

https://dspace.org/

Mutation rates in mycobacterial hosts with altered Dna metabolic activity

Barichievy, Samantha

2006-02-08
http://hdl.handle.net/10539/163

Downloaded from a DSpace repository. https://dspace.org/

Mutation rates in mycobacterial hosts with altered DNA metabolic activity

Samantha Barichievy

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfillment of the requirements for the degree of Master of Science

Johannesburg, February 2005

Do not meddle in the affairs of Dragons, because you are crunchy and taste good with ketchup

Table of Contents

	Page
Declaration	i
Preface	ii
Abstract	iii
Acknowledgements	iv
List of Figures	v
List of Tables	xii
Glossary	xiii

Chapter 1: Introduction

1.1 The worldwide burden of tuberculosis	1
1.2 The BCG vaccine	3
1.3 Macrophages, granulomas and the host response	4
1.4 Chemotherapy	6
1.5 MDR-TB	9
1.6 Environment-dependent and heritable mutators	11
1.6.1 Environment-dependent mutators	11
1.6.2 Heritable global mutators	12
1.6.3 Heritable local mutators	13
1.7 Identification of the PE and PPE multigene families	13
1.8 Polymorphisms in the PE and PPE genes	16
1.9 Localisation of proteins encoded by PE and PPE genes	18
1.10 Potential biological functions of proteins encoded by PE and PPE genes	18
1.11 Genetic diversity of pathogens involving repeat sequences	20
1.11.1 Antigenic variation	20
1.11.2 Phase variation	21
1.12 Mutation rate analysis	26

1.12.1 The original mutation rate experiments of Luria and Delbrück..... **26**
1.12.2 The Fluctuation Test..... **29**
1.13 Aims of this study..... **31**

Chapter 2: Materials and Methods

2.1 Media and general solutions..... **33**
2.2 Storage of bacterial strains..... **33**
2.3 Culture conditions..... **33**
2.4 Medium supplements..... **34**
2.5 Plasmid DNA..... **34**
2.6 Transformation of cells..... **36**
2.6.1 *E. coli* DH5α..... **36**
2.6.2 Electroporation of *M. smegmatis*..... **36**
2.7 DNA Extraction Protocols..... **37**
2.7.1 Isolation of plasmid DNA..... **37**
2.7.2 Cetyltrimethylammoniumbromide (CTAB) method for isolation of chromosomal DNA from *M. smegmatis*..... **37**
2.7.3 Precipitation of DNA..... **38**
2.8 Agarose gel electrophoresis..... **38**
2.8.1 Preparation of gels..... **38**
2.8.2 Electrophoresis..... **38**
2.8.3 Purification of DNA from agarose gels..... **38**
2.9 DNA manipulations..... **39**
2.9.1 Restriction endonuclease digests..... **39**
2.9.2 Blunt-ending of 3' or 5' overhangs..... **39**
2.9.3 Removal of 5' phosphate from DNA..... **40**
2.9.4 Ligations..... **40**
2. 10 The Morf2 reporter assay..... **40**
2.10.1 Design of the Morf2 substrate..... **40**
2.10.2 Cloning of the pGRAK reporter..... **43**
2.11 Fluctuation Tests..... **43**
2.11.1 Experimental protocol for mutation rate assessment..... **43**
2.11.2 Adaptation of experimental protocol for mutation rate assessment of mc^2 155 *his*G380E strain..... **45**
2.11.3 Calculation of mutation rates..... **46**

2.12 PCR and sequencing..... **50**
2.12.1 PCR of pGRAK based strains..... **50**
2.12.2 PCR of *M. smegmatis* mc^2 155 *his*G380E strain..... **52**
2.12.3 PCR of *attL* and *attR* regions..... **52**
2.12.4 DNA sequencing..... **53**

Chapter 3: Results

3.1 Optimisation of experimental parameters..... **55**
3.1.1 Growth curve of *M. smegmatis* mc^2 155..... **55**
3.1.2 Confirmation of the integrity of the Morf2 region..... **56**
3.1.3 Confirmation of site-specificity of pGRAK reporter..... **58**
3.1.4 Gm concentration range..... **59**
3.1.5 Antibiotic cross-resistance..... **61**
3.1.6 Optimisation of methodology to minimise variability in N_t values..... **62**
3.1.7 Dependence of the μ value on the growth phase of the culture..... **65**
3.2 Mutation rate assessment: Fluctuation Test data..... **65**
3.2.1 Overall *m* and N_t values at log and saturation phase..... **65**
3.2.2 μ values in pGRAK reporter-based strains..... **68**
3.2.3 Rate of mutation to Rif^R **69**
3.2.4 Rate of A→G base substitution mutagenesis with the *his*G380E reporter... **69**
3.2.5 Statistical analysis of results..... **70**
3.3 Mutation frequencies..... **70**
3.4 Spectrum of mutations in pGRAK based strains..... **71**
3.4.1 From Fluctuation Test to mutant genotyping..... **71**
3.4.2 Multiple genotypes recovered from a single tube..... **73**

Chapter 4: Discussion and Conclusions

4.1 Optimisation of experimental parameters..... **79**
4.1.1 Confirmation of the integrity of the integrating and replicative vectors..... **79**
4.1.2 Optimisation of Fluctuation Test methodology for use in mycobacteria..... **79**
4.1.3 Deviations from Luria and Delbrück assumptions..... **80**
4.2 Mutation rate assessment..... **81**
4.2.1 μ value in the *recA*-deficient pGRAK reporter-based strain..... **81**
4.2.2 μ values in pGRAK reporter-based strains with altered levels of

expression of Y-family and C-family polymerases..... **82**
4.2.3 Rates of base substitution mutagenesis in *M. smegmatis*..... **83**
4.2.4 Limitations to reporter-based assays..... **84**
4.3 Spectrum of mutations associated with PE-PGRS genes..... **84**
4.4 Conclusions..... **86**

Appendices

Appendix A : Media and some general solutions..... **87**
Appendix B : Plasmid maps..... **89**
Appendix C : Specific experimental requirements for Fluctuation Tests..... **92**
Appendix D : Techniques to limit clumping of mycobacterial cultures..... **94**
Appendix E : Additional sequencing data from Morf2 region..... **96**

References..... **98**

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the degree of Master of Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.

Samantha Barichievy

Date

Preface

Some aspects of the work conducted for this dissertation have been presented as a poster elsewhere:

- Machowski, E.E., Barichievy, S., Brackin, R., Warner, D.F., Boshoff, H., Stoker, N. G., Mizrahi, V. (2004). "Application of fluctuation analysis to determine mycobacterial mutation rates". Molecular and Cell Biology Group (MCBG) Symposium. University of the Witwatersrand Medical School Campus.

Abstract

The completion of the genome sequence of *Mycobacterium tuberculosis* strain H37Rv revealed that 10% of the coding capacity is devoted to two, large multigene families that are characterised by repeat sequences. These are the PE and PPE families that code for acidic, glycine rich proteins. A subgroup of the PE family is the polymorphic GC rich sequence (PGRS) gene subfamily. Genome comparisons of clinical isolates of *M. tuberculosis* have confirmed the polymorphic character of some of these genes suggesting they may be analogous to the contingency loci found in other pathogenic bacteria. Certain PE-PGRS proteins play a direct role in virulence in *M. marinum*, other PE-PGRS genes are cell surface associated, and some PE-PGRS proteins are variable surface antigens, supporting a potential role in host pathogen interactions. A reporter assay designed to investigate mutations in a PE-PGRS repeat-containing sequence was used to assess mutation rates in various *M. smegmatis* host strains by fluctuation analysis. A wide spectrum of mutations was observed and the evidence suggests that slipped-strand mispairing between proximal and distal PGRS sequences located *in cis* is the predominant type of mutational event at such loci. Moreover, slipped-strand mispairing at such loci occurs at a moderately higher rate than base substitution mutagenesis and is mediated by the normal replicative polymerase.

Acknowledgements

There are many people who helped me during my MSc and I would like to acknowledge all of their individual contributions.

To my supervisor, Professor Valerie Mizrahi, thank you for all your input on my project, for your constructive criticism and for teaching me so much more than just science.

To my co-supervisor, Dr. Edith Machowski, thank you for all your guidance, for answering my endless questions, for your patience and kindness and for always being so bubbly and enthusiastic. I really appreciate all that you shared with me.

To my other office colleagues, Digby Warner and Limenako Matsoso, thank you for answering my unending questions, for explaining things more than once, for making me laugh and for putting up with the Lab Oinker.

To Dr. Bavesh Kana, Dr. Bhavna Gordhan and Dr. Stephanie Dawes, thank you for teaching me how to be competent in pouring gels (Bavesh), work in the TB lab (Bhav) and for all the Q and D's (Steph). Thanks too for the great chats, for making me laugh and for letting me act like a monkey.

To Sinah, Julia and Leah, thank you for making life easier. I was very spoilt to have the three of you and really appreciate all your hard work.

To my friend Robyn Brackin, thank you for the hours we spent chatting, for the tea and the moans and the support. You brought the fun.

To Dr. Marco Weinberg, thank you for all your constructive criticism regarding the first drafts of my write-up. You really gave of your time and I truly appreciate it.

To my mom and dad, thank you for all the support and the encouragement and for being interested when you had no clue what I was studying. Your guidance and support are priceless.

To my brother, C, thank you for helping with the stress, for making me laugh and for always reminding me of the big picture.

To my other folks, Bev and Graham, thank you for all the support, love and the endless chats. I really appreciate your time and for being there whenever I needed advice.

To my babe, Liam, thank you for the tickles, for the coffee and the hugs, thank you for the help whenever I was stuck or ready to leave, thank you for all that you are and remember that I appreciate you every day.

Thanks are also extended to Dr. Stephen Durbach for starting the ball rolling, for making plasmids 9/19 and pMorf2 and for the lab books – they were a challenge.

To anyone I have not mentioned, please just remember that you all helped in some way and I appreciate you.

List of Figures

Page

Figure 1: The mycobacterial cell envelope. The plasma membrane is composed of a phospholipid bilayer with integral proteins. The periplasmic space contains the peptidoglycan and arabinogalactan (indicated by orange chain) with long chain mycolic acids covalently bound to the latter polymers. The outer capsule-like coat is composed mostly of polysaccharides. PE-PGRS proteins that are localised at the cell surface are possibly bound to the peptidoglycan layer allowing them to interact with the extracellular milieu..... **8**

Figure 2: Chromosomal organisation of the conserved loci of *dinX* and *dinP* in *M. smegmatis* as well as a third homologous *umuC*-like gene, *dinP3*. The genes are indicated by dark red arrows with the gene numbers as annotated by TIGR (http://www.tigr.org) indicated at the bottom of the figure. These genes encode putative DNA polymerases belonging to the UmuC-DinB-Rad30-Rev1 or Y family of polymerases. The green arrows indicate other annotated genes in the *M. smegmatis* genome. The blue arrows indicate genes in *M. smegmatis* with homology to annotated genes from *M. tuberculosis*..... **12**

Figure 3: Members of the PE and PPE protein families. (A) The two main groups within the PE family of which the PE-PGRS proteins are the most abundant. Both groups share a common N-terminal domain sequence comprised of ~110 amino acids (indicated by orange blocks). The C-terminal domains are divergent and range in length from 1 to > 1400 amino acids (indicated by green blocks). The PE-PGRS proteins have a common amino acid sequence, GGAGG, repeated in tandem. (B) The three main groups within the PPE family. All three groups share a common N-terminal domain composed of ~180 amino acids (indicated by purple blocks). The C-terminal domains are divergent and range in length from 1 to > 3500 amino acids (indicated by yellow blocks). The PPE-MPTR proteins have a repeat sequence NXGXGNXG where X can be any amino acid..... **15**

Figure 4: Sequence variation that exists between *M. tuberculosis* H37Rv and *M. bovis* BCG in the PE-PGRS gene Rv0746. (A) The BCG protein has a deletion of 29 amino acids (indicated by blue arrow) and an insertion of 46 amino acids (indicated by pink arrow) compared to the sequence in the *M. tuberculosis* H37Rv counterpart. (B) Specific amino acid sequence variation between H37Rv (upper line of sequence) and BCG (lower line of sequence) in gene Rv0746. The 29 amino acid deletion in BCG is indicated by blue

letters. The 46 amino acid insertion in BCG is indicated by pink letters. The Morf2 substrate used in this study was based on the 29 amino acid deletion between the 2 sequences (refer to Section 2.10.1)..... **17**

Figure 5: Organisation of an *opa* gene showing the control mechanism of gene expression. The pink arrow indicates the functional promoter, part of each unique *opa* gene. The orange block indicates the ATG start codon of the *opa* gene. The blue block indicates the repeat region of the gene containing various numbers of the CTCTT repeat unit (coloured pink or black). Together with the ATG start codon, the repeat region forms the hydrophobic signal sequence. The green block indicates the downstream portion of the *opa* gene. The number of CTCTT repeat units within the repeat region affects the open reading frame of the *opa* gene: if the reading frame is incorrect (for example 5 or 6 CTCTT units are present), no Opa protein is produced. If the repeat units are in frame (for example 7 CTCTT units) a functional Opa protein will be translated. Alterations in the reading frame between the CTCTT repeats are RecA independent and hypothesized to be due to replication error..... **23**

Figure 6: RecA independent misalignment events involving repeat sequences. (A) Deletion misalignment events are formed when there is backward slippage between repeat sequences (indicated by red blocks). After replication of the slippage intermediate, one repeat and the intervening DNA sequence (indicated by green line) are deleted. (B) Duplication misalignment events are formed when there is forward slippage between the repeat sequences. Replication of this slippage intermediate produces an expanded number of repeats including a duplication of the intervening sequence..... **25**

Figure 7: Schematic diagrams indicating the differences between the Mutation Hypothesis (A and B) and the Acquired Immunity Hypothesis (C and D) tested by Luria and Delbrück. (A) According to the Mutation Hypothesis, resistant bacteria will occur as a group of closely related individuals as each mutant gives rise to clonal progeny. Furthermore if resistance is due to random genetic events occurring prior to the onset of a selective pressure, the number of resistant cells within a bacterial population will increase over time (B) because the existing mutants grow and divide while new mutants are also added to the resistance pool. Alternately, the Acquired Immunity Hypothesis states that cells which mutate to survive the selective pressure do so at random under the selective conditions. Their descendants may not do so and their subsequent resistance distributions will also be random (C). In addition, if resistance mutations are acquired only once a selective pressure is present, the number of resistant cells within a population will remain constant over time (D). This is because each cell has a small, constant and independent probability of sustaining a resistance mutation prior to selection within its lifetime. For simplicity, each cell in Figure 7C only gives rise to a single progeny. Resistant bacteria are indicated by green circles while non-resistant bacteria are indicated by purple circles..... **27**

Figure 8: Design of Km^R based reporter, pGRAK that integrates at the *attB* locus of *M. smegmatis* mc^2 155 strains. The RAK fragment is defined by a *Xba*I site at the 5' end and an *Asp*718 site at the 3' end. The pink arrow indicates the Antigen 85A promoter (Ag85A) with ribosome binding site (RBS) and ATG start codon. The Morf2 region is defined by a *Bam*HI site at the 5' end and a *Nhe*I site at the 3' end. This region is comprised of three PE-PGRS repeat sequences, called R1 (green block), R3 (yellow block) and R2 (blue block) respectively, and an engineered TAG stop codon (red star) in the intervening sequence (grey block) between R1 and R3. The 3' end of the Morf2 region is fused in frame to a cryptic Km^R cassette ('aph) indicated by the bright green arrow. A closer look at the DNA sequence making up the Morf2 region indicates that the PE-PGRS repeats, R1, R2 and R3 share an 11 base pair (bp) core repeat sequence (indicated by underlining of bases). R1 and R2 share a 15 bp repeat sequence, except for one base difference. The red bases and stars indicate the TAG stop codon. The green bases and arrows indicate repeat 1 (R1). The yellow bases and arrows indicate repeat 3 (R3). The blue bases and arrows indicate repeat 2 (R2)..... **42**

Figure 9: Number of culture tubes (*C*) versus *m*, the number of mutational events per culture tube. For any *m* value between 10 and 100, the number of cultures tubes required to achieve a theoretical precision of 20% using the Lea-Coulsen Method (LC – shown in light blue) or MSS Maximum Likelihood Method (MSS – shown in dark blue) can be determined from the graph. Using either method to generate *m*, the number of cultures tubes required decreases as the *m* value increases. However, due to the iterative nature of the MSS Method, the number of culture tubes required by this method is lower overall..... **45**

Figure 10: The power of the rank of a data set (Equation 17) is plotted against iterative *m* values to generate a 2 tailed curve. The *m* value that correlates with the maximum Equation 17 value is read from the graph and this value is used to generate μ. In the example illustrated, the initial *m* value is 1.6 while the *m* value that maximises Equation 17 from the graph is 1.4..... **47**

Figure 11: Classical representation of the Luria-Delbrück model. The data indicates the relationship between the number of mutational events (*m*) within a tube and the number of Km^R CFU's that were scored (r). The relationship between r and Pr is tight, i.e. the R^2 value is close to 1 if the Luria-Delbrück model is a good representation of the underlying biological events. The slope of the graph indicates the time point at which the data was collected. (A) During log phase growth, fewer mutational events lead to the observed number of Km^R mutants and so the slope of the graph is steep. (B) At saturation phase growth, the slope is less steep as more mutational events occurred to give rise to the observed Km^R mutants..... **49**

Figure 12: Region of the pGRAK reporter plasmid showing the Morf2 region (between *Bam*HI and *Nhe*I sites) and downstream cryptic 'aph cassette (*Nhe*I to *Kpn*I) with relevant PCR and sequencing primers. ANAF and ANAR are forward and reverse PCR primers respectively that amplify a 282 bp product. MFSF is a forward sequencing primer. RK aph is a second reverse PCR primer used in conjunction with ANAF to amplify the Morf2 region and the cryptic 'aph cassette. The red star indicates the TAG stop codon within the Morf2 region..... **51**

Figure 13: Mechanism of integration of L5 derived vectors into the *M. smegmatis* mc^2 155 chromosome. The *attP* site on the plasmid (indicated by blue blocks on the closed green circle) lines up with the corresponding *attB* site on the bacterial chromosome (indicated by the purple blocks on the orange line). Using the phage integrase enzyme (encoded on the plasmid) the *attP* and *attB* sites are recombined, generating the *attL* and *attR* sites. All genetic material contained within the plasmid is linearised and becomes part of the bacterial chromosome. PCR primers (indicated by blue and red arrows) were designed to flank the *attL* and *attR* regions to ensure correct integration of the pGRAK vector into the *M. smegmatis* based strains used in this study..... **53**

Figure 14: Growth parameters of *M. smegmatis* mc^2 155 grown in MADC-Tw. Log CFU / mL versus Time (Hours). Data points are averages of two biological replicates assessed in duplicate. Log phase growth, as defined by CFU / mL values of $Log_{10}7$ to $Log_{10}8$, was observed between 12 and 20 hours post inoculation. Thereafter, stationary phase growth occurred..... **55**

Figure 15: Confirmation of the integrity of the Morf2 region. (A) Restriction digestion of the Gm^R cassette in the pGRAK reporter. Thirty different restriction endonucleases were used to screen the Gm^R cassette from pGRAK. pHINT was digested with the same endonucleases as a control. All restriction digests were separated on agarose gels and selected examples are illustrated. From these and other restriction digests (data not shown), it was deduced that the Gm^R cassette contains *Eco*RV, *Bgl*II, *Sph*I, *Bst*BI and *Not*I restriction sites, the first two sites being previously non-annotated. MW λIII: Molecular Weight Marker λIII, MW λVI: Molecular Weight Marker λVI, Uncut: uncut control plasmids pGRAK or pHINT. (B) Chromatogram data of the Morf2 region within the pGRAK reporter. The top panel indicates a control sample of the Morf2 region while the lower panel indicates a sample sequence selected for comparison. The 5' end of the Morf2 region is defined by a *Bam*HI site (G↓GATCC highlighted by blue block) and the 3' end is defined by a *Nhe*I site (G↓CTAGC highlighted by pink block). The stop codon (TAG highlighted by red block) is present in both sequences as are the direct repeat sequences R1 (highlighted by green block), R2 (highlighted by yellow block) and R3 (highlighted by blue block) respectively ($^{5'}$GGCCGGGGCCGG$^{3'}$)..... **57**

Figure 16: PCR products of *attL* and *attR* regions of selected pGRAK based strains used in this study. Only correct integration events would have yielded the expected PCR product. The left hand gel indicates *attL* PCR products (320 bp) and the right hand gel indicates *attR* PCR products (282 bp). Lanes 1: Molecular Weight Marker λVI, Lanes 2: control PCR product of 282 bp in length, Lanes 3: control PCR product of 263 bp in length and Lanes 4: control PCR product of 245 bp in length. Lanes 5 to 10: *attL* and *attR* PCR products from separate colonies that were picked from Gm^5 plates post electroporation..... **58**

Figure 17: Growth of mycobacterial cultures with and without Gm based plasmids over a range of antibiotic concentrations. (A) Growth of cultures in liquid medium. Wild type *M. smegmatis* (indicated as mc2 in key) did not grow at Gm concentrations higher than 1 µg / mL. pGINTO based cultures grew poorly while pGRAK based cultures grew well at Gm concentrations of 2 µg / mL. Neither Gm based plasmid facilitated growth in liquid medium at Gm concentrations higher than 2 µg / mL. (B) Growth of cultures on solid agar medium. Wild type *M. smegmatis* did not grow at Gm concentrations higher than 0.1 µg / mL. pGINTO based cultures grew poorly at any Gm concentration higher than 0.5 µg / mL. pGRAK based cultures grew well at Gm concentrations up to and including 5 µg / mL. At higher concentrations, pGRAK did not facilitate growth on solid agar. Poor growth was defined as an increase in OD_{600} values but few countable CFUs when plated on agar (indicated by arbitrary value of 0.5 graphs in A and B). Good growth was defined as CFUs of $Log_{10}7$ when plated on agar (indicated by arbitrary value of 1 on graphs in A and B)..... **60**

Figure 18: Cross-resistance between two aminoglycoside antibiotics, Gm and Km. Cultures grown in liquid media containing Gm at a concentration of 2 µg / mL developed cross resistance to low Km concentrations (indicated by 'background' growth on agar plate containing Km^{10} on left hand side of figure). When the same Gm^R cultures were plated on agar plates containing a higher Km concentration, the 'background' growth no longer appeared and only discrete Km^R CFUs grew (right hand side of figure). Sequence analysis of these Km^R CFUs revealed various mutations within the Morf2 region of the pGRAK reporter to by-pass the TAG stop codon and allow for expression of a functional Aph fusion protein..... **62**

Figure 19: Fluctuation in N_t values was dependant on the growth medium used. Ten parallel tubes that contained the same culture were grown in three different media that each contained Gm at a concentration of 2 µg / mL to maintain the plasmid within the bacteria. (A) N_t values of mc^2 155 (pGRAK) cultures grown in general LB media. (B) N_t values of mc^2 155 (pGRAK) cultures grown in enriched 2TY media. (C) N_t values of mc^2 155 (pGRAK) cultures grown in mycobacterium-specific MADC-Tw..... **63**

Figure 20: N_t values assessed at different time points during growth of mc^2 155 (pGRAK) in MADC-Tw. Early log cultures were assessed after 24 hours, mid log cultures after 48 hours and saturation cultures

after 114 hours of growth. In each case, N_t values were assessed from the same tube that was then cultured further. The N_t values (Log CFU / mL) increased over time from $Log_{10}6$ CFU / mL at early log phase to $Log_{10}9$ CFU / mL at saturation phase, however at each time point, the N_t values were comparable between tubes..... **64**

Figure 21: Chromatogram data showing the wild type (GAG) Histidine auxotroph (His-) sequence (upper panel) and the revertant (GGG) Histidine prototroph (His+) sequence (lower panel) in selected mc^2 155 *his*G380E clones. The reversion mutation from GAG to GGG is highlighted by the black box. The base substitution mutation results in a codon change from glutamate (E) to glycine (G)..... **69**

Figure 22: Flow diagram outlining the experimental procedures followed to generate μ values and to assess the spectrum of mutations per mycobacterial host background. Fluctuation Tests generated a μ value and Km^R CFU's. A random selection of these mutants were subsequently subjected to PCR to determine the type of mutation that had occurred in the Morf2 region of the pGRAK reporter that led to Km^R. Between 80 and 300 Km^R CFUs per Fluctuation Test were assessed by PCR, depending on the number of jackpots that had occurred. Once PCR was complete, a select group of PCR products were sequenced to verify the specific genotypic mutation that had occurred..... **72**

Figure 23: Spectrum of Km^R mutations observed in mc^2 155 (pGRAK) CFUs as shown by gel electrophoresis. Lane 1: Molecular Weight Marker λVI. Lanes 2 to 12: PCR products from separate Km^R colonies that were picked from a single tube. The PCR products vary in size, indicating a number of underlying mutational events..... **73**

Figure B1: Plasmids p9/19, pHINT, pGINTO and pGRAK..... **89**

Figure B2: Plasmids pOLYGaa, pOLYGaadinP and pOLYGaadinX..... **90**

Figure D1: Clumping of mycobacterial cultures. (A) Culture tubes used in Fluctuation Tests were placed in a beaker partly filled with water and secured with paper towel. Cultures grown in this manner for 2 or 6 days (B) did not clump, therefore N_t values assessed from these tubes were reproducible..... **95**

Figure E1: Sequence data of Morf2 region indicating a spectrum of mutations resulting in Km^R. G↓GATCC at the 5' end of the sequences indicates the *Bam*HI site at the start of the Morf2 region. G↓CTAGC at the 3' end of the sequence indicates the *Nhe*I site at the start of the 'aph cassette. R1 is highlighted by green blocks. The TAG stop codon is highlighted by red blocks. R2 and R3 are highlighted by blue and yellow blocks respectively. The point mutation within TAG to form a codon encoding Leucine is highlighted by the

bright green block. The point mutation to form an alternative in frame start codon (GTG) is highlighted by the orange block. IV = intervening sequence..... **96**

Figure E2: Sequence data indicating a Km^R clone with duplication in the 'aph cassette. The upper sequence indicates the wild type 'aph cassette from the *NheI* site at the 3' end of the Morf2 sequence (indicated by pink block at the 5' end of the sequence). The lower sequence indicates the Km^R mutant with the duplicated portion of the 'aph cassette (indicated by the purple blocks) that in turn created a new in frame ATG start codon (indicated by the orange block)..... **97**

List of Tables

Page

Table 1: Luria and Delbrück's Fluctuation Test confirming the Mutation Theory of spontaneous mutation. The mean is the average number of phage resistant bacteria within a set of parallel cultures. The variance is a measure of the fluctuating spread of variables about the mean..... **28**

Table 2: *M. smegmatis* strains used in this study..... **34**

Table 3: Plasmids used in this study (refer to Appendix B for plasmid maps)..... **35**

Table 4: Oligonucleotide sequences used for PCR and sequencing..... **50**

Table 5: *m* values, average N_t values and μ values assessed at different time points in various mycobacterial strains..... **67**

Table 6: Comparison of mutation frequencies and mutation rates for Km^R in mc^2 155 (pGRAK)..... **71**

Table 7: Spectrum of mutational events in the Morf2 region of the pGRAK reporter as well as mutations within the 'aph cassette that resulted in Km^R..... **75**

Table 8: Relative percentages of three different groups of mutations per pGRAK-based host strain used in this study..... **77**

Table B1: Symbol definitions of the plasmid maps in Appendix B..... **91**

Table C1: Specific experimental parameters used for each Fluctuation Test in this study..... **92**

Glossary

aph	aminoglycoside phosphotransferase gene that encodes kanamycin resistance
‘aph	cryptic *aph* cassette that does not encode its own ATG start codon
ADC	albumin, dextrose, catalase supplement for MADC-Tw
Amp	ampicillin antibiotic
attB	attachment site within the bacterial chromosome used during plasmid integration
attP	phage attachment site within the plasmid that integrates at the bacterial *attB* site
BCG	bacille Calmette-Guérin
bla	gene encoding β-lactamase for ampicillin resistance
bp	base pairs
C	number of culture tubes per Fluctuation Test
CD4+	subgroup of CTLs that form part of the human immune system
CFU	colony forming unit
CL	confidence limit
CTL	cytotoxic T lymphocyte
dinP	gene encoding DinP error prone polymerase
dinX	gene encoding DinX error prone polymerase
dNTP’s	deoxyribonucleotide triphosphates
DOTS	directly observed treatment, short course
EBV	Epstein-Barr Virus
EBNA	Epstein-Barr virus nuclear antigen
EMB	ethambutol antibiotic
EP	error-prone
G	guanine
Gm	gentamicin antibiotic
GTP	guanosine triphosphatase
his D	gene encoding histidine
His-	histidine auxotrophy
His+	histidine prototrophy
HIV	Human Immunodeficiency Virus
hyg	gene encoding hygromycin B resistance
Hyg	hygromycin B antibiotic
IFN-γ	interferon γ
INH	isoniazid antibiotic
int	gene encoding phage integrase enzyme

IV	intervening sequence
kDa	kilo Dalton
Km	kanamycin antibiotic
LA	Luria Bertani Agar
LB	Luria Bertani Broth
lacZ	gene encoding β-galactosidase
μ	mutation rate = probability of mutation per cell per division or generation
μg	microgram
μL	microlitre
m	number of mutational events per culture tube
MADC-Tw	Middlebrook 7H9 broth supplemented with ADC and 0.1% Tween 80
MDR-TB	multi-drug resistant *Mycobacterium tuberculosis*
MHC	major histocompatibility complex
MIC	minimal inhibitory concentration of drug used to prevent growth of a pathogen
Morf2	reporter construct containing 3 repeat elements and engineered stop codon
MMR	mismatch repair
MPTR	major polymorphic tandem repeats
MSP-1	major surface protein 1 antigen located on *Aniplasma marginale*
MSS	Ma-Sandri-Sarkar Maximum Likelihood Method for calculating μ
N_o	initial population numbers within a single fluctuation assay tubes
N_t	final population numbers within a single fluctuation assay tubes
NO	nitric oxide
NOS2	nitric oxide synthase
OADC	oxalic acid, albumin, dextrose, catalase supplement for Middlebrook 7H10 agar
OD_{600}	optical density at 600 nm
Opa	opacity proteins expressed on the surface of *Neisseria gonorrhoeae*
OriM	origin of replication used in *Mycobacterium smegmatis* plasmids
PE	proline glutamate
PGRS	polymorphic GC rich repetitive sequence
PknG	protein kinase G
P_o	distribution of culture tubes yielding no mutants
POA	pyrazinoic acid
PPE	proline proline glutamate
PZA	pyrazinamide antibiotic
r	observed number of mutants scored per selective plate
^{R}	resistant
RAK	Repeat Aminoglycoside phosphotransferase Kanamycin
RBS	ribosome binding site

RecA	protein involved in DNA recombination events
RFLP	restriction fragment length polymorphism
Rif	rifampicin antibiotic
RNI's	reactive nitrogen intermediates
ROI's	reactive oxygen intermediates
rpm	revolutions per minute
S	sensitive
Str	streptomycin antibiotic
TB	tuberculosis
Tw	Tween 80
U	unit
Vaa	variable adherence associated proteins found on *Mycoplasma hominis*