)_2SO_4$, 0.2g of $MgSO_4$×7H_2O, 0.04g of $CaCl_2$×2H_2O. pH was adjusted to 6.5-7. Volume was brought to 1L with sterile, de-ionized, distilled H_2O and the 1L was autoclaved. Autoclaved separately in 50mL of de-ionized, distilled H_2O was 0.8g of KH_2PO_4, 1.2g of Na_2HPO_4, and after cooling, was combined with the 1L of media.

Media was supplemented with either 1g/L glycerol, 1g/L sodium acetate, 1g/L sodium propionate, 1g/L glucose, or .01% cholesterol dissolved in ethanol as described elsewhere[64]. Controls for cholesterol experiments were performed in the same concentration of ethanol.

Construction of deletion mutants and complementing strains

Deletion of *ung, uvrC*, *fpg*, and *icd* from *M. smegmatis* mc^2155 was performed using recombineering as previously described[111]. 3' and 5' flanks were 250bp long, designed such to

preserve the transcriptional start site, and synthesized by Genscript (Piscataway, NJ) into puc57-simple. Flank constructs and corresponding antibiotic markers flanked by *loxP* sites were double-digested using PacI and SpeI and cloned together using T4 Ligase, transformed into DH5α competent *E. coli*. Antibiotic markers for each deletion construct are as follows: *ung:* NAT *uvrC:* NAT *fpg:* zeocin *icd:* hygromycin Plasmids were extracted using Qiagen mini-prep kits, and the recombineering substrate was PCR-amplified using custom primers. The substrate was then electroporated into *M. smegmatis* expressing the recombineering plasmid; transformants were selected on kanamycin. DNA was isolated from transformants using a phenol:chloroform:isoamyl alcohol extraction with beadbeating, and an isopropanol precipitation. Deletions were confirmed by PCR validation using the following 4 sets of primers: within the gene to be deleted (no product expected), 5' within the MSm genome and 3' within the antibiotic marker; 5' within the antibiotic marker and 3' within the MSm genome; and 5' and 3' within the antibiotic marker. Complementing strains were created by gene synthesis by Genscript (Piscataway, NJ) and Gibson assembly (New England Biolabs, Ipswich, MA) to clone constructs into the integrating vector, pJEB402[112]. Knockout strains were cured of the recombinase plasmid using counter-selection on 10% sucrose followed by confirmation of loss of plasmid by plating on kanamycin(+) and (-) plates with the appropriate selection marker for each deletion strain. pJEB402 containing the complementing gene was electroporated into cured deletion strains and transformants were screened by growth on kanamycin, and for restoration of phenotype by UV susceptibility assays (for *uvrC*) or by fluctuation analysis (for *ung*).

Fluctuation analysis

Fluctuation analysis was performed as described previously[14,78]. In brief, a starter culture was grown from a freezer stock to an optical density (OD600) = 1 in 7H9 media. The same set of

freezer stocks was used for all strains for all experiments to ensure reproducibility. The starter culture was used to inoculate 140mL of media to an effective cell concentration of 2.4×10^5 cells, and this culture was divided into 25 5mL cultures grown in 15mL conical Falcon tubes (BD Biosciences). Cultures were grown until they reached an approximate OD of 0.8-1, and all but 3-5 cultures were harvested by centrifugation at 4000RPM. Pellets were plated onto 7H10 plates containing 200ug/mL Rifampicin (Sigma) and spread using plastic loop spreaders. The remaining cultures were used to perform 5 10-fold serial dilutions, and these dilutions were then plated on 7H10 plates containing no antibiotic. CFU plates were enumerated after 3 days of incubation at 37 degrees and Rifampicin plates were enumerated after 5 and 7 days of incubation. Drug resistance rate was determined by calculating the estimated number of mutations per culture using the analytic and statistical methods and MATLAB (Natick, MA) scripts described elsewhere[14,78,113].

CHAPTER FOUR

Whole-genome Sequencing to Investigate the Cause of Recurrent Tuberculosis Infection in HIV-infected Individuals

Authors: Richa Gawande[1], Kogieleum Naidoo[2], Osman Abdullahi[1], Michael Chase[1], Nesri Padayatchi[2], Preshnie Moodley[2], Salim Abdool Karim[2], Sarah Fortune[1]

[1]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA

[2] Center for the AIDS Programme of Research in South Africa (CAPRISA), KwaZulu Natal, South Africa

Author contributions R.G., **K.N., O.A., and S.F.** designed the study. **O.A., N.P., and P.M** collected and processed patient samples, performed drug sensitivity assays, and genotyping. **O.A.** isolated DNA from patient samples. **R.G.** constructed sequencing libraries, performed sequencing, and developed and performed data filtering and analysis. **M.C.** developed and performed data analysis. **R.G., K.N., S.A.K., and S.F.** drafted the manuscript.

4.1 Abstract

The risk of recurrent tuberculosis infection is highest in HIV-infected populations where TB is endemic. In South Africa, many patients with HIV who are apparently successfully treated for TB will develop TB again. Little is understood about why patients develop recurrent infection following apparently successful treatment. Here we investigated the cause of recurrent infection in the first 13 HIV-infected individuals enrolled in a prospective cohort study, TB Recurrence Upon Treatment with Highly Active Antiretroviral Therapy (TRuTH), in KwaZulu Natal, South Africa.

The TRuTH study seeks to establish the incidence of recurrent infection in HIV-TB coinfectected patients who have been initiated on highly active antiretroviral therapy (HAART) and received apparently adequate treatment for TB infection. Patients were followed for up to 60 months after TB treatment for evidence of recurrent TB. Here we analyzed the presenting and recurrent Mtb strains for the first 13 patients who developed recurrent disease using used 24-locus mycobacterial interspersed repetitive units and variable number tandem repeat (MIRU-VNTR) typing and whole-genome sequencing (WGS). We found that 7 of the 13 recurrent episodes appeared to represent relapse of the primary infection based on a high degree of genetic similarity between the presenting and recurrent strains. Results from MIRU-VNTR typing were concordant with WGS for highly related and highly divergent strains, but only WGS analysis could resolve intermediate genetic distances. In contrast with previous reports that relapsed infections are most likely to occur soon after the end of treatment, patients presented with relapsed infection and reinfection up to 4 years after completion of treatment and no difference was found between the timing of relapse and reinfection cases ($p < .55$). Finally, while we found no differences in phenotypic drug susceptibility as measured by the 1% proportion method

between relapse and reinfecting strains, Mtb strains that caused relapsed infection were more likely to harbor genetic polymorphisms associated with changes in susceptibility to isoniazid (INH) or the acquisition of INH resistance ($p < .03$).

Despite the force of TB transmission in KwaZulu Natal and the successful HIV and TB treatment history, over a half of the 13 recurrent episodes represented relapse of the initial infection. Strains causing relapsed infection are more likely to harbor genetic polymorphisms associated with changes in INH susceptibility or acquisition of INH resistance.

4.2 Introduction

The introduction of directly observed therapy for tuberculosis (TB) has cured approximately 56 million people, and saved 22 million lives[114]. Yet, patients who receive successful TB treatment can develop recurrent tuberculosis infection. The WHO estimates that recurrent infection represents 5% of all new tuberculosis infections, though the rate can vary tremendously and may be as high as 20% in some places[7,8,9,10,11,12,115]. HIV infected individuals who present with TB infection, representing approximately 12% of new TB cases, are at greater risk than HIV-negative individuals for recurrent TB infection[115,116]. Recurrent infection also poses a greater treatment challenge in HIV-infected individuals, in whom even standard first-line TB treatment regimens are not uniformly effective[115].

Despite the burden of recurrent infection, little is understood about why some patients develop recurrent infection following apparently successful treatment. Studies to identify risk factors for recurrence are complicated by the fact that recurrent infection represents both reinfection with an exogenous strain of TB and relapse of the primary infecting strain. The risk of relapsed infection, reflecting treatment failure, is thought to be highest soon after the end of treatment. By contrast, the rate of reinfection is thought to be steady over time and driven by the local rates of ongoing TB transmission[116].

Here we investigated the cause of recurrent infection in the first 13 HIV-infected individuals enrolled in a prospective cohort in KwaZulu-Natal, South Africa, where 60% of TB patients are HIV-positive, and whose TB incidence is among the highest in the world[114]. Where estimated, the burden of recurrent infection in South Africa is high, ranging from 15-20% of successfully treated patients[7,12]. In this study, we assessed the genetic relatedness of presenting and recurrent Mtb strains first using standard molecular typing, and because these typing

techniques are limited by low discriminatory power, then through whole-genome sequencing (WGS). Despite apparently successful treatment, we found that over half of the recurrent episodes appeared to represent relapse of the primary infection. Patients presented with relapsed infection and reinfection up to 4 years after completion of treatment; indeed, no difference was found between the timing of relapse and reinfection cases. Only one patient in this study presented with drug-resistant Mtb, as assessed phenotypically by drug-susceptibility testing. However, as compared to the Mtb strains causing reinfections, the Mtb strains in relapsed infections were more likely to harbor genetic polymorphisms associated with either changes in susceptibility to isoniazid (INH) or the acquisition of INH resistance.

4.3 Results

MIRU-VNTR typing to estimate genetic distance between presenting and recurrent Mtb strains

We investigated the cause of recurrent TB infection in 13 HIV-infected patients who were enrolled in the prospective cohort study, TB Recurrence Upon Treatment with Highly Active Antiretroviral Therapy (TRuTH), at the Center for the AIDS Programme of Research in South Africa (CAPRISA) in KwaZulu Natal. The TRuTH study aims to determine the incidence of recurrent TB infection in patients being treated with highly active antiretroviral therapy (HAART) who were previously treated for a TB infection. Patients were treated for an initial TB infection during their participation in the Starting AIDS Treatment at Three Points in TB Treatment (SAPIT) study, which was designed to determine the most effective timing of HAART during TB treatment[117]. 700 patients with a known TB outcome were enrolled and then followed up for up to 60 months in the TRuTH study. We assessed the cause of recurrent TB in

first 13 patients who developed recurrent disease by analyzing the *M. tuberculosis* (Mtb) isolates from presenting and recurrent episodes (Table 4.1).

Table 4.1 Patient characteristics

Category	Relapse	Reinfection	p-value
Total	7	6	
HIV positive	7	6	1.00
Months until recurrent infection (25%,75%)	28.3 *(19.9,39.7)*	24.08 *(15.4, 30.9)*	.582
HIV suppression at the end of treatment	4	3	1.00
Previous history of TB	3	2	1.00
Number cured	6	6	1.00
Number receiving delayed ART in SAPIT study	1	4	.102
Lung infiltrates (both lungs)	6	4	.559
Lung cavitation (one lung)	3	3	1.00
SPOT sputum (+++)	4	1	.266
EMS sputum (+++)	2	0	.46

We sought to determine whether episodes of recurrent infection represented a relapse of the initial infection or reinfection with an exogenous strain. We first performed 24-locus MIRU-VNTR typing on each patient's presenting and recurrent Mtb strains. MIRU-VNTR typing separated the patients into two populations: 7 patients had 0 MIRU-VNTR loci differences between their initial and recurrent isolates and 6 patients had 10+ MIRU-VNTR loci differences (Figure 4.1a). In molecular epidemiology applications, strains with 0-1 MIRU-VNTR loci differences are interpreted as being identical, while strains differing by multiple loci are considered distinct strains. According to this classification, 7 of the 13 patients developed recurrent tuberculosis due to the same strain of Mtb, apparent relapsed infections. The remaining 6 appeared to have been reinfected with distinct strains of Mtb.

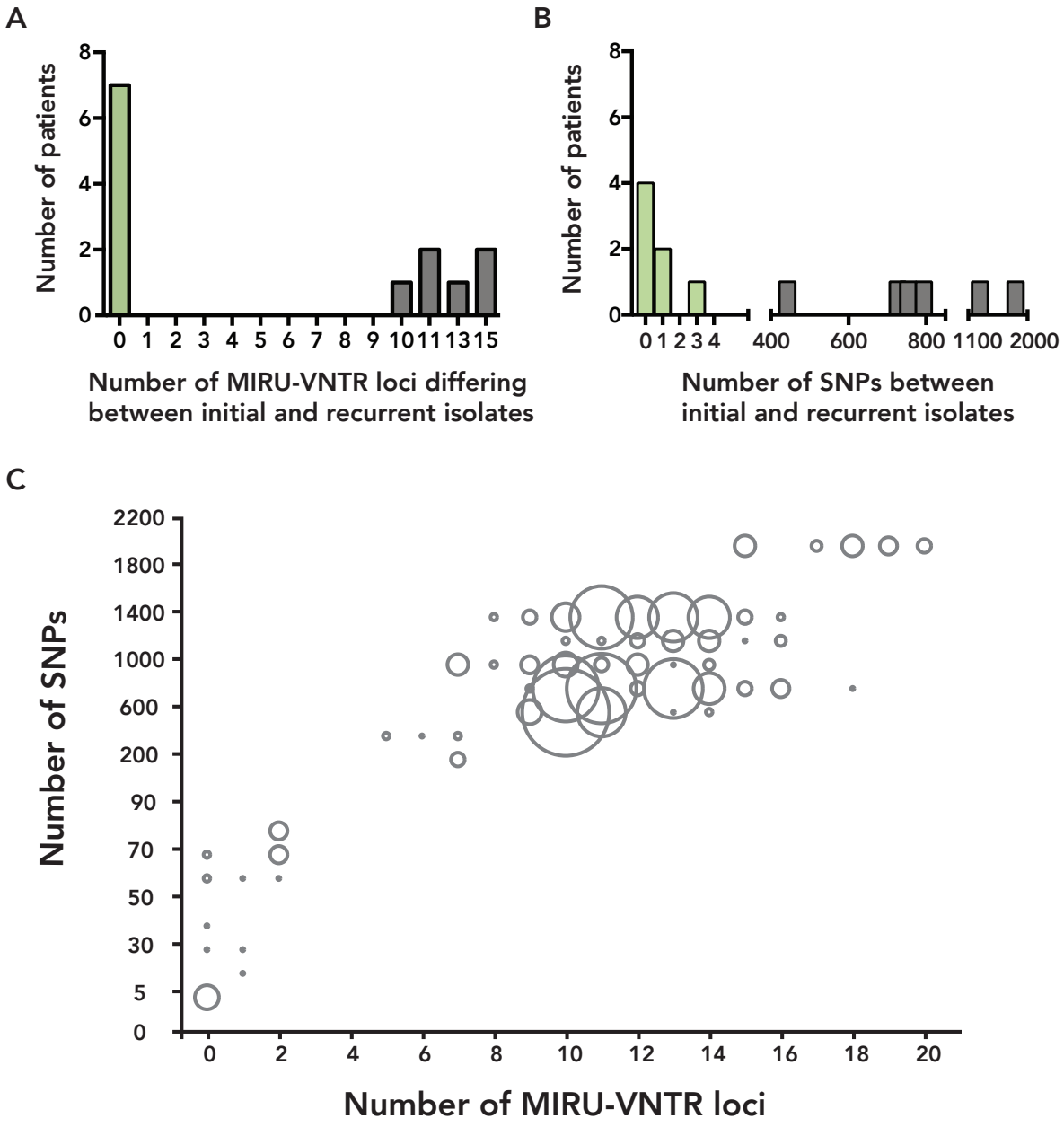


Figure 4.1 MIRU-VNTR typing and whole-genome sequencing are concordant in estimating genetic diversity between initial and recurrent strain. a) MIRU loci difference between initial and recurrent infection; b) SNP difference between initial and recurrent infection; c) Pairwise comparison of all isolates by MIRU type and corresponding SNP distance. Circle size corresponds to the number of patient pairs that matched the given (MIRU, SNP) distances.

No differences in timing of recurrent and relapsed infection

Only one of the 7 patients with apparently relapsed infection had documented interruptions in TB treatment. Given the successful HIV and TB treatment history of the remaining patients and the force of transmission in KwaZulu Natal, the relatively high proportion of relapsed infection was surprising. In addition, despite expectations that risk of relapse would be highest soon after completion of treatment, which would be interpreted as treatment failure, no difference was found between the timing of relapse and reinfection (Figure 4.2). Indeed, we found apparently relapsed infections that occurred up to 50 months after apparently successful treatment for both TB and HIV.

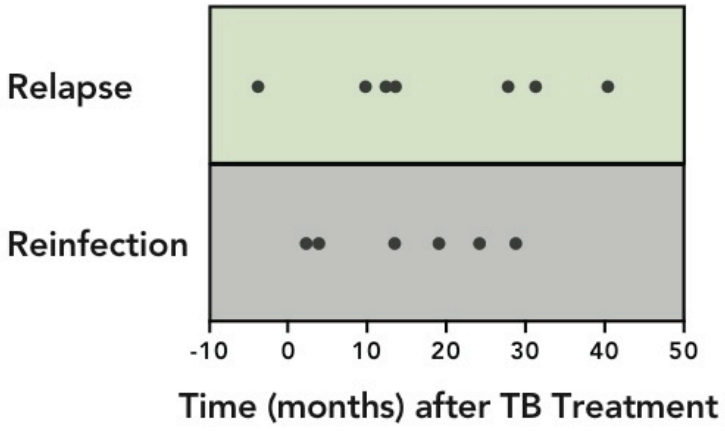


Figure 4.2 Rate of relapse is equivalent to rate of reinfection over time. The time from the end of treatment for the initial TB infection is indicated for each patient (black circles). The x-axis is time (months) after TB treatment when the patient presented with recurrent infection. Negative values indicate the patient was still in treatment when they had relapsed TB. Patients who had relapsed TB did not have significantly different timing of recurrent infection than patients who were reinfected by a new strain ($p = .55$, unpaired t-test).

Whole-genome sequencing offers higher resolution than MIRU-VNTR typing in estimating genetic distance between strains

Based on these considerations, we thought it possible that MIRU-VNTR typing, which relies on a limited number of loci, misidentified relapsed infections. We therefore performed whole-genome sequencing, in order to define the differences between primary and recurrent infecting strain with higher genetic resolution. Sequencing was performed in technical replicates for each Mtb isolate at an average depth of 80-fold, with an average genome coverage of 95%. While we could document some instances of within host evolution of the infecting Mtb strain, we found no evidence of mixed infection in these samples. Indeed, most heterologous base calls were a result of mis-alignment and could not be confirmed by resequencing.

Whole genome sequencing also separated patients into two distinct populations, one with few single-nucleotide polymorphism (SNP) differences between their initial and recurrent isolates (< 4), and one with hundreds of SNP differences (400-1900) (Figure 4.1b). These SNP distances were concordant with the MIRU-VNTR typing results for these patient pairs, again suggesting that the 7 patients with nearly identical initial and recurrent strains represented relapsed infections, while the 6 remaining patients were exogenously reinfected.

Though MIRU-VNTR typing performed similarly to whole genome sequencing in estimating the relatedness of two isolates from the same patient, when pairwise comparisons were made between all isolates (676 possible pairwise comparisons among 26 isolates), the differences in the resolution of the typing methods became apparent. MIRU-VNTR typing and

whole-genome sequencing were in agreement for isolate pairs differing by less than 5 SNPs or greater than 200 SNPs. However, MIRU-VNTR incorrectly identified 6 isolate pairs differing by intermediate SNP differences (30-70 SNPs) as identical.

Investigation of phenotypic and genotypic drug-resistance in presenting Mtb strains

The number of polymorphisms that can arise within a clonal infection of a given individual is unclear and similarly it is unclear how many polymorphisms are expected to separate isolates of a given strain as it is transmitted between individuals. Small SNP differences (0-3) have been deemed consistent with both within-host evolution and clonal transmission in previous molecular epidemiology studies[32,33]. Although we interpreted genetic distances of less than 5 SNPs as evidence of relapsed infection because they were consistent with previous estimates of within-host evolution, we could not exclude the possibility that these individuals had been reinfected with a highly related strain. We therefore sought to identify clinical or biologic features that were specific to putative relapse cases, focusing on the 12 of 13 patients who completed treatment without interruption.

Previous studies have reported that lung cavitation and high bacterial burden in sputum are predictors of treatment failure[118]. In this small group of patients, we found no major differences between the clinical course of primary infection in patients who we classified as having developed recurrent tuberculosis as a result of relapse versus reinfection. However, it is important to note that in this early analysis, we lacked the statistical power to robustly detect clinical differences (Table 4.1).

We considered instead whether the presenting strains among patients who appeared to have relapsed were biologically predisposed to treatment failure. At the time of treatment, we

performed drug-susceptibility testing (DST) using the using the 1% direct proportion method. One patient was identified as being infected with multi-drug-resistant Mtb and was treated as such, but all other isolates were deemed phenotypically drug-susceptible. Conventional DST identifies strains that have high-level drug resistance. Strains may be otherwise poised to fail therapy because they have low-level drug resistances. We reasoned that polymorphisms that have been associated with the acquisition of drug resistance might be associated with treatment failure in our study, even in the absence of high-level phenotypic drug resistance. To test this hypothesis, we assembled a database of polymorphisms associated with altered susceptibility to first and second line antibiotics as reported in the TB Drug Resistance Database, including polymorphisms demonstrated to alter minimum inhibitory concentration (MIC) and polymorphisms whose effect on MIC has not been determined but have been previously found to be enriched in or harbored by drug resistant strains (Supplementary Table 4.1-4.3)[119,120,135]. In this analysis, we excluded synonymous polymorphisms and polymorphisms located in repetitive regions that are prone to sequencing and alignment error. We also excluded lineage specific polymorphisms as we found that though isolates belonging to Lineage 2 harbored the highest number of drug resistance associated polymorphisms consistent with a previous study[122], Lineage 2 and Lineage 4 were similarly represented across putative relapse and reinfection isolates (Figure 4.3a).

We tested the hypothesis that relapse-associated strains were predisposed to treatment failure because they carried polymorphisms specifically associated with resistance to first line antibiotics. We restricted our analysis to the primary isolates from the 12 patients who had successfully completed treatment, but included the one isolate assessed as phenotypically drug-resistant. We found that INH-associated polymorphisms were significantly enriched ($p < .03$) in

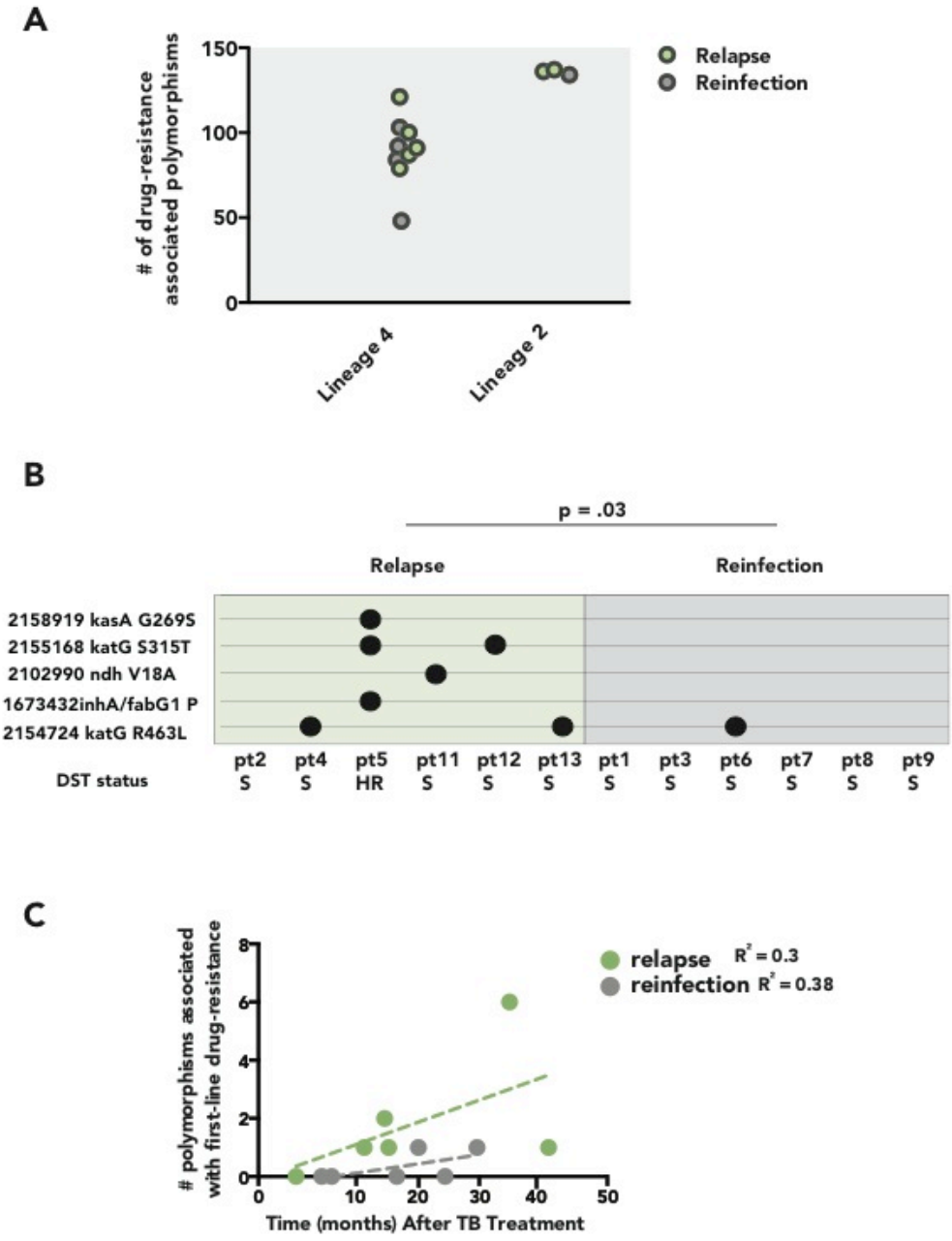


Figure 4.3 Drug-resistance-associated polymorphisms harbored by infecting strains. a) Lineage 2 strains have more drug-resistance associated polymorphisms than Lineage 4 strains (p = .007, Mann-Whitney U Test); b) Polymorphisms associated with INH-resistance are enriched in relapse patients (p = .03, Mann Whitney U test); c) Timing of recurrent infection does not correlate with the number of first-line drug-resistance associated polymorphisms harbored by the primary strain for both relapse and reinfection cases (p = .25; p = .19, respectively, for test of departure from 0 for slope of linear regression fits).

the primary isolates associated with relapse as compared to the primary strains in which recurrent disease was due to exogenous reinfection (Figure 4.3b). There was a trend for enrichment of polymorphisms associated with altered susceptibility to first-line drugs more broadly in the strains associated with apparent treatment failure though this did not reach statistical significance ($p<0.06$) (Table 4.2). By comparison, polymorphisms associated with altered susceptibility to second line drugs were uniformly distributed across the primary strains associated with recurrence due to both relapse and reinfection ($p = 0.5$). Interestingly, we also find no correlation between the timing of recurrent infection and the number of drug-resistant polymorphisms harbored by the presenting strain (Figure 4.3c); indeed, the patient with multi-drug resistant TB relapsed 42 months after apparently successfully completing appropriate therapy.

4.4 Discussion

In this study we assessed whether recurrent TB episodes in HIV-infected patients enrolled in the TRuTH study represented relapse or reinfection by analyzing the primary and recurrent Mtb isolates from these patients.

We used both conventional MIRU-VNTR typing and whole-genome sequencing to assess the relatedness of bacterial strains. One goal of our work was to define the added value of whole genome sequencing in Mtb strain analysis. Both methods unambiguously divided the patients into two groups, characterized by small and large genetic differences. These data suggest that MIRU-VNTR typing provides similar information to WGS in terms of distinguishing relapse from reinfection in a single patient over time. However, consistent with previous studies, when we performed pairwise comparisons in which the 26 strains in this study are treated as independent to mimic population scale datasets, we find that MIRU-VNTR typing significantly underestimates genetic distance between strains that differ by intermediate SNP distances. These results are concordant with a previous study of Mtb transmission chains in which MIRU-VNTR typing was found to both underestimate genetic diversity in pairs that differed by intermediate SNP differences, and in some cases, overestimate genetic diversity in pairs that differed by 5 or fewer SNPs[32,33].

Table 4.2 Drug-resistance-associated polymorphisms in patient strains

Polymorphism	# Relapse patients	# Reinfection patients
Polymorphisms with impact on drug MIC or harbored by resistant strains in known drug-resistance associated genes		
2154724 katG R463L	2	1
1673432 inhA/fabG1 P	1	0
2102990 ndh V18A	1	0
2155168 katG S315T	2	0
2158919 kasA G269S	1	0
4242803 embC V981L	1	1
4247429 embB M306V	1	0
761161 rpoB L452P	1	0
2288785 pncA T153+G (Asn)	1	0
4407640 gid V188G	1	0
4408156 gid L16R	2	3
3505298 fadE24 P	1	1
1793769 Rv1592c P	2	3
7362 gyrA E21Q	6	6
7585 gyrA S95T	6	5
9304 gyrA G668D	6	5
5520 gyrB P133L	1	1
6140 gyrB V340L	0	1
8040 gyrA G247S	1	0
9138 gyrA Q613E	1	0
3073868 thyA T202A	2	3
		INH: p = .03 first-line: p = .06 second line: p = 0.5

In our patient group, roughly half of the cases of recurrent TB were clearly due to exogenous reinfection. The other patients developed recurrent TB with a strain of Mtb that was nearly identical to their primary strain. While this provided strong presumptive evidence for relapse, we could not from these data alone rule out reinfection with a highly similar strain. Previous studies suggest a cutoff of 5 SNPs for the amount of diversity that can evolve in an Mtb population over 5-15 years, reflecting an evolution rate of 0.3-1 SNPs/year, estimated in these studies and elsewhere[32,33]. However, these criteria apply equally to the diversification of an Mtb population within a given individual and as the strain is transmitted between individuals. Indeed, there are data to suggest that the purifying selection associated with transmission might actually restrict the diversity that emerges during a transmission chain as compared to that within a single individual[121]. Thus, these criteria do not *a priori* distinguish between relapse of the primary strain and exogenous reinfection by a derivative of the primary strain circulating in a private transmission chain (for example within a family).

However, analysis of the genome sequencing data indicates that in cases where the recurrent strains differs by <5 SNPs from the primary strains, the primary strains are also biologically distinct from strains in which the recurrent case was clearly reinfection, supporting their classification as true relapses. Across relapsed isolates, there was a significant enrichment for polymorphisms associated with either an increase in the MIC to INH, or with clinical INH resistance although 6 of these 7 strains were phenotypically drug susceptible. It is possible that the constellation of polymorphisms in each patient may contribute individually, or in combination, to changes in susceptibility to first line antibiotics that are not manifest as high level resistance as detected by the 1% proportion method, thereby predisposing to treatment failure in the face of first-line drug-treatment. Two of the phenotypically drug-sensitive strains

harbored the R463L mutation in the katG gene, and one strain harbored the V18A mutation in the ndh gene; these are polymorphisms that have been detected in INH-resistant strains but whose direct impact on INH MIC is still unclear. One phenotypically drug-susceptible strain harbored the S315T mutation in katG, which has been reported to result in high-level INH resistance, indicating a discordance between the phenotypic drug-susceptibility data and the genome sequencing data. A diagnosis of INH mono-resistant TB would not have altered the treatment regimen at the time that this patient was treated, as WHO recommendations for standard of care was equivalent for INH-mono-resistant and pan-susceptible TB. Recently, in a study investigating recurrent infection after successful treatment in a Malawi population cohort, phenotypic INH resistance was associated with a higher risk of recurrent disease, where the primary strains in relapse cases were slightly more likely to be INH resistant than in reinfection cases. Genotypic analyses suggested that the S315T katG mutation was significantly enriched in primary strains causing relapse as compared to strains isolated from patients for whom a recurrent infection could not be documented. Our findings contribute to the growing body of evidence that INH resistance may be a driver of recurrent disease, and more specifically implicate polymorphisms associated with both low- and high-level INH resistance in relapse as compared with reinfection[10].

Although we did not detect significant enrichment of other drug resistance associated polymorphisms among relapse-associated strains (Supplementary Table 4.1-4.3), as this analysis was limited by a small sample size, we cannot exclude the influence of other polymorphisms that have been reported as significantly enriched in drug resistant strains according to three recent sequencing studies, or polymorphisms that fall in regions that have a high level of genetic

diversity in drug-resistant strains[119,120,135]. A larger prospective study could contribute to our understanding of the role of drug-resistance polymorphisms in treatment failure.

The timing of relapse and reinfection cases was distributed similarly in our study; both occurred up to 4 years after the completion of treatment. The late relapses do not neatly fit the model set forth in recent studies where relapse is thought to be driven by incomplete drug treatment and occurs soon after the apparent completion of treatment, and while reinfection is thought to be driven by the local TB incidence rate and therefore occurs uniformly over time[7,116]. Our study includes only patients who were carefully monitored and successfully treated, and thus we would not expect to see the early relapses. Nevertheless, our study suggests that after apparently successful treatment, patients can harbor Mtb bacteria for years and then go on to reactivate.

Recent studies indicate that INH mono-resistance, prevalent at 7% in new TB cases, is associated with higher rates of treatment failure, recurrent infection, and acquired drug-resistance, compared to infection with pan-susceptible strains[10,116,122]. The recent WHO recommendation that INH preventative therapy (IPT) be incorporated into standard care for HIV-infected individuals[123] aims to decrease TB incidence based on evidence that IPT reduces the risk of progression from latent to active disease in this population[124]. IPT, while not expected to increase prevalence of high-level INH mono-resistance in treated patients[72,125], may impose purifying selection on the Mtb population enriching for isolates that have altered susceptibility to INH, ultimately serving to increase the circulation of low- or high-level INH-resistant strains in the community. Indeed, a mathematical model suggested that community-wide rollout of IPT could drive prevalence and incidence of drug-resistant TB at the community-level, even if the individual-level risk of developing INH-resistant TB is not altered[73]. In view of this, our

findings may expand the scope of poor treatment outcomes associated with INH mono-resistance, and suggest that the role of both low- and high-level changes in INH susceptibility and potential causative genetic polymorphisms merits further consideration, particularly given the expected increase in organisms with altered INH susceptibility in the setting of IPT.

Finally, our data indicate that completion of the standard treatment regimen was not sufficient to fully cure TB in some patients. There are data to suggest that altering treatment regimens may reduce the risk of treatment failure, both in the settings of HIV and INH mono-resistance. Post-treatment INH prophylaxis reduced rates of recurrent infection in HIV-infected individuals[126], and daily dosing during the intensive phase of treatment and longer duration of PZA treatment reduced risks of treatment failure in patients with pretreatment isoniazid resistance[127]. Our study highlights the need for further research into altered treatment regimens designed to suppress relapse in HIV-infected patients, particularly as the scope of INH mono-resistance increases in this population.

4.5 Materials and Methods

Study description

Patients were enrolled in the prospective cohort study, TB Recurrence Upon Treatment with Highly Active Antiretroviral Therapy (TRuTH), at the Center for the AIDS Programme of Research in South Africa (CAPRISA) in KwaZulu Natal. The TRuTH study aims to determine the incidence of recurrent TB infection in patients on highly active antiretroviral therapy (HAART) who were previously treated for a TB infection (registered with ClinicalTrials.gov, number NCT01539005). As described previously, patients were treated for an initial TB infection during their participation in the Starting AIDS Treatment at Three Points in TB Treatment (SAPIT) study, which was designed to determine the most effective timing of HAART during TB treatment[117].

Written informed consent was obtained from all study participants, and the University of KwaZulu-Natal Biomedical Research Ethics Committee approved this study (reference no BF 051/09).

Drug susceptibility testing

The 1% direct proportion method was used to assess phenotypic drug-resistance. Culture conditions and the following critical drug concentrations were used: INH (0.2ug/mL); RIF (40ug/mL); EMB (2ug/mL);

Genomic DNA isolation

Patient sputum was cultured using the BACTEC MGIT 960? system. Cultures were plated on Löwenstein Jensen medium, and colonies were harvested by scraping the entire plate, and resuspending in 7H9 medium. Genomic DNA was isolated from each sample by the CTAB method. Briefly, scraped colonies were heat-killed and incubated in lysozyme, followed by

sodium-dodecyl-sulfate and Proteinase K incubation. Samples were incubated in NaCl-CTAB buffer, followed by chloroform:isoamyl alcohol extraction and isopropanol precipitation.

MIRU-VNTR Typing

24-locus MIRU-VNTR typing was performed by GenoScreen, Inc. (Campus de l'Institut Pasteur de Lille, France).

Illumina sequencing library construction

Libraries were constructed in technical replicates using the Illumina Nextera XT sample preparation kit. Paired-end sequencing was performed in-house on the Illumina MiSeq with a read length of 101 bases, with an average read-depth of 80X and an average genome coverage of 95% for all samples. No discrepancies were found between technical replicates.

Bioinformatic analysis

Reads were processed according to the Broad Institute's GATK guidelines. Briefly, reads were mapped to H37Rv (NC_000962) using bwa mem. Aligned output was coordinate-sorted and duplicate reads were marked using Picard. GATK was used for base-quality recalibration, using a list of known high-quality SNPs for calibration. SAMtools mpileup was used to generate pileup files. Mpileup output was processed in custom Python scripts to call high-confidence SNPs as follows. Repetitive regions of the genome such as PE/PPE, IS*6110* and transposable elements were removed. SNPs and small insertions/deletions were defined as positions in which the variant comprised at least 90% of reads, with a variant read depth of greater than 10, a mapping quality of 60, and a base quality of 30 or higher. There was an average of 80 heterogeneous positions per patient sample. Heterogeneous positions meeting the following criteria were considered for PCR-resequencing validation: positions with a mapping quality of 60 and a base quality of 30 or higher, in which less than 85% and greater than 10% of reads

differed from the reference genome, with the minor variant having a read depth of at least than 7, represented by at least 3 reads on either strand. Positions were discarded if they were within 10bp of other heterogeneous positions or insertions or deletions. This pipeline was validated in a mixing experiment in which H37Rv and Erdman strains were combined at known ratios, and whole-genome sequencing was performed to an average read depth of 80X. 98% of expected heterogeneous sites were identified, and no heterogeneous sites were identified that were not expected. Candidate heterogeneous positions from patient samples that passed the above filters were PCR-validated and those positions that appeared homogeneous by PCR-validation were then additionally checked for unique mapping to the genome by BLAST analysis; reads that mapped to multiple positions were discarded. Remaining positions were visually inspected in Integrated Genome Viewer for any additional artifacts such as abrupt drops in read depth or proximity to many heterogeneous sites. 0-3 heterogeneous sites per patient sample passed the above filters and were confirmed by PCR-validation as heterogeneous.

Drug resistance database

A drug-resistance polymorphisms database (Supplementary Table 4.1-4.3) was compiled with three categories: (1) polymorphisms with experimentally proven impact on MICs, as listed in the TBDreamDB[128] (2) polymorphisms statistically associated with phenotypic drug-resistance in three recent large-scale sequencing studies[119,120,135], and (3) genomic regions associated with a high degree of genetic polymorphism in drug-resistant strains. Lineage-specific polymorphisms and SNPs which fell into repetitive regions were removed. Only isolates from patients who completed successful treatment were considered for genotypic drug-resistance analysis.

CHAPTER FIVE: CONCLUSION

The rapid emergence of drug-resistant strains of Mtb calls for new ways of understanding the processes that lead to drug resistance. The observations that underlie this work suggest that the capacity of mycobacteria to evolve resistance to antibiotics is greater and more complex than previously expected[30,31,32,33,129]. Specifically, recent whole genome sequencing studies showed that mutation rate may not vary with replication rate[30,31,32,33]. However, it has remained to be understood whether and how mutations can be driven by replication-independent processes. Here, we harnessed technology that has not been used widely in mycobacteria to study whether mutation rate is replication-independent and how this might be an important driver of mutation.

We showed that the rate of accumulation of antibiotic-resistant mutants does not vary in accordance with replication rate in a chemostat model of *Mycobacterium smegmatis*. To rule out the possibility that selection against less fit mutants would differentially affect our ability to detect mutants at fast versus slow growth rates, we used genetic barcoding to track the growth kinetics during competition of mutants of different genotypes at two different growth rates. When we applied a mathematical model to fit the resulting growth kinetics, mutant fitness was found to be lower than wild-type but the selection force did not vary with growth rate. These findings suggest that that the mutant accumulation rate is similar at fast versus slow growth rates, supporting a time-based model of mutation put forth by preceding *in vivo* studies[30,31,32,33].

We went on to investigate one possible driver of time-based mutation in our study of metabolically-induced mutation. Our work revealed a new DNA adduct induced by fatty acid metabolism, and an accompanying mutagenesis when nucleotide excision repair was absent. We further showed that metabolism of cholesterol alone was sufficient to induce mutagenesis. Our

work demonstrates that endogenous metabolic processes can drive DNA damage, and that these conditions are also associated with increased mutation rate.

Together our findings suggest that mycobacterial genomic fidelity may be governed by enzymes not strictly involved in replication and repair, but rather could be expanded to processes that can contribute to time-based mutation such as those involved in DNA damage repair, oxidative stress defense, stress-induced mutagenesis, and carbon metabolism. Insight into the molecular mechanisms driving time-based mutation will elaborate the role of contributing pathways. In these studies, it will be important to focus on infection models and conditions that mimic *in vivo* conditions in order to reveal the balance of damage, stress, repair, and mutation over the course of infection. Previous findings in both Msm and Mtb suggest that this balance shifts across conditions; essentiality of DNA damage repair components has been shown to be conditional on environment, strains deficient in these components exhibit conditional mutagenesis, and different mutational spectrums have been observed *in vitro* compared with those observed in clinical isolates *in vivo*[50, 56, 61, 64,65,130].

One biological implication of this expanded view of genomic fidelity is that a wide range of genetic polymorphisms may affect the balance of damage, repair, metabolism, and mutation. Mutator strains of other pathogens are characterized by increases by orders of magnitude in mutation rate and are commonly associated with loss of one genetic process, mismatch repair[131,132,133,134]. Clinical strains of mycobacteria are instead associated with a more modest continuum of mutation rates, but may harbor polymorphisms in a multitude of contributing pathways. The East Asian lineage has been associated with increased mutation rates compared with the Euro American lineage and increased mutation frequencies compared with the East African Indian lineage[14,115]. Beijing lineage strains were associated with higher rates of drug

resistance in the presence of rifampicin and in the mouse model of non-compliance, compared with East African Indian strains[115]. There also exist phenotypically resistant strains for which no genetic resistance has been detected[135]. Further investigation into the molecular mechanisms driving mutation rate will improve our ability to predict whether and where other strains lie on the continuum of increased mutation rates. The clinical consequences of even modest increases in mutation rate become clear when considered in the context of a time-based model of mutation. When Ford *et al* incorporated time-based mutation into a mathematical model to estimate the effects of 5-40X increases in mutation rate, the probability of emergence of multi-drug-resistance in humans was found to be orders of magnitude higher than previously expected based on a replication-based model[14,17] .

Time-based mutation could allow Mtb to maintain genetic diversity during periods of slow growth, enabling continued adaptability in a changing fitness landscape. Clearly Mtb populations withstand major bottlenecks in both space and time. Theories about the co-evolution of Mtb and humans suggest that Mtb has withstood major changes in the structure of its host population as humans have evolved from small, isolated communities to more interconnected, larger, more dense communities[136]. Is time-based mutation a general evolutionary strategy to balance both the need to preserve sites essential for survival, and the need to adapt to a changing environment? An analysis of mutation rates in mammalian species has challenged the prevailing view of the growth-rate dependence of mutation rate[137]. The mutation rate at sites that are predicted to be truly neutral was found to be constant per year across 326 species characterized by considerable differences in generation times. While variability was found across lineages, this variability did not correlate with generation time, suggesting an overall model of time-based mutation. Further work will be required to determine the source of variability in estimates of

mutation rate within and across studies, such as systematic differences in analytic methods, the effect of selection on presence of variants in a population, and sources of true heterogeneity in mutation rates within a population[138].

In mycobacteria, the combined effect of fitness and the ability to detect genetic variants is likely to influence estimates of mutation rate. Work on antibiotic-resistant mutants highlights that mutant fitness can change based on the condition, as the spectrum of rpoB mutants was shown to change upon reduction in pH[139]. Acquisition of antibiotic resistance itself has been associated with increases in mutation rates. For example, upregulation of the error-prone polymerase DnaE2 upon acquisition of Rif resistance may cause further downstream genetic polymorphisms that could affect fitness[140]. Bryant *et al.* showed that an estimate of the mutation rate using whole genome sequencing of isolates from human transmission chains was significantly affected by the inclusion of drug-resistance mutations, suggesting that a clock-like signal of mutation accumulation can be obscured by the presence of variants whose fitness is significantly different than others[33]. Efforts to use neutral reporters of mutation and to gain a better understanding of the effects of fitness will improve our ability to estimate the mutation rate in mycobacteria. In addition, it is unknown the extent to which sputum samples are representative of the entire population diversity within a patient; deeper sampling across multiple sites over the course of an infection will help elucidate the total "cloud" of genetic diversity within a patient. Recent studies in both human and macaque Mtb infections revealed additional heterogeneity detected when performing deep sequencing of sputum samples versus sequencing of individual colonies, suggesting that greater resolution allows detection of additional variants[4,129]. It is likely that sampling from other sites not proximal to airways will reveal variants that are not present in the sputum[4].

Clearly, high-level drug resistance is a major driver of treatment failure in Mtb. Yet further work is required to understand whether mutations that lead to low-level drug resistance also drive treatment failure. In our analysis of recurrent infection in HIV-infected patients in South Africa, we found that over half of cases were true relapses of an earlier infection, ranging from 6 months to 5 years from the start of treatment. There was an enrichment in relapse-associated strains for polymorphisms associated with first-line drug resistance, suggesting that strains associated with relapse may be biologically predisposed to treatment failure because of undiagnosed drug resistance. In a different population, phenotypic INH resistance was found to be enriched in strains associated with relapse, and has been previously associated with treatment failure[10,141]. Deeper genotype/phenotype analyses will be required to establish the genetic basis of drug resistance and the relationship between drug resistance and various parameters governing a strain's success, such as transmission, severity of disease, and mutation rate.

The recent progress in developing new antibiotics for TB represents renewed hope for treating strains recalcitrant to current therapies, and shortening and simplifying existing treatments[142]. Critical to predicting the success of these treatments will be a more thorough analysis of the rates and genetic processes driving drug resistance. Our work and the work of others suggests that the capacity for mycobacteria to acquire genetic diversity during different conditions may be higher and more complex than previously appreciated.

References

[1] World Health Organization Global Tuberculosis Report 2010

[2] Gengenbacher M, Kaufmann SH. Mycobacterium tuberculosis: success through dormancy. FEMS Microbiol Rev. 2012 May;36(3):514-32.

[3] Barry CE 3rd, Boshoff HI, Dartois V, Dick T, Ehrt S, Flynn J, Schnappinger D, Wilkinson RJ, Young D. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies.Nat Rev Microbiol. 2009 Dec;7(12):845-55.

[4] Lin PL, Ford CB, Coleman MT, Myers AJ, Gawande R, Ioerger T, Sacchettini J, Fortune SM, Flynn JL. Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing.Nat Med. 2014 Jan;20(1):75-9.

[5] Robertson BD, Altmann D, Barry C et al. Detection and treatment of subclinical tuberculosis. Tuberculosis (Edinb) 2012; 92: 447–52.

[6] WHO Global Tuberculosis Report 2013 (http://apps.who.int/iris/bitstream/10665/91355/1/9789241564656_eng.pdf)

[7] Sonnenberg P, Murray J, Glynn JR, Shearer S, Kambashi B, Godfrey-Faussett P., HIV-1 and recurrence, relapse, and reinfection of tuberculosis after cure: a cohort study in South African mineworkers. Lancet. 2001 Nov 17;358(9294):1687-93

[8] Cox H, Kebede Y, Allamuratova S, Ismailov G, Davletmuratova Z, Byrnes G, Stone C, Niemann S, Rüsch-Gerdes S, Blok L, Doshetov D. Tuberculosis recurrence and mortality after successful treatment: impact of drug resistance.PLoS Med. 2006 Oct;3(10):e384.

[9] Datiko DG, Lindtjørn B. Tuberculosis recurrence in smear-positive patients cured under DOTS in southern Ethiopia: retrospective cohort study. BMC Public Health. 2009 Sep 18;9:348.

[10] Guerra-Assunção JA, Houben RM, Crampin AC, Mzembe T, Mallard K, Coll F, Khan P, Banda L, Chiwaya A, Pereira RP, McNerney R, Harris D, Parkhill J, Clark TG, Glynn JR. Recurrence due to Relapse or Reinfection With Mycobacterium tuberculosis: A Whole-Genome Sequencing Approach in a Large, Population-Based Cohort With a High HIV Infection Prevalence and Active Follow-up.J Infect Dis. 2014 Oct 21. pii: jiu574.

[11] Crampin AC, Mwaungulu JN, Mwaungulu FD, Mwafulirwa DT, Munthali K, Floyd S, Fine PE, Glynn JR. Recurrent TB: relapse or reinfection? The effect of HIV in a general population cohort in Malawi. AIDS. 2010 Jan 28;24(3):417-26. doi: 10.1097/QAD.0b013e32832f51cf.

[12] Verver S, Warren RM, Beyers N, Richardson M, van der Spuy GD, Borgdorff MW, Enarson DA, Behr MA, van Helden PD. Rate of reinfection tuberculosis after successful treatment is

higher than rate of new tuberculosis. Am J Respir Crit Care Med. 2005 Jun 15;171(12):1430-5. Epub 2005 Apr 14.

[13] Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature. 1998 Jun 11;393(6685):537-44.

[14] Ford CB, Shah RR, Maeda MK, Gagneux S, Murray MB, Cohen T, Johnston JC, Gardy J, Lipsitch M, Fortune SM. Mycobacterium tuberculosis mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. Nat Genet. 2013 Jul;45(7):784-90.

[15] McGrath M, Gey van Pittius NC, van Helden PD, Warren RM, Warner DF. J Antimicrob Chemother. Mutation rate and the emergence of drug resistance in Mycobacterium tuberculosis. 2014 Feb;69(2):292-302.

[16] Mizrahi V1, Andersen SJ. DNA repair in Mycobacterium tuberculosis. What have we learnt from the genome sequence?Mol Microbiol. 1998 Sep;29(6):1331-9.

[17] Lipsitch M, Levin BR. Population dynamics of tuberculosis treatment: mathematical models of the roles of non-compliance and bacterial heterogeneity in the evolution of drug resistance. Int J Tuberc Lung Dis. 1998 Mar;2(3):187-99.

[18] Zhao Y1, Xu S, Wang L, Chin DP, Wang S, Jiang G, Xia H, Zhou Y, Li Q, Ou X, Pang Y, Song Y, Zhao B, Zhang H, He G, Guo J, Wang Y. National survey of drug-resistant tuberculosis in China. N Engl J Med. 2012 Jun 7;366(23):2161-70.

[19] Ioerger TR, Feng Y, Chen X, Dobos KM, Victor TC, Streicher EM, Warren RM, Gey van Pittius NC, Van Helden PD, Sacchettini JC. The non-clonality of drug resistance in Beijing-genotype isolates of Mycobacterium tuberculosis from the Western Cape of South Africa. BMC Genomics. 2010 Nov 26;11:670.

[20] Victor TC1, Streicher EM, Kewley C, Jordaan AM, van der Spuy GD, Bosman M, Louw H, Murray M, Young D, van Helden PD, Warren RM. Spread of an emerging Mycobacterium tuberculosis drug-resistant strain in the western Cape of South Africa. Int J Tuberc Lung Dis. 2007 Feb;11(2):195-201.

[21] Ioerger TR1, Koo S, No EG, Chen X, Larsen MH, Jacobs WR Jr, Pillay M, Sturm AW, Sacchettini JC. Genome analysis of multi- and extensively-drug-resistant tuberculosis from KwaZulu-Natal, South Africa. PLoS One. 2009 Nov 5;4(11):e7778.

[22] Colijn C, Cohen T, Ganesh A, Murray M. Spontaneous emergence of multiple drug resistance in tuberculosis before and during therapy. PLoS One. 2011 Mar 30;6(3):e18327.

[23] Ford C1, Yusim K, Ioerger T, Feng S, Chase M, Greene M, Korber B, Fortune S. Mycobacterium tuberculosis--heterogeneity revealed through whole genome sequencing. Tuberculosis (Edinb). 2012 May;92(3):194-201

[24] Gagneux S1, Small PM. Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. Lancet Infect Dis. 2007 May;7(5):328-37.

[25] Achtman M Evolution, Population Structure, and Phylogeography of Genetically Monomorphic Bacterial Pathogens. Annu Rev Microbiol. 2008;62:53-70.

[26] de Steenwinkel JE, ten Kate MT, de Knegt GJ, Kremer K, Aarnoutse RE, Boeree MJ, Verbrugh HA, van Soolingen D, Bakker-Woudenberg IA. Drug susceptibility of Mycobacterium tuberculosis Beijing genotype and association with MDR TB. Emerg Infect Dis. 2012 Apr;18(4):660-3.

[27] Drobniewski F, Balabanova Y, Nikolayevsky V, Ruddy M, Kuznetzov S, Zakharova S, Melentyev A, Fedorin I. Drug-resistant tuberculosis, clinical virulence, and the dominance of the Beijing strain family in Russia. JAMA. 2005 Jun 8;293(22):2726-31.

[28] Glynn JR, Alghamdi S, Mallard K, McNerney R, Ndlovu R, Munthali L, Houben RM, Fine PE, French N, Crampin AC. Changes in Mycobacterium tuberculosis genotype families over 20 years in a population-based study in Northern Malawi. PLoS One. 2010 Aug 17;5(8):e12259.

[29] Capuano SV 3rd, Croix DA, Pawar S, Zinovik A, Myers A, Lin PL, Bissel S, Fuhrman C, Klein E, Flynn JL. Experimental Mycobacterium tuberculosis infection of cynomolgus macaques closely resembles the various manifestations of human M. tuberculosis infection. Infect Immun. 2003 Oct;71(10):5831-44.

[30] Ford CB, Lin PL, Chase MR, Shah RR, Iartchouk O, Galagan J, Mohaideen N, Ioerger TR, Sacchettini JC, Lipsitch M, Flynn JL, Fortune SM. Use of whole genome sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection.Nat Genet. 2011 May;43(5):482-6.

[31] Gardy JL1, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodkin E, Rempel S, Moore R, Zhao Y, Holt R, Varhol R, Birol I, Lem M, Sharma MK, Elwood K, Jones SJ, Brinkman FS, Brunham RC, Tang P. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N Engl J Med. 2011 Feb 24;364(8):730-9.

[32] Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, Eyre DW, Wilson DJ, Hawkey PM, Crook DW, Parkhill J, Harris D, Walker AS, Bowden R, Monk P, Smith EG, Peto TE. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis. 2013 Feb;13(2):137-46.

[33] Bryant JM, Schürch AC, van Deutekom H, Harris SR, de Beer JL, de Jager V, Kremer K, van Hijum SA, Siezen RJ, Borgdorff M, Bentley SD, Parkhill J, van Soolingen D. Inferring patient to patient transmission of Mycobacterium tuberculosis from whole genome sequencing data. BMC Infect Dis. 2013 Feb 27;13:110.

[34] Lenaerts A, Barry CE 3rd, Dartois V. Heterogeneity in tuberculosis pathology, microenvironments and therapeutic responses. Immunol Rev. 2015 Mar;264(1):288-307.

[35] Gill WP, Harik NS, Whiddon MR, Liao RP, Mittler JE, Sherman DR. A replication clock for Mycobacterium tuberculosis. Nat Med. 2009 Feb;15(2):211-4.

[36] Muñoz-Elías EJ, Timm J, Botha T, Chan WT, Gomez JE, McKinney JD. Replication dynamics of Mycobacterium tuberculosis in chronically infected mice. Infect Immun. 2005 Jan;73(1):546-51.

[37] Barry S1, Breen R, Lipman M, Johnson M, Janossy G. Impaired antigen-specific CD4(+) T lymphocyte responses in cavitary tuberculosis. Tuberculosis (Edinb). 2009 Jan;89(1):48-53.

[38] Lin PL, Flynn JL. Understanding latent tuberculosis: a moving target. J Immunol. 2010 Jul 1;185(1):15-22.

[39] Nicholson S, Bonecini-Almeida Mda G, Lapa e Silva JR, Nathan C, Xie QW, Mumford R, Weidner JR, Calaycay J, Geng J, Boechat N, Linhares C, Rom W, Ho JL. Inducible nitric oxide synthase in pulmonary alveolar macrophages from patients with tuberculosis. J Exp Med. 1996 May 1;183(5):2293-302.

[40] Nathan C, Shiloh MU. Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens. Proc Natl Acad Sci U S A. 2000 Aug 1;97(16):8841-8.

[41] VanderVen BC, Yates RM, Russell DG. Intraphagosomal measurement of the magnitude and duration of the oxidative burst. Traffic. 2009;10:372–378.

[42] MacMicking JD, North RJ, LaCourse R, Mudgett JS, Shah SK, Nathan CF. Identification of nitric oxide synthase as a protective locus against tuberculosis. Proc Natl Acad Sci U S A. 1997;94:5243–5248.

[43] Schnappinger D, Ehrt S, Voskuil MI, Liu Y, Mangan JA, Monahan IM, Dolganov G, Efron B, Butcher PD, Nathan C, et al. Transcriptional Adaptation of Mycobacterium tuberculosis within Macrophages: Insights into the Phagosomal Environment. J Exp Med. 2003;198:693–704.

[44] Shi L1, Sohaskey CD, North RJ, Gennaro ML. Transcriptional characterization of the antioxidant response of Mycobacterium tuberculosis in vivo and during adaptation to hypoxia in vitro. Tuberculosis (Edinb). 2008 Jan;88(1):1-6.

[45] Warner DF, Mizrahi V. Tuberculosis chemotherapy: the influence of bacillary stress and damage response pathways on drug efficacy. Clin Microbiol Rev. 2006 Jul;19(3):558-70.

[46] Ng VH, Cox JS, Sousa AO, MacMicking JD, McKinney JD. Role of KatG catalase-peroxidase in mycobacterial pathogenesis: countering the phagocyte oxidative burst. Mol Microbiol. 2004 Jun;52(5):1291-302.

[47] Piddington DL1, Fang FC, Laessig T, Cooper AM, Orme IM, Buchmeier NA. Cu,Zn superoxide dismutase of Mycobacterium tuberculosis contributes to survival in activated macrophages that are generating an oxidative burst. Infect Immun. 2001 Aug;69(8):4980-7.

[48] Kurthkoti et al. A distinct physiological role of MutY in mutation prevention in mycobacteria. Microbiology (2010) vol. 156 (1) pp. 88-93

[49] Jain et al. A distinct role of formamidopyrimidine DNA glycosylase (MutM) in down-regulation of accumulation of G, C mutations and protection against oxidative stress in mycobacteria. DNA Repair (Amst) (2007). vol. 6 (12) pp. 1774-85

[50] Wanner et al. The Uracil DNA Glycosylase UdgB of Mycobacterium smegmatis Protects the Organism from the Mutagenic Effects of Cytosine and Adenine Deamination. Journal of Bacteriology (2009) vol. 191 (20) pp. 6312-6319

[51] Sidorenko VS1, Rot MA, Filipenko ML, Nevinsky GA, Zharkov DO. Novel DNA glycosylases from Mycobacterium tuberculosis. Biochemistry (Mosc). 2008 Apr;73(4):442-50.

[52] Guo Y1, Bandaru V, Jaruga P, Zhao X, Burrows CJ, Iwai S, Dizdaroglu M, Bond JP, Wallace SS. The oxidative DNA glycosylases of Mycobacterium tuberculosis exhibit different substrate preferences from their Escherichia coli counterparts. DNA Repair (Amst). 2010 Feb 4;9(2):177-90.

[53] Darwin KH, Ehrt S, Gutierrez-Ramos JC, Weich N, Nathan CF. The proteasome of Mycobacterium tuberculosis is required for resistance to nitric oxide. Science. 2003 Dec 12;302(5652):1963-6.

[54] Güthlein C1, Wanner RM, Sander P, Davis EO, Bosshard M, Jiricny J, Böttger EC, Springer B. Characterization of the mycobacterial NER system reveals novel functions of the uvrD1 helicase. J Bacteriol. 2009 Jan;191(2):555-62.

[55] Springer B, Sander P, Sedlacek L et al. Lack of mismatch correction facilitates genome evolution in mycobacteria. Mol Microbiol 2004; 53: 1601 – 9.

[56] Venkatesh J, Kumar P, Krishna PS, Manjunath R, Varshney U. Importance of uracil DNA glycosylase in Pseudomonas aeruginosa and Mycobacterium smegmatis, G+C-rich bacteria, in mutation prevention, tolerance to acidified nitrite, and endurance in mouse macrophages. J Biol Chem. 2003 Jul 4;278(27):24350-8.

[57] Kurthkoti et al. A distinct physiological role of MutY in mutation prevention in mycobacteria. Microbiology (2010) vol. 156 (1) pp. 88-93

[58] Dos Vultos T, Blázquez J, Rauzier J, Matic I, Gicquel B. Identification of Nudix hydrolase family members with an antimutator role in Mycobacterium tuberculosis and Mycobacterium smegmatis. J Bacteriol. 2006 Apr;188(8):3159-61.

[59] Darwin KH, Nathan CF. Role for nucleotide excision repair in virulence of Mycobacterium tuberculosis. Infect Immun. 2005 Aug;73(8):4581-7.

[60] Houghton J, Townsend C, Williams AR, Rodgers A, Rand L, Walker KB, Böttger EC, Springer B, Davis EO. Important role for Mycobacterium tuberculosis UvrD1 in pathogenesis and persistence apart from its function in nucleotide excision repair. J Bacteriol. 2012 Jun;194(11):2916-23.

[61] Sassetti CM, Rubin EJ. Genetic requirements for mycobacterial survival during infection. Proc Natl Acad Sci U S A. 2003 Oct 28;100(22):12989-94.

[62] Dutta NK, Mehra S, Didier PJ, Roy CJ, Doyle LA, Alvarez X, Ratterree M, Be NA, Lamichhane G, Jain SK, Lacey MR, Lackner AA, Kaushal D. Genetic requirements for the survival of tubercle bacilli in primates. J Infect Dis. 2010 Jun 1;201(11):1743-52.

[63] Rengarajan J1, Bloom BR, Rubin EJ. Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages. Proc Natl Acad Sci U S A. 2005 Jun 7;102(23):8327-32. Epub 2005 May 31.

[64] Griffin JE, Gawronski JD, Dejesus MA, Ioerger TR, Akerley BJ, Sassetti CM. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog. 2011 Sep;7(9):e1002251.

[65] Beste DJ1, Espasa M, Bonde B, Kierzek AM, Stewart GR, McFadden J. The genetic requirements for fast and slow growth in mycobacteria. PLoS One. 2009 Apr 28;4(4):e5349.

[66] Boshoff HI, Reed MB, Barry CE 3rd, Mizrahi V. DnaE2 polymerase contributes to in vivo survival and the emergence of drug resistance in Mycobacterium tuberculosis. Cell. 2003 Apr 18;113(2):183-93.

[67] Halliwell, B. & J. M. C. Gutteridge, (2007) *Free Radicals in Biology and Medicine,* p. 851. Oxford University Press, Oxford.

[68] Jain R, Kumar P, Varshney U. A distinct role of formamidopyrimidine DNA glycosylase (MutM) in down-regulation of accumulation of G, C mutations and protection against oxidative stress in mycobacteria. DNA Repair. 2007;6:1774–1785.

[69] MDR Tuberculosis Update (WHO, 2014) http://www.who.int/tb/challenges/mdr/mdr_tb_factsheet.pdf?ua=1

[70] Gutierrez-Vazquez, J.M. Studies on the rate of growth of mycobacteria. I. Generation time of *Mycobacterium tuberculosis* on several solid and liquid media and effects exerted by glycerol and malachite green. *Am. Rev. Tuberc.* 74, 50–58 (1956).

[71] Lee, J., Remold, H.G., Ieong, M.H. & Kornfeld, H. Macrophage apoptosis in response to high intracellular burden of *Mycobacterium tuberculosis* is mediated by a novel caspase-independent pathway. *J. Immunol.* 176, 4267–4274 (2006).

[72] Balcells ME1, Thomas SL, Godfrey-Faussett P, Grant AD. Isoniazid preventive therapy and risk for resistant tuberculosis. Emerg Infect Dis. 2006 May;12(5):744-51.

[73] H. L. Mills, T. Cohen, C. Colijn, Community-wide isoniazid preventive therapy drives drug-resistant tuberculosis: A model-based analysis. Sci. Transl. Med. 5, 180ra49 (2013).

[74] NOVICK and SZILARD. Experiments with the Chemostat on Spontaneous Mutations of Bacteria. Proc Natl Acad Sci (1950) USA vol. 36 (12) pp. 708-19

[75] Rosenberg SM, Shee C, Frisch RL, Hastings PJ. Stress-induced mutation via DNA breaks in Escherichia coli: a molecular mechanism with implications for evolution and medicine. Bioessays. 2012 Oct;34(10):885-92.

[76] Saumaa J. Oxidative DNA damage defense systems in avoidance of stationary-phase mutagenesis in Pseudomonas putida.Bacteriol. 2007 Aug;189(15):5504-14.

[77] Dos Vultos T1, Mestre O, Tonjum T, Gicquel B. DNA repair in Mycobacterium tuberculosis revisited. FEMS Microbiol Rev. 2009 May;33(3):471-87.

[78] Foster PL1. Methods for determining spontaneous mutation rates. Methods Enzymol. 2006;409:195-213.

[79] Jenkins C1, Bacon J, Allnutt J, Hatch KA, Bose A, O'Sullivan DM, Arnold C, Gillespie SH, McHugh TD. Enhanced heterogeneity of rpoB in Mycobacterium tuberculosis found at low pH. J Antimicrob Chemother. 2009 Jun;63(6):1118-20.

[80] Monod J. The Growth of Bacterial Cultures. Annu. Rev. Microbiol. 1949.3:371-394.

[81]Panikov, N (1995) *Microbial Growth Kinetics*, Chapman & Hall

[82] Wick LM1, Weilenmann H, Egli T. The apparent clock-like evolution of Escherichia coli in glucose-limited chemostats is reproducible at large but not at small population sizes and can be explained with Monod kinetics. Microbiology. 2002 Sep;148(Pt 9):2889-902.

[83] Koch A, Mizrahi V., Warner, DF. The impact of drug resistance on *Mycobacterium tuberculosis* physiology: what can we learn from rifampicin? *Emerging Microbes & Infections* (2014) 3, e17;

[84] Gagneux S, Long CD, Small PM, et al. The competitive cost of antibiotic resistance in Mycobacterium tuberculosis. Science 2006;312:1944-6.

[85] Rock JM, Lang UF, Chase MR, Ford CB, Gerrick ER, Gawande R, Coscolla M, Gagneux S, Fortune SM, Lamers MH. DNA replication fidelity in Mycobacterium tuberculosis is mediated by an ancestral prokaryotic proofreader. Nat Genet. 2015 Apr 20. doi: 10.1038/ng.3269. [Epub ahead of print]

[86] Herring CD, Raghunathan A, Honisch C *et al*. Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nat Genet* 2006; 38: 1406–1412.

[87] Katz S, Hershberg R. Elevated mutagenesis does not explain the increased frequency of antibiotic resistant mutants in starved aging colonies. *PLOS Genet* 2013; 9: e1003968.

[88] Blumenthal A, Trujillo C, Ehrt S, Schnappinger D. Simultaneous analysis of multiple Mycobacterium tuberculosis knockdown mutants in vitro and in vivo. PLoS One. 2010 Dec 22;5(12):e15667.

[89] Monod J. The Growth of Bacterial Cultures. Annu. Rev. Microbiol. 1949.3:371-394.

[90] Moellering RE, Cravatt BF. Functional lysine modification by an intrinsically reactive primary glycolytic metabolite. Science. 2013 Aug 2;341(6145):549-53.

[91] Weinert BT1, Schölz C, Wagner SA, Iesmantavicius V, Su D, Daniel JA, Choudhary C. Lysine succinylation is a frequently occurring modification in prokaryotes and eukaryotes and extensively overlaps with acetylation. Cell Rep. 2013 Aug 29;4(4):842-51.

[92] Byrns, M.C., et al., Detection of DNA adducts derived from the reactive metabolite of furan, cis- 2-butene-1,4-dial. Chemical Research in Toxicology, 2006. 19(3): p. 414-420.

[93] Blair, I.A., DNA adducts with lipid peroxidation products. J Biol Chem, 2008. 283(23): p. 15545-9.

[94] McKinney, J.D., et al., Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. Nature, 2000. 406(6797): p. 735-8.

[95] Eastmond, P.J., et al., Postgerminative growth and lipid catabolism in oilseeds lacking the glyoxylate cycle. Proc Natl Acad Sci U S A, 2000. 97(10): p. 5669-74.

[96] Kornberg, H.L. and H. Beevers, The Glyoxylate Cycle as a Stage in the Conversion of Fat to Carbohydrate in Castor Beans. Biochimica Et Biophysica Acta, 1957. 26(3): p. 531-537.

[97] Moeller, B.C., et al., Biomarkers of exposure and effect in human lymphoblastoid TK6 cells following [13C2]-acetaldehyde exposure. Toxicol Sci, 2013. 133(1): p. 1-12.

[98] Lu, K., et al., Molecular dosimetry of N2-hydroxymethyl-dG DNA adducts in rats exposed to formaldehyde. Chem Res Toxicol, 2011. 24(2): p. 159-61.

[99] Sayato Y1, Nakamuro K, Ueno H. Mutagenicity of products formed by ozonation of naphthoresorcinol in aqueous solutions. Mutat Res. 1987 Nov;189(3):217-22.

[100] Jumpathong, Watthanachai (2014). The dynamic interplay between DNA damage and metabolism: The metabolic fate and transport of DNA lesions and novel DNA damage derived from intermediary metabolism (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

[101] Campbell W. J., Ogren W. L. Glyoxylate inhibition of ribulosebisphosphate carboxylase/oxygenase activation in intact, lysed, and reconstituted chloroplasts. Photosynth. Res. 1990;23:257–268.

[102] Hausler R. E., Bailey K. J., Lea P. J., Leegood R. C. Control of photosynthesis in barley mutants with reduced activities of glutamine synthetase and glutamate synthase. III. Aspects of glyoxylate metabolism and effects of glyoxylate on the activation state of ribulose-1,5-bisphosphate carboxylase-oxygenase. Planta. 1996;200:388–396.

[103] Allan W. L., Simpson J. P., Clark S. M., Shelp B. J. γ-Hydroxybutyrate accumulation in *Arabidopsis* and tobacco plants is a general response to abiotic stress: putative regulation by redox balance and glyoxylate reductase isoforms. J. Exp. Bot. 2008;59:2555–2564.

[104] Wanner RM1, Castor D, Güthlein C, Böttger EC, Springer B, Jiricny J. The uracil DNA glycosylase UdgB of Mycobacterium smegmatis protects the organism from the mutagenic effects of cytosine and adenine deamination. J Bacteriol. 2009 Oct;191(20):6312-9.

[105] Malshetty VS1, Jain R, Srinath T, Kurthkoti K, Varshney U. Synergistic effects of UdgB and Ung in mutation prevention and protection against commonly encountered DNA damaging agents in Mycobacterium smegmatis. Microbiology. 2010 Mar;156(Pt 3):940-9.

[106] Griffin JE, Pandey AK, Gilmore SA, Mizrahi V, McKinney JD, Bertozzi CR, Sassetti CM. Cholesterol catabolism by Mycobacterium tuberculosis requires transcriptional and metabolic adaptations. Chem Biol. 2012 Feb 24;19(2):218-27.

[107] Savvi S1, Warner DF, Kana BD, McKinney JD, Mizrahi V, Dawes SS. Functional characterization of a vitamin B12-dependent methylmalonyl pathway in Mycobacterium

tuberculosis: implications for propionate metabolism during growth on fatty acids. J Bacteriol. 2008 Jun;190(11):3886-95.

[108] Upton AM, McKinney JD. Role of the methylcitrate cycle in propionate metabolism and detoxification in Mycobacterium smegmatis. Microbiology. 2007 Dec;153(Pt 12):3973-82.

[109] Chang JC, Miner MD, Pandey AK, Gill WP, Harik NS, Sassetti CM, Sherman DR. igr Genes and Mycobacterium tuberculosis cholesterol metabolism. J Bacteriol. 2009 Aug;191(16):5232-9.

[110] Muñoz-Elías EJ1, Upton AM, Cherian J, McKinney JD. Role of the methylcitrate cycle in Mycobacterium tuberculosis metabolism, intracellular growth, and virulence. Mol Microbiol. 2006 Jun;60(5):1109-22.

[111] van Kessel JC1, Hatfull GF. Mycobacterial recombineering. Methods Mol Biol. 2008;435:203-15.

[112] Lee M.H., Pascopella,L., Jacobs,W.R. and Hatfull,G.F. (1991) Site specific integration of mycobacteriophage L5: integration-proficient vectors for *Mycobacterium smegmatis,Mycobacterium tuberculosis*, and bacilli Calmette–Guerin. Proc. Natl Acad. Sci. USA, 88, 3111–3115.

[113] Lang, GI, Murray, AM. Estimating the Per-Base-Pair Mutation Rate in the Yeast *Saccharomyces cerevisiae*. Genetics. 2008 Jan; 178(1): 67–82.

[114] WHO Global Tuberculosis Report 2013 (http://apps.who.int/iris/bitstream/10665/91355/1/9789241564656_eng.pdf)

[115] McGreevy J1, Jean Juste MA, Severe P, Collins S, Koenig S, Pape JW, Fitzgerald DW. Outcomes of HIV-infected patients treated for recurrent tuberculosis with the standard retreatment regimen. Int J Tuberc Lung Dis. 2012 Jun;16(6):841-5.

[116] Korenromp EL1, Scano F, Williams BG, Dye C, Nunn P. Effects of human immunodeficiency virus infection on recurrence of tuberculosis after rifampin-based treatment: an analytical review. Clin Infect Dis. 2003 Jul 1;37(1):101-12. Epub 2003 Jun 23.

[117] Abdool Karim SS1, Naidoo K, Grobler A, Padayatchi N, Baxter C, Gray A, Gengiah T, Nair G, Bamber S, Singh A, Khan M, Pienaar J, El-Sadr W, Friedland G, Abdool Karim Q. Timing of initiation of antiretroviral drugs during tuberculosis therapy. N Engl J Med. 2010 Feb 25;362(8):697-706.

[118] Panjabi R1, Comstock GW, Golub JE. Recurrent tuberculosis and its risk factors: adequately treated patients are still at high risk. Int J Tuberc Lung Dis. 2007 Aug;11(8):828-37.

[119] Casali N, Nikolayevskyy V, Balabanova Y, Harris SR, Ignatyeva O, Kontsevaya I, Corander J, Bryant J, Parkhill J, Nejentsev S, Horstmann RD, Brown T, Drobniewski F. Evolution and

transmission of drug-resistant tuberculosis in a Russian population. Nat Genet. 2014 Mar;46(3):279-86.

[120] Zhang H1, Li D, Zhao L, Fleming J, Lin N, Wang T, Liu Z, Li C, Galwey N, Deng J, Zhou Y, Zhu Y, Gao Y, Wang T, Wang S, Huang Y, Wang M, Zhong Q, Zhou L, Chen T, Zhou J, Yang R, Zhu G, Hang H, Zhang J, Li F, Wan K, Wang J, Zhang XE, Bi L. Genome sequencing of 161 Mycobacterium tuberculosis isolates from China identifies genes and intergenic regions associated with drug resistance. Nat Genet. 2013 Oct;45(10):1255-60.

[121] Pepperell CS, Casto AM, Kitchen A, Granka JM, Cornejo OE, Holmes EC, Birren B, Galagan J, Feldman MW. The role of selection in shaping diversity of natural M. tuberculosis populations. PLoS Pathog. 2013 Aug;9(8):e1003543.

[122] WHO, 2008 Anti-Tuberculosis Drug Resistance in the World Report (http://apps.who.int/iris/bitstream/10665/43889/1/WHO_HTM_TB_2008.394_eng.pdf?ua=1&ua=1)

[123] WHO 2011 Report (http://www.who.int/hiv/pub/tb/9789241500708/en/)

[124] Akolo C, Adetifa I, Shepperd S, Volmink J. Treatment of latent tuberculosis infection in HIV infected persons. Cochrane Database Syst Rev. 2010 Jan 20;(1):CD000171.

[125] C. L. van Halsema, K. L. Fielding, V. N. Chihota, E. C. Russell, J. J. Lewis, G. J. Churchyard, A. D. Grant, Tuberculosis outcomes and drug susceptibility in individuals exposed to isoniazid preventive therapy in a high HIV prevalence setting. AIDS 24, 1051–1055 (2010)

[126] Fitzgerald DW, Desvarieux M, Severe P, Joseph P, Johnson WD Jr, Pape JW. Lancet. 2000 Oct 28;356(9240):1470-4.

[127] Menzies D1, Benedetti A, Paydar A, Royce S, Madhukar P, Burman W, Vernon A, Lienhardt C. Standardized treatment of active tuberculosis in patients with previous treatment and/or with mono-resistance to isoniazid: a systematic review and meta-analysis. PLoS Med. 2009 Sep;6(9):e1000150.

[128] Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB. (2009) Tuberculosis Drug Resistance Mutation Database. PLoS Med 6(2): e1000002. doi:10.1371/journal.pmed.1000002.

[129] Sun G1, Luo T, Yang C, Dong X, Li J, Zhu Y, Zheng H, Tian W, Wang S, Barry CE 3rd, Mei J, Gao Q. Dynamic population changes in Mycobacterium tuberculosis during acquisition and fixation of drug resistance in patients. J Infect Dis. 2012 Dec 1;206(11):1724-33.

[130] Bergval I, Coll F, Schuitema A, de Ronde H, Mallard K, Pain A, McNerney R, Clark TG, Anthony RM. A proportion of mutations fixed in the genomes of in vitro selected isogenic drug-

resistant Mycobacterium tuberculosis mutants can be detected as minority variants in the parent culture. FEMS Microbiol Lett. 2015 Jan;362(2):1-7.

[131] Chopra, I., A. J. O'Neill, and K. Miller. 2003. The role of mutators in the emergence of antibiotic-resistant bacteria. Drug Resist. Updates 6:137-145.

[132] LeClerc, J. E., B. Li, W. L. Payne, and T. A. Cebula. 1996. High mutation frequencies among Escherichia coli and Salmonella pathogens. Science 274:1208-1211

[133] Oliver, A., R. Canton, P. Campo, F. Baquero, and J. Blazquez. 2000. High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection. Science 288:1251-1254.

[134] Tanaka, M. M., C. T. Bergstrom, and B. R. Levin. 2003. The evolution of mutator genes in bacterial populations: the roles of environmental change and timing.

[135] Farhat MR, Shapiro BJ, Kieser KJ, Sultana R, Jacobson KR, Victor TC, Warren RM, Streicher EM, Calver A, Sloutsky A, Kaur D, Posey JE, Plikaytis B, Oggioni MR, Gardy JL, Johnston JC, Rodrigues M, Tang PK, Kato-Maeda M, Borowsky ML, Muddukrishna B, Kreiswirth BN, Kurepina N, Galagan J, Gagneux S, Birren B, Rubin EJ, Lander ES, Sabeti PC, Murray M. Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. Nat Genet. 2013 Oct;45(10):1183-9.

[136] Brites D, Gagneux S. Co-evolution of Mycobacterium tuberculosis and Homo sapiens. Immunol Rev. 2015 Mar;264(1):6-24.

[137] Kumar S1, Subramanian S. Mutation rates in mammalian genomes. Proc Natl Acad Sci U S A. 2002 Jan 22;99(2):803-8.

[138] Callaway E. DNA clock proves tough to set. Nature. 2015 Mar 12;519(7542):139-40.

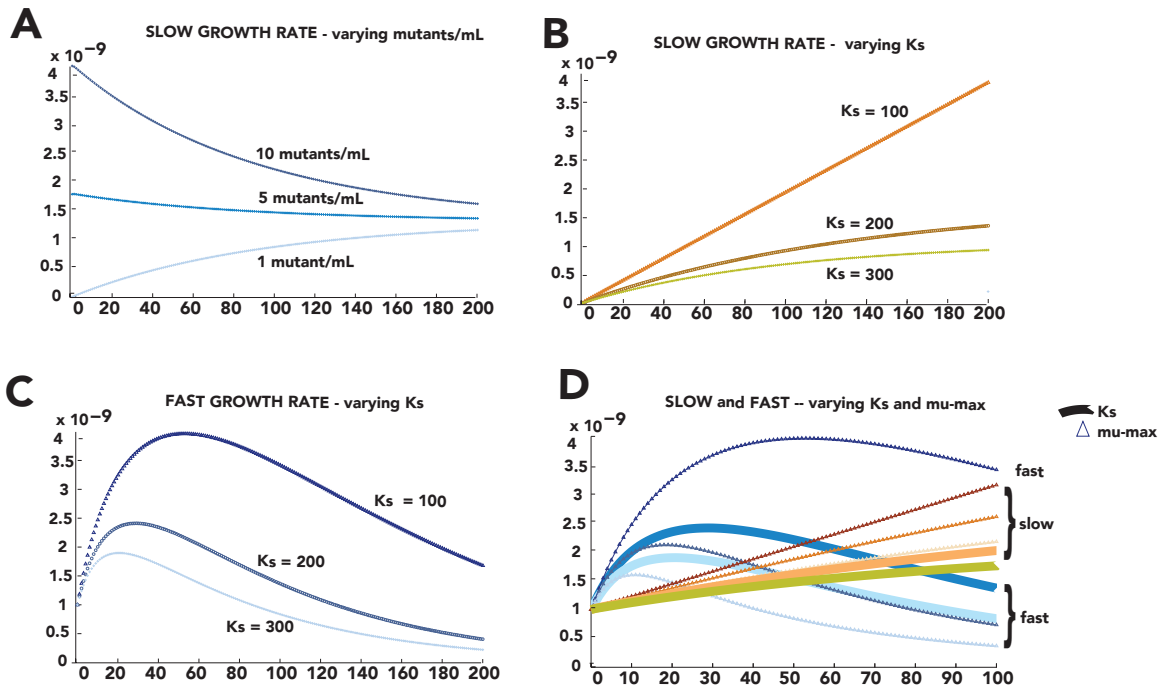
[139] Jenkins C1, Bacon J, Allnutt J, Hatch KA, Bose A, O'Sullivan DM, Arnold C, Gillespie SH, McHugh TD. Enhanced heterogeneity of rpoB in Mycobacterium tuberculosis found at low pH. J Antimicrob Chemother. 2009 Jun;63(6):1118-20. doi: 10.1093/jac/dkp125. Epub 2009 Apr 15.

[140] FEMS Microbiol Lett. 2007 Oct;275(2):338-43. Epub 2007 Sep 14.
Specific mutations in the Mycobacterium tuberculosis rpoB gene are associated with increased dnaE2 expression.
Bergval IL1, Klatser PR, Schuitema AR, Oskam L, Anthony RM.

[141] Mitchison DA, Nunn AJ. Influence of initial drug resistance on the response to short- course chemotherapy of pulmonary tuberculosis. Am Rev Respir Dis 1986;133:423-30.

[142] Zumla AI, Gillespie SH, Hoelscher M, Philips PP, Cole ST, Abubakar I, McHugh TD, Schito M, Maeurer M, Nunn AJ. New antituberculosis drugs, regimens, and adjunct therapies: needs, advances, and future prospects. Lancet Infect Dis. 2014 Apr;14(4):327-40.

APPENDIX 1



Supplementary Figure 2.1 Growth kinetics of mutants in a simple two-strain simulation of bacterial growth using the Monod model. a) Initial concentration of mutants (as indicated) determines collapse of population. b and c) Rate of approach to steady state is proportional to fitness (Ks and μ_{max} were varied as indicated) in fast b) and slow c) growth rates. d) Varying Ks and mu-max at fast and slow growth rates shows that each parameter is differentially sensitive to growth rate – fast growth rate is more sensitive to varying μ_{max}, slow growth rate is more sensitive to varying Ks. Y-axis is biomass of the mutant strain in mg.

Supplementary Table 4.1 Polymorphisms listed in TBDreamDB for which there have been documented increases in MIC (Ref. 128)

Gene Name	Nucleotide Position	Polymorphism	Estimated Codon Position
rpoB		CCT/CTT	564
rpoB		GAA/GGA	562
rpoB		ACC/CCC	508
rpoB		GGC/AGC	507
rpoB		ACC/GCC	508
katG	823	ACC/CCC	275
rpoB		TTC/TTG	505
rpoB		TTC/TTA	505
pncA	357	TGG/TGC	119
rpoB		ATC/TTC	572
katG	85-1059	del	29
katG	823	ACC/TCC	275
katG	2180	GCC/GAC	727
katG	920	GGA/GAA	307
katG	745	CGC/TGC	249
katG	982	TGG/GGG	328
katG	1054	CAA/TAA	352
katG	984	TGG/TCG	328
pncA	158	GAC/GCC	53
pncA	428	ins GG	143
rpoB		CAC/TTC	526
rpoB		GGG/GCG	523
rpoB		CAA/TAA	513
rpoB		CAA/AAT	513
rpoB		ACC/CAC	508
rpoB		GGC/GAT	507
rrs	904	C/A	904
rpoB		del GAC	517
rpoB		ACC/AGC	508
katG	1464	G/C	488
katG	1360	GAG/CGA	454
katG	1069	GAC/CAC	357
katG	1069	GAC/AAC	357
katG	984	TGG/TGT	328
katG	961	TGG/TAG	321

Supplementary Table 4.1 (continued)

katG	962	TGG/TTG	321
katG	962-963	TGG/TCC	321
katG	963	TGG/TGT	321
katG	946	GGC/AGC	316
katG	941	ACC/AAC	314
katG	941-942	ACC/ACG	314
katG	931	GAC/TAC	311
katG	931-932	GAC/TTC	311
katG	926	GGT/GTT	309
katG	926	GGT/GCT	309
katG	926-927	GGT/GTC	309
katG	919	GGA/CGA	307
katG	920	GGA/GCA	307
katG	913	GGC/GCC	305
inhA	581	ATC/ACC	194
katG	943	AGC/CTA	315
katG	103	AAC/GAC	35
pncA	263	TCG/TAG	88
pncA	38	TTC/TCC	13
pncA	215	ins 29bp	72
pncA	1	del 8bp	1
ethA, aka, etaA	167	GAC/GCC	55
ethA, aka, etaA	1322-1323	del GC	441
ethA, aka, etaA	127	GGC/AGC	43
ethA, aka, etaA	1174	ACG/GCG	392
ethA, aka, etaA	1154	GGC/GAC	385
ethA, aka, etaA	110	del A	37
ethA, aka, etaA	768	del G	256
ethA, aka, etaA	736	CAG/TAG	246
ethA, aka, etaA	703	del T	234
rpsL	262	AAG/CAG	88
embB	880	GGC/AGC	294
embB	878-879	ATT/AAC	293
embB	860-861	TTT/TGC	287
embB	859-861	TTT/GTG	287
embB	844	GTG/CTG	282
embB	841	GCG/CCG	281
embB	841-843	TTT/CTG	281

Supplementary Table 4.1 (continued)

embB	814	AGC/TGC	272
embB	815-816	AGC/ATT	272
embB	812	GCG/GTG	271
embB	680-681	GTG/GGC	227
embB	673-674	GCG/AGC	225
embB	662-663	GCG/GGC	221
katG	1300	CAG/TAG	434
katG	372	ins G	124
katG	1559	ins 64bp	521
pncA	446	del 8bp	149
pncA	206	CCA/CGA	69
rrs	426	G/C	426
katG	836	GGC/GAC	279
katG	270	TGG/TGA	90
katG	1513	TGG/CGG	505
katG	1271	GCG/GTG	424
katG	413	AAC/ACC	138
ndh	37	CGT/TGT	13
katG		del 8bp	10
rpoB		CTG/CCG	533
inhA	62	ATC/ACC	21
embB	1889	ACC/ATC	630
katG	1759	CTG/ATG	587
pncA	137	GCA/GTA	46
pncA	137	GCA/GAA	46
pncA	415	GTG/ATG	139
pncA	413	TGT/TCT	138
pncA	341	del C	114
pncA	301	ins G	100
katG	311	CGG/CTG	104
pncA	187	GAC/CAC	63
pncA	162	del G	54
pncA	422	CAG/CCG	141
pncA	413	TGT/TAT	138
pncA	288	del G	96
iniA			501
iniA			308
rmlD			257

Supplementary Table 4.1 (continued)

rmlA2, hddC, manB			152
iniB	-89	C/T	-89
rmlD	-71	G/T	-71
rpoB		TCG/TTG	574
pncA	297	TAC/TAG	99
pncA	233	GGC/GAC	78
pncA	212	CAT/CGT	71
pncA	185	CCG/CGG	62
pncA	416	GTG/GGG	139
pncA	403	ACC/CCC	135
pncA	395	GGT/GAT	132
pncA	388	ins GAGGTCGAT	129
pncA	136	ins G	46
pncA	152	CAC/CCC	51
pncA	152	CAC/CGC	51
pncA	128	CAC/CCC	43
pncA	340	ACG/CCG	114
pncA	287	ins T	96
pncA	188	GAC/GGC	63
pncA	50	GGC/GAC	17
pncA	297	TAC/TAA	99
pncA	20	GTC/GGC	7
pncA	199	TCG/CCG	67
pncA	123	TAC/TAA	41
pncA	101	TAC/TCC	34
pncA	515	CTG/CCG	172
pncA	424	ACG/GCG	142
pncA	389	GTG/GGG	130
pncA	312	AGC/AGG	104
pncA	29	CAG/CCG	10
pncA	395-411 del	del 16bp	132
pncA	437	GCG/GTG	146
ahpC	7	CTG/AAG	3
rpoB		TCG/TAT	531
rpoB		CAC/CAG	526
pncA	203	TGG/TCA	68
pncA	52	ins GG	17

Supplementary Table 4.1 (continued)

pncA	29	CAG/CGA	10
pncA	289	GGT/AGT	97
pncA	213	CAT/CAG	71
pncA	420	ins GG	140
pncA	420	ins G	140
pncA	288	AAG/AAT	96
pncA	287	AAG/ACG	96
pncA	254	CTG/CCG	85
pncA	24	GAC/GAG	8
pncA	214	TGC/CGC	72
pncA	162	ins T	54
pncA	14	ATC/AGC	5
pncA	100	ins T	34
pncA	98-216	del 118bp	33
pncA	77	GCG/GGG	26
pncA	518	ins 5bp	173
pncA	475	del C	158
pncA	446	ins 8bp	149
pncA	443	del G	148
pncA	425	ACG/AAG	142
pncA	425	ACG/ATG	142
pncA	40	TGC/CGC	14
pncA	416	del TG	139
pncA	410	CAT/CCT	137
pncA	394	GGT/AGT	132
pncA	391	del G	130
pncA	35	GAC/GCC	12
pncA	34	GAC/AAC	12
pncA	312	del C	104
pncA	28	del C	10
pncA	1	del 11bp	1
pncA	-11	A/C	-11
katG	4	CCC/TCC	2
katG	1649	GCC/GAC	550
katG	50	AGC/AAC	17
katG	727	GCC/TCC	243
katG	56	GGC/GAC	19
katG	926	GGT/GAT	309

Supplementary Table 4.1 (continued)

katG	790	GCG/ACG	264
katG	196	GCG/CCG	66
katG	193	GCC/ACC	65
katG	1010	TAC/TTC	337
katG	773	AAC/AGC	258
mabA, fabG1	-8	T/G	-8
inhA	280	TCG/GCG	94
mabA, fabG1	-24	G/T	-24
katG	514	GCG/ACG	172
katG	412	AAC/CAC	138
katG	412	AAC/GAC	138
katG	413	AAC/AGC	138
katG	419	AGC/AAC	140
katG	1471	GGC/TGC	491
katG	1574	CAG/CCG	525
katG	350	GAC/GCC	117
katG	1760	CTG/CCG	587
katG	583	G/A	195
katG	2	GTG/GCG	1
katG	944	AGC/ACC	315
katG	2098	TCC/CCC	700
katG	1849	ins AC	617
katG	1048	GCT/TCT	350
katG	898	TGG/GGG	300
katG	785	ACA/AGA	262
katG	944-945	AGC/ACA	315
katG	464	TAC/TCC	155
katG	1003	ATC/GTC	335
katG	945	AGC/AGA	315
rpoB		GTC/TTC	146
rpoB		GTC/TTC	146
katG	670	CAG/GAG	224
inhA	233	GTG/GCG	78
inhA	140	ATT/ACT	47
katG	157-168	del CCGTCGCTGACC	52
inhA	47	ATC/ACC	16
rrs	865	C/G	865

Supplementary Table 4.1 (continued)

katG	820	ACC/GCC	275
rpoB		del AAC	519
embA	2498	GAC/GGC	833
embA	1385	GCG/GTG	462
embA	1049	GGC/GAC	350
embA	961	GGC/AGC	321
embA	-16	C/T	-16
embB	2876	GAC/GCC	959
embB	983	GAT/GGT	328
embB	889	TCG/GCG	297
embB	3070	GAC/AAC	1024
embB	2999	ATG/AGG	1000
pncA	415	GTG/CTG	139
pncA	309	TAC/TAA	103
pncA	-7	T/C	-7
pncA	7	GCG/CCG	3
pncA	419	del 68bp	140
pncA	341	ins 1355bp	114
pncA	353	AAT/ACT	118
pncA	403	ins CC	134
pncA	368	ins AG	123
pncA	436	del G	145
pncA	260	ACG/ATG	87
pncA	192	ins A	64
pncA	485	GGT/GAT	162
mabA, fabG1	-17	G/T	-17
ethA, aka, etaA	338	ins A	113
pncA	224	ATG/ACG	75
mabA, fabG1	-17	G/T	-17
Rv3124	-16	A/G	-16
katG	31	ACC/GCC	11
katG	98	ins A	33
katG	109	del G	37
katG	185	ins C	62
katG	181	GCG/ACG	61
katG	30	del C	10
rpoB		TTC/TTG	514
pncA	206	CCA/CGA	69

Supplementary Table 4.1 (continued)

embB	1381-1383	CCG/AGC	461
embB	1193	TTT/TAT	398
embB	1123-1124	CCG/GCG	375
embB	1106	GTG/GCG	369
embB	931-932	GAC/CGT	311
embB	928-930	GCG/CGT	310
embB	926-927	GTG/GGC	309
embR	718-720	GGC/CGT	240
embB	931	GAC/CAC	311
embC	761	GCC/GGC	254
embC	752	CTG/CGT	251
embB	1105	GTG/CTG	369
embB	845	GTG/GGG	282
gid, gidB		TGG/TGC	45
gid, gidB		TGG/CGG	148
pncA	464	GTG/GGG	155
pncA	308	TAC/TGC	103
pncA	139	ACC/CCC	47
embC	2941	GTG/CTG or GTG/TTG	981
rrs	904	C/G	904
embB	352	AAC/CAT	318
embB	888	AAT/AAA	296
embR	669	GAT/GAA	223
embB	1203	ACG/ACA	401
embB	1078	GTG/ATG	360
embB	991-993	GGC/CGT	331
pncA	59	del C	20
katG	195	ins CCCC	65
katG	1888	GGC/AGC	629
katG	1010	TAC/TGC	337
katG	842	GCC/GTC	281
katG	506	GGC/GCC	169
katG	?	GAC/AAC	73
katG	34	ACC/CCC	12
katG	932	GAC/GGC	311
katG	271	TGG/CGG	91
embB	897	GAC/GAA	299

Supplementary Table 4.1 (continued)

katG	1225-1226	GCC/CGC	409
gid, gidB	223	C/G	75
gid, gidB	118	G/T	40
gid, gidB	47	CCT/CGT	16
gid, gidB	115	del C	39
gid, gidB	212	G/T	71
rrs	887	G/A	887
gid, gidB	208	A/C	70
gid, gidB	164	T/G	55
gid, gidB	563	T/G	188
gid, gidB	276	GAA/GAC	92
gid, gidB	210	C/A	70
gid, gidB	140	G/A	47
gid, gidB	413	C/T	138
gid, gidB	102	del G	34
rpsL	257	CGG/CCG	86
rrs	907	A/G	907
rrs	907-908	ins T	907
rrs	906	A/C	906
rrs	513	A/T	513
embA	-43	del CG	-43
embR	1136	CAG/CGG	379
embR	-273	ins A	-273
rpoB		GCG/GTG	381
rpoB		GCG/GTG	381
ethA, aka, etaA	1013	ATC/AGC	338
embB	1217	GGC/GCC	406
embB	918	ATG/ATA	306
embB	918	ATG/ATT	306
embB	918	ATG/ATC	306
embA	-11	C/A	-11
embR	599-600	GAA/GCG	200
embR	556-558	CGT/GGC	186
embB	716	CTC/CCC	239
embB	1354	GTG/CTG	452
embB	1200	AAC/AAA	400
embB	1150-1152	TAT/AAC	384
embB	1130-1131	GTG/GGC	377

Supplementary Table 4.1 (continued)

embB	953	AAC/AGC	318
embB	943-945	TAT/CTG	315
embB	917-918	ATG/ACG	306
embB	916	ATG/TTG	306
embB	1412-1413	CGT/CCG	471
embB	1406-1407	CGT/CCG	469
embB	1393-1394	ATT/GAT	465
embB	1337-1338	CCG/CAT	446
embB	1103-1104	GAA/GCG	378
embC	925-927	TAT/AAC	309
embC	859-861	GTG/TTT	287
embB	1291-1293	GCG/ACC	431
embB	1289	CCG/CTG	430
embB	1135-1137	GCG/ACC	379
embR	1106-1107	ATT/ACC	369
embB	1069-1071	GCG/AGC	357
embR	1066-1068	CGT/GGC	356
embR	1053	CAT/CAG	351
embR	1049-1050	GTG/GGC	350
embR	977	AGC/AAC	326
embR	958	CTG/GTG	320
embR	631	CTG/ATG	211
embR	541-543	AGC/CGT	188
katG	2053	GGC/CGC	685
gyrA	265	GAC/AAC	89
gyrA	263	GGC/GCC	88
embB	1490	CAG/CGG	497
embB	1490	CAG/AAG	497
embB	1216	GGC/TGC	406
pncA	245	CAT/CGT	82
pncA	503	ACC/AAC	168
iniB			47
embB	1216	GGC/AGC	406
embB	1000	TAC/CAC	334
embB	982	GAT/TAT	328
embA	601	GCG/ACG	201
embA	-12	C/T	-12
rpoB		TCG/GCG	553

Supplementary Table 4.1 (continued)

rpoB		GAG/GAT	541
rpoB		TCG/CAG	522
embB	1190	GGC/GCG	406
embB	1190	CCG/CAG	397
embB	1138-1140	AGC/CGT	380
embB	1103-1104	GAA/GCG	368
embR	157	GAT/CAT	53
embB	1376-1377	GGC/GCG	459
embR	94-96	ACC/CCG	32
embB	954	AAC/AAA	318
katG	1765	CCC/ACC	589
ndh	53	GTG/GCG	18
embC	976-978	TGG/CGT	326
embC	889-891	ATT/TTG	297
embB	1343-1344	GGC/GTG	448
embB	1217	GGC/GAT	406
embB	1189-1190	CCG/ACC	397
embB	1139	AGC/AAC	380
embB	1102-1104	GAA/CAG	368
embB	1073-1074	GGC/GTG	358
katG	2084	GAC/GCC	695
rpoB		AAG/AAT	527
embA	2737	CCG/TCG	913
rpoB		CGC/CCC	528
mabA, fabG1	-15	C/T	-15
pncA	290	GGT/GAT	97
pncA	174	TTC/TTG	58
embB	2234	GGC/GAC	745
Rv3125c			54
embB	1217	GGC/GAC	406
embA	10	GAC/AAC	4
iniC			351
embB	988	TTC/GTC	330
embB	916	ATG/CTG	306
embB	916	ATG/GTG	306
embR	20-21	GTG/GGC	7
embC	889-891	ATT/CTG	297
embB	991	GAT/TAT	331

Supplementary Table 4.1 (continued)

embB	965	TAT/TGT	322
embB	941	GGC/GCG	314
embB	718	GAC/CAC	240
embB	1198-1200	AAC/CCG	400
embB	1185	TGG/TGC	395
embB	1121-1122	GGC/GTG	374
embC	809-810	ACC/ATT	270
embB	1379-1380	CGT/CTG	460
embR	1187-1188	ATT/ACC	396
embB	1104	GAA/GAT	368
embB	1040	AGT/ATT	347
embC	890-891	ATT/ACC	297
embR	688-690	CGT/TGG	230
embR	559-560	GCG/GTG	187
embR	515-516	CTG/CGT	172
katG	324	CAC/CAG	108
rpoB		AAC/AAG	519
embB		TAT/GAT	319
embB		TAT/CAT	333
rpoB		GAC/CAC	516
pncA	391	ins GG	130
pncA	306	ins C	102
pncA	545	TTG/TCG	182
pncA	181	ACA/CCA	61
embC	1180	AAC/GAC	394
Rv3126			276
katG	189	GAC/GAA	63
ethA, aka, etaA	1387	CGT/AGT	463
ethA, aka, etaA	1290	del C	430
ethA, aka, etaA	1238	GGT/GAT	413
ethA, aka, etaA	667	GAG/AAG	223
katG	2146	GCG/CCG	716
katG	1022	TGG/TCG	341
iniC			248
embB	383	GTC/GGG	128
pncA	121	TAC/CAC	41
pncA	71	del G	24
rpoB		CGC/CAC	528

Supplementary Table 4.1 (continued)

embC	908-909	GTG/GGC	303
embC	862-864	GGC/TGG	288
embC	730-732	GCG/ACC	244
embC	919-921	GCG/ACC	307
embC	898-900	ATG/CGT	300
embC	853	CAT/TAT	285
embC	814-815	GGC/GAC	272
embB	1378-1380	CGT/TGC	460
embB	1309-1311	ACC/GCG	437
embB	1307-1308	GTG/GGC	436
embB	1304-1305	GTG/GGC	435
embB	1301-1302	GCG/GGC	434
embB	1213-1214	GAA/CCG	405
embB	1195-1196	TTT/CAT	399
embB	1192-1193	TTT/CAT	398
embB	1183-1185	TGG/CGT	395
embB	1132	GAA/AAA	378
embB	1129	GTG/ATG	377
embB	1112-1113	CCG/CGT	371
embB	1109-1110	CTG/CGT	370
embB	1096	AGC/CCG	366
embB	1066-1068	GCG/AGC	356
embB	1067	GCG/GTG	356
embC	973	GGC/AGC	325
embC	923-924	GGC/GAT	308
embR	907-909	CTG/GAA	303
embC	904-906	CGT/GGC	302
embC	886-888	TAT/CAT	296
embC	886-888	TAT/AGC	296
embR	874	CTG/ATG	292
embB	769-771	CGT/TGG	257
embR	527	GCG/GTG	176
gid, gidB	199	G/C	67
rpoB		ACC/ATC	525
rpoB		GGG/TGG	523
pncA	11	TTG/TCG	4
pncA	106	del 5.3kb	35
pncA	458	ACC/AAC	153

Supplementary Table 4.1 (continued)

pncA	398	ATT/AAT	133
pncA	29	CAG/CCA	10
embC	739-741	GCC/CCG	247
embB	1216	GGC/CCG	406
embB	1195-1196	TTT/GAT	399
embB	1163-1164	GCG/GGC	388
embB	1136-1137	GCG/GAT	379
embB	1070	GCG/GTG	357
embR	329-330	TGC/TAT	110
katG	1543	CGC/TGC	515
katG	906	AGC/AGG	302
rpoB		TCG/TGG	531
rpoB		CAC/CCC	526
rpoB		TCG/TTG	522
gid, gidB	351	del G	117
inhA	61	ATC/GTC	21
pncA	301	ins A	100
pncA	501	ins CG	167
embB	1100-1101	CGT/CCG	367
embB	1075-1077	CTG/ATT	359
embB	1066-1068	GCG/TTT	356
embC	979-981	TAT/AAC	327
embB	950	AGC/ACC	317
embB	935	CAT/CGT	312
embC	929-930	ATG/AAA	310
embC	863-864	GGC/GTG	288
inhA	283-284	ATT/CCT	95
tlyA	272	GCA/GAA	91
tlyA	64	CAG/TAG	22
tlyA	550	C/T	184
tlyA	477	del G	159
tlyA	7	CGA/TGA	3
tlyA	758	del C	253
tlyA	586	del G	195
tlyA	548	C/T	183
tlyA	397	ins C	132
tlyA	353	CTG/CCG	118
tlyA	218	ins C	73

Supplementary Table 4.1 (continued)

tlyA	23	del A	8
gyrA	280	GAC/CAC	94
gyrA	281	GAC/GCC	94
gyrA	271	TCG/CCG	91
gyrA	269	GCG/GTG	90
gyrA	280	GAC/AAC	94
gyrA	280	GAC/TAC	94
gyrA	281	GAC/GGC	94
gyrA	269	GCG/GGG	90
gyrB	1528	AAC/GAC	510
tlyA	449	CTG/CCG	150
tlyA	708	AAT/AAG	236
rrs	1484	G/T	1484
rrs	1402	C/T	1402
rrs	1402	C/A	1402
rrs	1408	A/G	1408
tlyA	40	CGG/TGG	14
rrs	1401	A/G	1401
tlyA	67-69	del GCC	23
tlyA		ins C	218
rrs	491	C/T	491
rrs	512	ins C	512
rpsL	263	AAG/AGG	88
rrs	513	A/C	513
rpsL	128	AAG/AGG	43
rpoB		TCG/TTG	531
rpoB		CAC/GAC	526
rpoB		CAC/TAC	526
rpoB		CAC/CGC	526
rpoB		CAC/CTC	526
rpoB		GAC/GTC	516
rpoB		GAC/TAC	516
rpoB		CAA/CTA	513
rpoB		CTG/CCG	511
rpoB		CAC/AAC	526
gyrA	220	GCC/TCC	74
gyrA	281	GAC/GTC	94
gyrA	325	CTG/GTG	109

Supplementary Table 4.1 (continued)

gyrB	1414	GAC/CAC	472
gyrB	1647	CAA/CAT	549
gyrB	1543	GCG/ACG	515
gyrB	1544	GCG/GTG	515
gyrB	1530	AAC/AAA	510
ahpC	-34	T/A	-34
ahpC	-34	del T	-34
ahpC	-20	C/T	-20
katG	2096	GGG/GAG	699
katG	149	del A	50
katG	1469	CGC/TGC	490
katG	1450	CGT/AGT	484
katG	723	CCC/CCG	241
katG	701	GGG/GAG	234
katG	571	TGG/CGG	191
katG	321	TGG/TGA	107
mabA, fabG1	-8	T/C	-8
mabA, fabG1	-24	G/T	-24
katG	-1	ins C	-1
kasA	805	GGT/AGT	269
ahpC	-39	C/T	-39
katG	944	AGC/AAC	315
katG	944	AGC/ATC	315
pncA	226	ACT/CCT	76
pncA	202	TGG/CGG	68
pncA	202	TGG/GGG	68
pncA	203	TGG/TTG	68
pncA	196	TCG/CCG	66
pncA	160	CCG/ACG	54
pncA	139	ACC/GCC	47
pncA	123	TAC/TAG	41
pncA	70	del G	24
pncA	62	GTA/GGA	21
pncA	56	CTG/CCG	19
pncA	512	GCG/GAG	171
pncA	481	GCG/CCG	161
pncA	478	ACA/CCA	160
pncA	418	CGC/AGC	140

Supplementary Table 4.1 (continued)

pncA	416	GTG/GCG	139
pncA	410	CAT/CGT	137
pncA	391	ins GG	130
pncA	388	ins AGGTCGATG	129
pncA	384	GTC/GGC	128
pncA	362	CGG/CCG	121
pncA	347	CTG/CGG	116
pncA	309	TAC/TAG	103
pncA	305	GCG/GTG	102
pncA	-11	A/G	-11
pncA	401	GCC/GTC	134
pncA	172	del T	58
pncA	169	CAC/TAC	57
pncA	104	CTG/CGG	35
pncA	554	AGC/ACC	185
pncA	476	CTG/CCG	159
pncA	40	TGC/CGC	14
gyrA	305	CCC/CAC	102

Supplementary Table 4.2 Intergenic regions and genes associated with a high degree of polymorphism in drug-resistant isolates (Ref. 120, 135)

Gene	Start	End	Strand
922	1275551	1276299	-
1058	1471578	1471845	-
1292	1836237	1836386	-
1899	2715333	2715471	-
2867	4120956	4121197	-
MTB000019	1471846	1473382	
Rv0005	5240	7267	
Rv0006	7302	9818	positive
Rv0006	7302	9818	
Rv0010c Rv0011c	13558	13714	
Rv0026	29722	31068	
Rv0039c	42004	42351	
Rv0050	53663	55699	positive
Rv0064	68620	71559	positive
Rv0109	131382	132872	positive
Rv0109	131382	132872	positive
Rv0147	173238	174758	
Rv0165c	194144	194815	
Rv0179c	209703	210812	
Rv0218	260924	262252	positive
Rv0218	260924	262252	positive
Rv0244c	293798	295633	
Rv0265c	316511	317503	
Rv0278c	333437	336310	negative
Rv0279c	336560	339073	negative
Rv0280	339364	340974	positive
Rv0280	339364	340974	positive
Rv0323c	390580	391251	
Rv0388c	467459	468001	negative
Rv0401	479789	480160	
Rv0402c	480355	483231	
Rv0404	483977	485734	
Rv0466 Rv0467	557252	557527	
Rv0532	622793	624577	positive
Rv0565c	656010	657470	

Supplementary Table 4.2 (continued)

Rv0588	685928	686815	
Rv0600c	697904	698410	
Rv0605	701406	702014	
Rv0608	703244	703489	
Rv0646c	740234	741139	
Rv0658c	753693	754409	negative
Rv0658c	753693	754409	negative
Rv0667	759807	763325	positive
Rv0667	759807	763325	positive
Rv0667	759807	763325	
Rv0668	763370	767320	positive
Rv0668	763370	767320	
Rv0682	781560	781934	positive
Rv0682	781560	781934	positive
Rv0682	781560	781934	
Rv0744c Rv0745	834946	835154	
Rv0746	835701	838052	positive
Rv0747	838451	840856	positive
Rv0812	906423	907292	
Rv0840c Rv0841	937317	937593	
Rv0859	955077	956288	
Rv0878c Rv0879c	978203	978481	
Rv0893c	995318	996295	
Rv0908 Rv0909	1014124	1014681	
Rv0920c Rv0921	1026816	1027104	
Rv1042c Rv1043c	1165499	1165781	
Rv1058	1180684	1182315	
Rv1070c	1194270	1195043	
Rv1080c Rv1081c	1205798	1205984	
Rv1090	1215599	1216054	
Rv1096	1224385	1225260	
Rv1112	1238255	1239328	
Rv1129c	1253074	1254534	
Rv1144	1271156	1271908	
Rv1180	1313725	1315191	positive
Rv1192	1334927	1335754	
Rv1194c Rv1195	1338513	1339003	

Supplementary Table 4.2 (continued)

Rv1218c	1361798	1362733	
Rv1286	1438907	1440751	
Rv1302 Rv1303	1459509	1459766	
Rv1319c	1480894	1482501	negative
Rv1347c Rv1348	1512605	1513047	
Rv1380	1553232	1554191	
Rv1387	1561769	1563388	
Rv1393c	1568109	1569587	
Rv1446c	1624454	1625365	negative
Rv1465	1652768	1653256	
Rv1482c Rv1483	1673299	1673440	
Rv1484	1674202	1675011	
Rv1520	1711028	1712068	
Rv1662	1881704	1886512	
Rv1663	1886512	1888020	
Rv1699	1923829	1925589	
Rv1704c	1929786	1931456	
Rv1736c	1962228	1964186	
Rv1741	1967917	1968165	
Rv1746	1972138	1973568	
Rv1816 Rv1817	2058960	2059595	
Rv1885c	2134273	2134872	
Rv1900c Rv1901	2147633	2147662	
Rv1908c	2153889	2156111	negative
Rv1908c	2153889	2156111	
Rv1909c	2156149	2156592	
Rv1969	2212855	2214126	
Rv2024c	2268693	2270240	negative
Rv2043c	2288681	2289241	negative
Rv2043c	2288681	2289241	
Rv2048c	2294531	2306986	negative
Rv2068c Rv2069	2326809	2326944	
Rv2077c	2333323	2334294	
Rv2080	2337306	2337869	
Rv2082	2338709	2340874	positive
Rv2155c	2414934	2416394	negative
Rv2155c	2414934	2416394	negative

Supplementary Table 4.2 (continued)

Rv2208 Rv2209	2473242	2473400	
Rv2215	2481965	2483626	
Rv2274c	2546488	2546805	
Rv2287	2559703	2561331	
Rv2314c	2585917	2587290	
Rv2340c Rv2341	2618908	2619597	
Rv2416c Rv2417c	2715332	2715472	
Rv2427c Rv2428	2725477	2726193	
Rv2436	2733230	2734144	positive
Rv2447c	2746135	2747598	
Rv2530c	2854267	2854686	
Rv2650c	2973795	2975234	negative
Rv2729c	3041570	3042475	
Rv2733c Rv2734	3046524	3046821	
Rv2741	3053914	3055491	positive
Rv2752c	3064515	3066191	
Rv2754c Rv2755c	3067945	3068189	
Rv2764c	3073680	3074471	
Rv2764c Rv2765	3074471	3074636	
Rv2771c	3080581	3081033	
Rv2807	3113658	3114812	
Rv2853	3162268	3164115	positive
Rv2853	3162268	3164115	positive
Rv2896c	3205265	3206434	negative
Rv2896c	3205265	3206434	negative
Rv2931	3245445	3251075	positive
Rv2947c	3296350	3297840	
Rv3021c	3379376	3380452	negative
Rv3067	3431428	3431838	
Rv3071	3434464	3435573	
Rv3087	3452925	3454343	
Rv3090	3458211	3459098	
Rv3093c	3461760	3462764	negative
Rv3121	3486509	3487711	
Rv3185 Rv3186	3552542	3552764	
Rv3210c Rv3211	3587539	3587798	
Rv3245c	3624910	3626613	negative

Supplementary Table 4.2 (continued)

Rv3260c Rv3261	3640141	3640543	
Rv3343c	3729364	3736935	negative
Rv3345c	3738158	3742774	negative
Rv3347c	3743711	3753184	negative
Rv3446c	3863317	3864531	negative
Rv3462c Rv3463	3880653	3880907	
Rv3478	3894426	3895607	positive
Rv3507	3926569	3930714	positive
Rv3564	4005247	4006203	
Rv3651 Rv3652	4092878	4093632	
Rv3711c	4155740	4156729	negative
Rv3756c	4201894	4202613	
Rv3765c Rv3766	4211784	4212293	
Rv3793 Rv3794	4243147	4243233	
Rv3794	4243233	4246517	
Rv3795	4246514	4249810	positive
Rv3795	4246514	4249810	positive
Rv3795	4246514	4249810	
Rv3796 Rv3797	4251005	4251085	
Rv3806c	4268925	4269833	
Rv3825c	4293225	4299605	
Rv3854c	4326004	4327473	negative
Rv3854c	4326004	4327473	
Rv3862c	4338171	4338521	
Rv3862c Rv3863	4338521	4338849	
Rv3877	4355007	4356542	
Rv3881c	4360543	4361925	
Rv3889c	4372800	4373630	
Rv3894c	4376262	4380452	
Rv3919c	4407528	4408202	negative
Rvnr01	1471846	1473382	positive
Rvnr01	1471846	1473382	positive

Supplementary Table 4.3 Polymorphisms listed as enriched in drug-resistant strains (Ref. 119, 120, 135)

RvNumber	Position	Database
Rv0001	698	Ref. 119
Rv0006	7362	Ref. 135
Rv0006	7363	Ref. 135
Rv0006	7364	Ref. 135
Rv0006	7587	Ref. 135
Rv0006	8430	Ref. 135
Rv0006	9306	Ref. 135
Rv0015c	18147	Ref. 119
Rv0033	36853	Ref. 119
Rv0050	53670	Ref. 135
Rv0050	53671	Ref. 135
Rv0050	53672	Ref. 135
Rv0050	53785	Ref. 135
Rv0050	53786	Ref. 135
Rv0050	53787	Ref. 135
Rv0050	53964	Ref. 135
Rv0050	53965	Ref. 135
Rv0050	53966	Ref. 135
Rv0050	54292	Ref. 135
Rv0050	54293	Ref. 135
Rv0050	54294	Ref. 135
Rv0050	54308	Ref. 135
Rv0050	54309	Ref. 135
Rv0050	54310	Ref. 135
Rv0050	54608	Ref. 135
Rv0050	54609	Ref. 135
Rv0050	54610	Ref. 135
Rv0050	54757	Ref. 135
Rv0050	54758	Ref. 135
Rv0050	54759	Ref. 135
Rv0050	55549	Ref. 119
Rv0064	69338	Ref. 135
Rv0064	69339	Ref. 135
Rv0064	69340	Ref. 135
Rv0064	69871	Ref. 135

Supplementary Table 4.3 (continued)

Rv0064	69872	Ref. 135
Rv0064	69873	Ref. 135
Rv0064	69991	Ref. 135
Rv0064	70267	Ref. 135
Rv0064	70268	Ref. 135
Rv0064	70269	Ref. 135
Rv0064	70300	Ref. 135
Rv0064	70301	Ref. 135
Rv0064	70302	Ref. 135
Rv0064	70816	Ref. 135
Rv0064	70817	Ref. 135
Rv0064	70818	Ref. 135
Rv0064	70924	Ref. 135
Rv0064	70925	Ref. 135
Rv0064	70926	Ref. 135
Rv0064	71338	Ref. 135
Rv0101	113793	Ref. 119
Rv0104	122334	Ref. 119
Rv0107c	128620	Ref. 119
Rv0120c	147250	Ref. 119
Rv0136	164315	Ref. 119
Rv0191	222486	Ref. 119
Rv0206c	244552	Ref. 119
Rv0236c	284783	Ref. 119
Rv0236c	285869	Ref. 135
Rv0236c	285870	Ref. 135
Rv0236c	285871	Ref. 135
Rv0252	304904	Ref. 119
Rv0277c	332737	Ref. 119
Rv0277c	332918	Ref. 119
Rv0277c	333046	Ref. 119
Rv0277c	333088	Ref. 119
Rv0277A	333211	Ref. 119
Rv0278c	334981	Ref. 135
Rv0278c	334982	Ref. 135
Rv0278c	334983	Ref. 135
Rv0278c	335297	Ref. 135
Rv0278c	335298	Ref. 135

Supplementary Table 4.3 (continued)

Rv0278c	335299	Ref. 135
Rv0278c	336189	Ref. 135
Rv0278c	336190	Ref. 135
Rv0278c	336191	Ref. 135
Rv0279c	337957	Ref. 135
Rv0279c	338098	Ref. 135
Rv0279c	338451	Ref. 135
Rv0279c	338452	Ref. 135
Rv0279c	338453	Ref. 135
Rv0279c	338717	Ref. 135
Rv0279c	338718	Ref. 135
Rv0279c	338719	Ref. 135
Rv0279c	338842	Ref. 135
Rv0279c	338901	Ref. 135
Rv0279c	338902	Ref. 135
Rv0279c	338903	Ref. 135
Rv0280	340132	Ref. 135
Rv0280	340133	Ref. 135
Rv0280	340134	Ref. 135
Rv0323c	390986	Ref. 119
Rv0371c	448551	Ref. 119
Rv0388c	467544	Ref. 135
Rv0388c	467545	Ref. 135
Rv0388c	467546	Ref. 135
Rv0388c	467562	Ref. 135
Rv0388c	467563	Ref. 135
Rv0388c	467564	Ref. 135
Rv0388c	467583	Ref. 135
Rv0388c	467584	Ref. 135
Rv0388c	467636	Ref. 135
Rv0388c	467637	Ref. 135
Rv0388c	467638	Ref. 135
Rv0388c	467983	Ref. 135
Rv0388c	467984	Ref. 135
Rv0388c	467985	Ref. 135
Rv0392c	471671	Ref. 119
Rv0404	485307	Ref. 119
Rv0411c	497388	Ref. 119

Supplementary Table 4.3 (continued)

Rv0420c	506572	Ref. 119
Rv0520	612556	Ref. 119
Rv0527	617698	Ref. 119
Rv0532	623021	Ref. 135
Rv0532	623022	Ref. 135
Rv0532	623023	Ref. 135
Rv0532	623474	Ref. 135
Rv0532	623510	Ref. 135
Rv0538	631400	Ref. 119
Rv0538	631403	Ref. 119
Rv0578c	672489	Ref. 135
Rv0578c	672490	Ref. 135
Rv0578c	672491	Ref. 135
Rv0591	690172	Ref. 119
Rv0667	760317	Ref. 135
Rv0667	761032	Ref. 135
Rv0667	761033	Ref. 135
Rv0667	761034	Ref. 135
Rv0667	761074	Ref. 135
Rv0667	761075	Ref. 135
Rv0667	761076	Ref. 135
Rv0667	761112	Ref. 135
Rv0667	761114	Ref. 119
Rv0667	761115	Ref. 119
Rv0667	761142	Ref. 135
Rv0667	761144	Ref. 119
Rv0667	761145	Ref. 119
Rv0667	761157	Ref. 135
Rv0667	761163	Ref. 135
Rv0667	761166	Ref. 119
Rv0667	761253	Ref. 135
Rv0667	761254	Ref. 135
Rv0667	761255	Ref. 135
Rv0667	761282	Ref. 119
Rv0667	761490	Ref. 135
Rv0667	761491	Ref. 135
Rv0667	761492	Ref. 135
Rv0667	762000	Ref. 135

Supplementary Table 4.3 (continued)

Rv0667	762315	Ref. 119
Rv0668	764582	Ref. 135
Rv0668	764665	Ref. 119
Rv0668	764719	Ref. 119
Rv0668	764827	Ref. 119
Rv0668	764845	Ref. 119
Rv0668	764846	Ref. 119
Rv0668	764936	Ref. 119
Rv0668	764953	Ref. 119
Rv0668	765464	Ref. 135
Rv0668	765465	Ref. 135
Rv0668	765466	Ref. 119
Rv0668	765467	Ref. 119
Rv0668	765468	Ref. 119
Rv0668	765619	Ref. 135
Rv0668	765620	Ref. 135
Rv0668	765621	Ref. 135
Rv0668	766027	Ref. 119
Rv0668	766467	Ref. 135
Rv0668	766468	Ref. 135
Rv0668	766469	Ref. 135
Rv0668	766490	Ref. 135
Rv0668	766492	Ref. 119
Rv0668	766493	Ref. 119
Rv0668	766646	Ref. 135
Rv0668	766647	Ref. 135
Rv0682	781689	Ref. 135
Rv0682	781692	Ref. 119
Rv0682	781824	Ref. 135
Rv0682	781827	Ref. 119
Rv0724A	817532	Ref. 119
Rv0746	836053	Ref. 135
Rv0746	836054	Ref. 135
Rv0746	836055	Ref. 135
Rv0746	836274	Ref. 135
Rv0746	836291	Ref. 135
Rv0746	836292	Ref. 135
Rv0746	836293	Ref. 135

Supplementary Table 4.3 (continued)

Rv0746	837928	Ref. 135
Rv0746	837929	Ref. 135
Rv0746	837930	Ref. 135
Rv0747	839123	Ref. 135
Rv0747	839124	Ref. 135
Rv0747	839125	Ref. 135
Rv0747	839129	Ref. 135
Rv0747	839130	Ref. 135
Rv0747	839131	Ref. 135
Rv0747	839336	Ref. 135
Rv0747	839684	Ref. 135
Rv0747	839685	Ref. 135
Rv0747	839686	Ref. 135
Rv0747	840179	Ref. 135
Rv0747	840180	Ref. 135
Rv0747	840181	Ref. 135
Rv0749	841331	Ref. 119
Rv0749	841382	Ref. 119
Rv0749	841383	Ref. 119
Rv0749	841453	Ref. 119
Rv0749	841487	Ref. 119
Rv0749	841634	Ref. 119
Rv0749	841655	Ref. 119
Rv0750	842049	Ref. 119
Rv0750	842056	Ref. 119
Rv0750	842061	Ref. 119
Rv0750	842062	Ref. 119
Rv0750	842063	Ref. 119
Rv0750	842068	Ref. 119
Rv0750	842070	Ref. 119
Rv0750	842116	Ref. 119
Rv0808	902554	Ref. 119
Rv0814c	908191	Ref. 119
Rv0818	911522	Ref. 119
Rv0823c	916768	Ref. 119
Rv0829	921813	Ref. 135
Rv0845	941757	Ref. 119
Rv0874c	972799	Ref. 119

Supplementary Table 4.3 (continued)

Rv0875c	974009	Ref. 119
Rv0933	1041451	Ref. 119
Rv0938	1048110	Ref. 119
Rv0979c	1094713	Ref. 119
Rv0987	1103338	Ref. 119
Rv1041c	1164577	Ref. 119
Rv1129c	1254138	Ref. 119
Rv1194c	1338071	Ref. 119
Rv1197	1340665	Ref. 119
Rv1197	1340675	Ref. 119
Rv1198	1341072	Ref. 119
Rv1198	1341081	Ref. 119
Rv1198	1341168	Ref. 119
Rv1198	1341182	Ref. 119
Rv1198	1341262	Ref. 119
Rv1206	1349383	Ref. 119
Rv1237	1380580	Ref. 119
Rv1256c	1404177	Ref. 119
Rv1288	1441541	Ref. 119
Rv1313c	1468208	Ref. 135
Rvnr01	1472344	Ref. 135
Rvnr01	1472345	Ref. 135
Rvnr01	1472346	Ref. 135
Rvnr01	1472359	Ref. 120, Ref. 135
Rvnr01	1472360	Ref. 135
Rvnr01	1472361	Ref. 135
Rvnr01	1472362	Ref. 135
Rvnr01	1472363	Ref. 135
Rvnr01	1472364	Ref. 135
Rvnr01	1472367	Ref. 119
Rvnr01	1472370	Ref. 119
Rvnr01	1472751	Ref. 135
Rvnr01	1472752	Ref. 135
Rvnr01	1472753	Ref. 135
Rvnr01	1472759	Ref. 119
Rvnr01	1473246	Ref. 120, Ref. 135

Supplementary Table 4.3 (continued)

Rvnr01	1473247	Ref. 135
Rvnr01	1473248	Ref. 135
Rvnr01	1473254	Ref. 119
Rv1319c	1481131	Ref. 135
Rv1319c	1481132	Ref. 135
Rv1319c	1481133	Ref. 135
Rv1319c	1481183	Ref. 135
Rv1319c	1481184	Ref. 135
Rv1319c	1481335	Ref. 135
Rv1319c	1481336	Ref. 135
Rv1319c	1481337	Ref. 135
Rv1319c	1482183	Ref. 135
Rv1319c	1482184	Ref. 135
Rv1319c	1482185	Ref. 135
Rv1326c	1490812	Ref. 119
Rv1378c	1552555	Ref. 119
Rv1400c	1576489	Ref. 119
Rv1437	1614824	Ref. 119
Rv1446c	1624789	Ref. 135
Rv1446c	1625252	Ref. 135
Rv1446c	1625253	Ref. 135
Rv1446c	1625254	Ref. 135
Rv1483	1674056	Ref. 119
Rv1484	1674489	Ref. 119
Rv1485	1675181	Ref. 119
Rv1557	1761773	Ref. 119
Rv1566c	1774087	Ref. 119
Rv1591	1791608	Ref. 119
Rv1614	1813611	Ref. 119
Rv1634	1840310	Ref. 119
Rv1635c	1841444	Ref. 119
Rv1792	2030363	Ref. 119
Rv1792	2030529	Ref. 119
Rv1793	2030856	Ref. 119
Rv1793	2030950	Ref. 119
Rv1808	2050913	Ref. 135
Rv1808	2050914	Ref. 135
Rv1808	2050915	Ref. 135

Supplementary Table 4.3 (continued)

Rv1860	2108611	Ref. 119
Rv1872c	2122403	Ref. 119
Rv1872c	2123141	Ref. 119
Rv1873	2123177	Ref. 119
Rv1873	2123189	Ref. 119
Rv1873	2123190	Ref. 119
Rv1882c	2132768	Ref. 119
Rv1908c	2154722	Ref. 135
Rv1908c	2154723	Ref. 135
Rv1908c	2154724	Ref. 135
Rv1908c	2155176	Ref. 119
Rv1908c	2155500	Ref. 135
Rv1908c	2155501	Ref. 135
Rv1908c	2155502	Ref. 135
Rv1908c	2155541	Ref. 120
Rv1908c	2155607	Ref. 120
Rv1908c	2155648	Ref. 120
Rv1908c	2155670	Ref. 135
Rv1908c	2155671	Ref. 135
Rv1908c	2155672	Ref. 135
Rv1912c	2158913	Ref. 119
Rv1917c	2163375	Ref. 135
Rv1922	2174224	Ref. 119
Rv1944c	2195930	Ref. 119
Rv1944c	2195931	Ref. 119
Rv1968	2212447	Ref. 119
Rv2024c	2269778	Ref. 135
Rv2030c	2277281	Ref. 119
Rv2043c	2288729	Ref. 135
Rv2043c	2288730	Ref. 135
Rv2043c	2288736	Ref. 119
Rv2043c	2288748	Ref. 119
Rv2043c	2288846	Ref. 135
Rv2043c	2288847	Ref. 135
Rv2043c	2288848	Ref. 135, Ref.121
Rv2043c	2288852	Ref. 135
Rv2043c	2288853	Ref. 135

Supplementary Table 4.3 (continued)

Rv2043c	2288854	Ref. 135
Rv2043c	2288868	Ref. 119
Rv2043c	2288883	Ref. 135
Rv2043c	2288884	Ref. 135
Rv2043c	2288931	Ref. 135
Rv2043c	2288932	Ref. 135
Rv2043c	2288947	Ref. 119
Rv2043c	2288971	Ref. 119
Rv2043c	2289052	Ref. 135
Rv2043c	2289053	Ref. 135
Rv2043c	2289054	Ref. 135
Rv2043c	2289071	Ref. 135
Rv2043c	2289072	Ref. 135
Rv2043c	2289088	Ref. 135
Rv2043c	2289210	Ref. 135
Rv2043c	2289211	Ref. 135
Rv2043c	2289240	Ref. 119
Rv2048c	2295683	Ref. 135
Rv2048c	2295684	Ref. 135
Rv2048c	2295685	Ref. 135
Rv2048c	2296051	Ref. 119
Rv2048c	2297137	Ref. 135
Rv2048c	2297138	Ref. 135
Rv2048c	2297139	Ref. 135
Rv2048c	2297974	Ref. 135
Rv2048c	2297975	Ref. 135
Rv2048c	2297976	Ref. 135
Rv2048c	2299194	Ref. 135
Rv2048c	2299195	Ref. 135
Rv2048c	2299196	Ref. 135
Rv2048c	2300235	Ref. 135
Rv2048c	2300236	Ref. 135
Rv2048c	2300237	Ref. 135
Rv2048c	2300246	Ref. 119
Rv2048c	2300544	Ref. 135
Rv2048c	2300545	Ref. 135
Rv2048c	2300546	Ref. 135
Rv2048c	2300561	Ref. 119

Supplementary Table 4.3 (continued)

Rv2048c	2300564	Ref. 119
Rv2048c	2302031	Ref. 135
Rv2048c	2302032	Ref. 135
Rv2048c	2302033	Ref. 135
Rv2048c	2304066	Ref. 135
Rv2048c	2304067	Ref. 135
Rv2048c	2304068	Ref. 135
Rv2048c	2304236	Ref. 135
Rv2048c	2304237	Ref. 135
Rv2048c	2304238	Ref. 135
Rv2048c	2306315	Ref. 119
Rv2082	2338773	Ref. 135
Rv2082	2338774	Ref. 135
Rv2082	2338775	Ref. 135
Rv2082	2338811	Ref. 135
Rv2082	2338812	Ref. 135
Rv2082	2338813	Ref. 135
Rv2082	2338921	Ref. 119
Rv2082	2338994	Ref. 135
Rv2082	2338995	Ref. 135
Rv2082	2338996	Ref. 135
Rv2082	2339255	Ref. 135
Rv2082	2339256	Ref. 135
Rv2082	2339257	Ref. 135
Rv2082	2339261	Ref. 135
Rv2082	2339262	Ref. 135
Rv2082	2339263	Ref. 135
Rv2082	2339269	Ref. 119
Rv2082	2339308	Ref. 119
Rv2082	2339309	Ref. 119
Rv2082	2339310	Ref. 119
Rv2082	2339525	Ref. 135
Rv2082	2339526	Ref. 135
Rv2082	2339527	Ref. 135
Rv2082	2339835	Ref. 135
Rv2082	2339836	Ref. 135
Rv2082	2339837	Ref. 135
Rv2082	2340104	Ref. 135

Supplementary Table 4.3 (continued)

Rv2082	2340105	Ref. 135
Rv2082	2340106	Ref. 135
Rv2082	2340623	Ref. 135
Rv2102	2363691	Ref. 119
Rv2112c	2372502	Ref. 119
Rv2112c	2372559	Ref. 119
Rv2155c	2415501	Ref. 135
Rv2155c	2415502	Ref. 135
Rv2155c	2415503	Ref. 135
Rv2155c	2416154	Ref. 135
Rv2155c	2416155	Ref. 135
Rv2155c	2416156	Ref. 135
Rv2182c	2444629	Ref. 119
Rv2205c	2470147	Ref. 135
Rv2205c	2470148	Ref. 135
Rv2205c	2470149	Ref. 135
Rv2258c	2531618	Ref. 119
Rv2291	2563472	Ref. 119
Rv2346c	2625932	Ref. 119
Rv2346c	2626113	Ref. 119
Rv2346c	2626116	Ref. 119
Rv2346c	2626118	Ref. 119
Rv2348c	2626686	Ref. 119
Rv2361c	2643278	Ref. 119
Rv2402	2700247	Ref. 119
Rv2427A	2725653	Ref. 120
Rv2427A	2725700	Ref. 120
Rv2427A	2725802	Ref. 120
Rv2436	2733749	Ref. 135
Rv2436	2733750	Ref. 135
Rv2436	2733751	Ref. 135
Rv2436	2734074	Ref. 135
Rv2436	2734075	Ref. 135
Rv2436	2734076	Ref. 135
Rv2447c	2747159	Ref. 119
Rv2447c	2747203	Ref. 119
Rv2447c	2747471	Ref. 120
Rv2447c	2747479	Ref. 119

Supplementary Table 4.3 (continued)

Rv2447c	2747488	Ref. 119
Rv2463	2765896	Ref. 119
Rv2476c	2778964	Ref. 119
Rv2540c	2863716	Ref. 119
Rv2544	2867764	Ref. 119
Rv2561	2881463	Ref. 119
Rv2561	2881480	Ref. 119
Rv2566	2886547	Ref. 119
Rv2605c	2932443	Ref. 119
Rv2665	2982964	Ref. 119
Rv2670c	2986835	Ref. 119
Rv2679	2995856	Ref. 119
Rv2722	3034771	Ref. 119
Rv2723	3035380	Ref. 119
Rv2741	3054665	Ref. 135
Rv2741	3054666	Ref. 135
Rv2741	3054667	Ref. 135
Rv2764c	3074238	Ref. 119
Rv2764c	3074408	Ref. 120
Rv2765	3075294	Ref. 119
Rv2782c	3089259	Ref. 119
Rv2787	3095501	Ref. 119
Rv2789c	3098318	Ref. 119
Rv2804c	3112885	Ref. 119
Rv2823c	3130012	Ref. 119
Rv2823c	3131473	Ref. 135
Rv2828A	3136343	Ref. 119
Rv2853	3162632	Ref. 135
Rv2853	3162633	Ref. 135
Rv2853	3162634	Ref. 135
Rv2853	3164083	Ref. 135
Rv2853	3164084	Ref. 135
Rv2853	3164085	Ref. 135
Rv2865	3177552	Ref. 119
Rv2916c	3224892	Ref. 119
Rv2931	3247317	Ref. 135
Rv2931	3247318	Ref. 135
Rv2931	3247324	Ref. 119

Supplementary Table 4.3 (continued)

Rv2931	3248074	Ref. 135
Rv2931	3248075	Ref. 135
Rv2931	3248076	Ref. 135
Rv2931	3248082	Ref. 119
Rv2931	3248083	Ref. 119
Rv2931	3248190	Ref. 135
Rv2931	3248191	Ref. 135
Rv2931	3248192	Ref. 135
Rv2931	3248308	Ref. 135
Rv2931	3248309	Ref. 135
Rv2931	3248310	Ref. 135
Rv2931	3249027	Ref. 135
Rv2931	3249411	Ref. 135
Rv2931	3249412	Ref. 135
Rv2931	3249413	Ref. 135
Rv2932	3251428	Ref. 119
Rv2932	3253560	Ref. 119
Rv2934	3266777	Ref. 119
Rv2947c	3297348	Ref. 119
Rv2986c	3343419	Ref. 119
Rv3021c	3379430	Ref. 135
Rv3021c	3379431	Ref. 135
Rv3021c	3379432	Ref. 135
Rv3021c	3379786	Ref. 135
Rv3021c	3380080	Ref. 135
Rv3021c	3380081	Ref. 135
Rv3021c	3380082	Ref. 135
Rv3064c	3430088	Ref. 119
Rv3081	3446707	Ref. 119
Rv3082c	3448009	Ref. 119
Rv3108	3477274	Ref. 119
Rv3134c	3500157	Ref. 119
Rv3165c	3534196	Ref. 119
Rv3169	3538160	Ref. 119
Rv3190c	3555465	Ref. 119
Rv3224	3600338	Ref. 119
Rv3228	3604821	Ref. 135
Rv3228	3604822	Ref. 135

Supplementary Table 4.3 (continued)

Rv3228	3604823	Ref. 135
Rv3245c	3625505	Ref. 135
Rv3245c	3625506	Ref. 135
Rv3245c	3625507	Ref. 135
Rv3245c	3625949	Ref. 119
Rv3245c	3626091	Ref. 135
Rv3245c	3626092	Ref. 135
Rv3245c	3626093	Ref. 135
Rv3245c	3626388	Ref. 135
Rv3245c	3626389	Ref. 135
Rv3245c	3626390	Ref. 135
Rv3245c	3626487	Ref. 135
Rv3245c	3626488	Ref. 135
Rv3245c	3626489	Ref. 135
Rv3296	3679971	Ref. 119
Rv3296	3680940	Ref. 119
Rv3297	3681535	Ref. 119
Rv3312c	3700229	Ref. 119
Rv3343c	3730392	Ref. 135
Rv3343c	3730393	Ref. 135
Rv3343c	3730394	Ref. 135
Rv3343c	3730614	Ref. 135
Rv3343c	3730615	Ref. 135
Rv3343c	3730616	Ref. 135
Rv3343c	3732342	Ref. 135
Rv3343c	3732343	Ref. 135
Rv3343c	3732344	Ref. 135
Rv3343c	3732859	Ref. 135
Rv3343c	3732860	Ref. 135
Rv3343c	3732861	Ref. 135
Rv3343c	3734187	Ref. 135
Rv3343c	3734188	Ref. 135
Rv3343c	3734189	Ref. 135
Rv3343c	3735079	Ref. 135
Rv3343c	3735080	Ref. 135
Rv3343c	3735081	Ref. 135
Rv3343c	3735538	Ref. 135
Rv3343c	3735539	Ref. 135

Supplementary Table 4.3 (continued)

Rv3343c	3735540	Ref. 135
Rv3343c	3735811	Ref. 135
Rv3343c	3735812	Ref. 135
Rv3343c	3735813	Ref. 135
Rv3343c	3736626	Ref. 135
Rv3343c	3736627	Ref. 135
Rv3343c	3736628	Ref. 135
Rv3345c	3738550	Ref. 135
Rv3345c	3738551	Ref. 135
Rv3345c	3738552	Ref. 135
Rv3345c	3738674	Ref. 135
Rv3345c	3738675	Ref. 135
Rv3345c	3738676	Ref. 135
Rv3345c	3738683	Ref. 135
Rv3345c	3738684	Ref. 135
Rv3345c	3738685	Ref. 135
Rv3345c	3740374	Ref. 135
Rv3345c	3740375	Ref. 135
Rv3345c	3740376	Ref. 135
Rv3345c	3741356	Ref. 135
Rv3345c	3741357	Ref. 135
Rv3345c	3741358	Ref. 135
Rv3345c	3742135	Ref. 135
Rv3345c	3742136	Ref. 135
Rv3345c	3742137	Ref. 135
Rv3345c	3742393	Ref. 135
Rv3345c	3742394	Ref. 135
Rv3345c	3742395	Ref. 135
Rv3347c	3744093	Ref. 135
Rv3347c	3744094	Ref. 135
Rv3347c	3744095	Ref. 135
Rv3347c	3746407	Ref. 135
Rv3347c	3746408	Ref. 135
Rv3347c	3746409	Ref. 135
Rv3347c	3747886	Ref. 135
Rv3347c	3747887	Ref. 135
Rv3347c	3747888	Ref. 135
Rv3347c	3750208	Ref. 135

Supplementary Table 4.3 (continued)

Rv3347c	3750209	Ref. 135
Rv3347c	3750210	Ref. 135
Rv3347c	3750415	Ref. 135
Rv3347c	3750416	Ref. 135
Rv3347c	3750417	Ref. 135
Rv3347c	3750419	Ref. 135
Rv3347c	3750420	Ref. 135
Rv3347c	3750421	Ref. 135
Rv3347c	3750826	Ref. 135
Rv3347c	3751255	Ref. 135
Rv3347c	3751256	Ref. 135
Rv3347c	3751257	Ref. 135
Rv3347c	3752819	Ref. 135
Rv3347c	3752820	Ref. 135
Rv3347c	3752821	Ref. 135
Rv3363c	3774269	Ref. 119
Rv3378c	3793110	Ref. 119
Rv3383c	3798070	Ref. 119
Rv3436c	3855220	Ref. 119
Rv3446c	3863346	Ref. 135
Rv3446c	3863347	Ref. 135
Rv3446c	3863348	Ref. 135
Rv3446c	3863679	Ref. 135
Rv3446c	3863680	Ref. 135
Rv3446c	3863681	Ref. 135
Rv3446c	3863812	Ref. 135
Rv3446c	3863813	Ref. 135
Rv3446c	3863814	Ref. 135
Rv3457c	3877946	Ref. 119
Rv3478	3894732	Ref. 135
Rv3478	3894733	Ref. 135
Rv3478	3894734	Ref. 135
Rv3478	3894784	Ref. 135
Rv3478	3894785	Ref. 135
Rv3478	3894786	Ref. 135
Rv3478	3894862	Ref. 135
Rv3478	3894863	Ref. 135
Rv3478	3894864	Ref. 135

Supplementary Table 4.3 (continued)

Rv3478	3895536	Ref. 135
Rv3478	3895537	Ref. 135
Rv3478	3895538	Ref. 135
Rv3478	3895585	Ref. 135
Rv3478	3895586	Ref. 135
Rv3478	3895587	Ref. 135
Rv3479	3895891	Ref. 119
Rv3507	3926968	Ref. 135
Rv3507	3926969	Ref. 135
Rv3507	3926970	Ref. 135
Rv3507	3926984	Ref. 135
Rv3507	3926985	Ref. 135
Rv3507	3926986	Ref. 135
Rv3507	3927578	Ref. 135
Rv3507	3927579	Ref. 135
Rv3507	3927580	Ref. 135
Rv3507	3929098	Ref. 135
Rv3507	3929099	Ref. 135
Rv3507	3929100	Ref. 135
Rv3507	3930476	Ref. 135
Rv3507	3930477	Ref. 135
Rv3507	3930478	Ref. 135
Rv3589	4030699	Ref. 119
Rv3711c	4156501	Ref. 135
Rv3711c	4156502	Ref. 135
Rv3711c	4156503	Ref. 135
Rv3768	4214150	Ref. 119
Rv3795	4247432	Ref. 135
Rv3795	4247433	Ref. 135
Rv3795	4247436	Ref. 119
Rv3795	4247476	Ref. 119
Rv3795	4247498	Ref. 135
Rv3795	4247553	Ref. 135
Rv3795	4247554	Ref. 135
Rv3795	4247555	Ref. 135
Rv3795	4247576	Ref. 135
Rv3795	4247732	Ref. 135
Rv3795	4247737	Ref. 119

Supplementary Table 4.3 (continued)

Rv3795	4248005	Ref. 135
Rv3795	4248009	Ref. 119
Rv3795	4248010	Ref. 119
Rv3795	4248320	Ref. 135
Rv3795	4248321	Ref. 135
Rv3795	4248322	Ref. 135
Rv3795	4249525	Ref. 119
Rv3795	4249590	Ref. 119
Rv3798	4254290	Ref. 135
Rv3806c	4269096	Ref. 119
Rv3806c	4269305	Ref. 119
Rv3806c	4269317	Ref. 120
Rv3839	4313163	Ref. 119
Rv3848	4322607	Ref. 119
Rv3854c	4326331	Ref. 135
Rv3854c	4326332	Ref. 135
Rv3854c	4326333	Ref. 135
Rv3854c	4326568	Ref. 135
Rv3854c	4326569	Ref. 135
Rv3854c	4326570	Ref. 135
Rv3854c	4326712	Ref. 135
Rv3854c	4326713	Ref. 135
Rv3854c	4326714	Ref. 135
Rv3854c	4326722	Ref. 135
Rv3854c	4326723	Ref. 135
Rv3854c	4326724	Ref. 135
Rv3854c	4326856	Ref. 135
Rv3854c	4326857	Ref. 135
Rv3854c	4326858	Ref. 135
Rv3854c	4327020	Ref. 135
Rv3854c	4327021	Ref. 135
Rv3854c	4327022	Ref. 135
Rv3854c	4327072	Ref. 119
Rv3854c	4327245	Ref. 135
Rv3854c	4327324	Ref. 135
Rv3854c	4327325	Ref. 135
Rv3862c	4338365	Ref. 120
Rv3862c	4338372	Ref. 119

Supplementary Table 4.3 (continued)

Rv3879c	4359172	Ref. 119
Rv3887c	4371255	Ref. 119
Rv3910	4398755	Ref. 119
Rv3919c	4407766	Ref. 135
Rv3919c	4407767	Ref. 135
Rv3919c	4407902	Ref. 135
Rv3919c	4407903	Ref. 135
Rv3919c	4407904	Ref. 135
Rv3919c	4407914	Ref. 135
Rv3919c	4407925	Ref. 135
Rv3919c	4407926	Ref. 135
Rv3919c	4407965	Ref. 135
Rv3919c	4408052	Ref. 135
Rv3919c	4408053	Ref. 135
Rv3919c	4408054	Ref. 135
Rv3919c	4408059	Ref. 135
Rv3919c	4408060	Ref. 135
Rv3919c	4408061	Ref. 135
Rv3919c	4408154	Ref. 135
Rv3919c	4408155	Ref. 135
Rv3919c	4408156	Ref. 135