Disabling the intrinsic resistome of Mycobacterium tuberculosis: elucidating hierarchies of DNA repair and mutagenesis that undermine current antibiotic efficacy.

Irene Gobe

Molecular Mycobacteriology Research Unit

Division of Medical Microbiology

Department of Pathology

A thesis submitted to the Faculty of Health Sciences, University of Cape Town, in fulfilment of the requirements for the degree of Doctor of Philosophy. March 2021

Dedication

I dedicate this work to my children, Ludo Sedilame Gobe and Gofaone Unaswi Gobe who gave me hope and strength to complete this work. I love you girls!

Children are like wet cement whatever falls on them makes an impression.” — Haim Ginott, child psychologist

Declaration

I declare that this thesis is my own unaided work, it is being submitted for the degree of Doctor of Philosophy at the University of Cape Town. It has not been submitted for any degree of examination at any other university**.**

Signed by candidate

Irene Gobe

March 15, 2021

Date

Abstract

DNA damage repair mechanisms are critical to the adaptive evolution of *Mycobacterium tuberculosis* as obligate human pathogen, including the emergence of drug-resistance during anti-tuberculosis (TB) chemotherapy. In experimental models, DnaE2-dependent translesion synthesis (TLS) and UvrB-dependent nucleotide excision repair (NER) have been identified as major mediators of DNA damage tolerance and repair, respectively. Given the inferred dominance of these pathways, this thesis aimed to elucidate otherwise cryptic repair mechanisms which might buffer loss of DnaE2 and UvrB in bacilli exposed to genotoxic stress. Using *dnaE2* and *uvrB* deletion mutants of the model mycobacterium, *M. smegmatis* (MSM) mc^2155, we applied genome-wide transposon (Tn) mutagenesis to identify conditionally essential repair pathways under treatment with genotoxins of different mechanistic classes. To this end, the DNA crosslinking agent, mitomycin C (MMC), and the gyrase inhibitor and clinically relevant TB drug, moxifloxacin (MOX), were used. The goal was to reveal potential targets for co-drugs that might shorten treatment duration and reduce the risk of drug resistance by severely limiting the intrinsic capacity of MTB to tolerate lethal drugs for extended periods.

Among others, our analysis identified GlgB, which is involved in glycogen biosynthesis, and mycothiol biosynthesis proteins, MSMEG_0933 and MSMEG_5261, as compensating the absence of UvrB during MMC treatment. Under MOX treatment, the absence of UvrB was compensated by the RecC/Single-strand Annealing pathway. In contrast, DnaE2 deficiency revealed the conditional essentiality of the PadR family transcriptional regulator, MSMEG_2868, under MMC exposure. Importantly, in all cases, results from the Tn screen were validated using CRISPR interference targeting the identified genes.

Of particular interest, we observed that UvrB was essential to compensate loss of DnaE2, whereas the reciprocal was less definitive: while DnaE2 appeared dispensable in MMC-treated Δ*uvrB*, Tn analyses suggested that *dnaE2* might be essential in the untreated Δ*uvrB* mutant. This result, which is consistent with very recent results suggesting the co-ordination of NER and DnaE2 functions in *Caulobacter crescentus*, is intriguing in potentially revealing a previously unappreciated role for DnaE2 in mycobacterial NER function. Taken together, these results support the utility of Tn-based whole-genome screens in revealing unexpected gene-gene interaction networks, and provide additional impetus to explore ancillary, non-essential metabolic functions as alternative targets for novel combination therapies designed to cripple intrinsic mechanisms of mycobacterial resistance.

Acknowledgements

I would like to thank **God** for the grace he granted me to complete this work.

Indeed, it wasn't by might but your power Lord!

I am grateful for the financial support I received from the **Molecular Mycobacteriology Research Unit (MMRU)** at the University of Cape Town. They covered all the bench fees and toped up the allowance I got from the University of Botswana. I am also thankful for the financial support and opportunity that the University of Botswana accorded me.

To my supervisor **Prof. Digby Warner**, Thank you for your support, encouragement, and patience. There were times I wanted to quit but the thought of disappointing you considering the support you were giving me kept me pushing forward.

To my co-supervisor **Prof. Valerie Mizrahi,** it was an honor to be under your leadership. The wealth of knowledge you have in the field was inspiring!

To my co-supervisor **Prof. Thomas R. Iorger**, I will forever be grateful for the help you offered during my data analysis.

I offer my gratitude to my colleagues at the MMRU. I have learnt the value of team work through you guys. **Dr Mandy Mason** and **Rendani Mbau**, the TnSeq group, thank you for your support. Thank you Rendani for training me in the early days of my training. **Dr Anastasia Koch**, thank you for your valuable help in analyzing my WGS data. Thank you **Timothy de Wet** for your guidance on CRISPRi technique and for the data that you generated which provided a useful comparison with my data.

To my husband, **Gobe Tanthuma**, thank you for your support, encouragement and for being a father and mother to our kids in my absence. You sacrificed a lot and I will forever be grateful. To my lovely daughters, **Ludo** and **Gofaone**, you kept me going gilrs! I hope that one day you will understand the sacrifice and hard work that mommy and daddy gave for the betterment of your welfare.

Presentations and publications

DE WET, T. J., **GOBE, I**., MHLANGA, M. M. & WARNER, D. F. 2018. CRISPRi-Seq for the Identification and Characterisation of Essential Mycobacterial Genes and Transcriptional Units. *bioRxiv***,** 358275

GOBE, I., MBAU, R.D., DE WET, T.J., MASON, M., MIZRAHI, V & WARNER, D. F. Disabling the intrinsic resistome of Mycobacterium tuberculosis; Elucidating hierarchies of DNA repair and mutagenesis that undermine current antibiotic efficacy. Poster presented at the Keystone symposia on Tuberculosis: mechanisms, pathogenesis and treatment, Banff, Alberta, Canada, January 17- 21, 2019.

GOBE, I., MBAU, R. D., MASON, M., MIZRAHI, V. & WARNER, D. F. Disabling the intrinsic resistome of Mycobacterium tuberculosis; Elucidating hierarchies of DNA repair and mutagenesis that undermine current antibiotic efficacy.Oral presented at HUB/IBMS/Pathology Post graduate Research Day. UCT, Cape Town, 26 November 2019.

GOBE, I., MBAU, R. D., MASON, M., MIZRAHI, V. & WARNER, D. F. Disabling the intrinsic resistome of Mycobacterium tuberculosis; Elucidating hierarchies of DNA repair and mutagenesis that undermine current antibiotic efficacy. Oral presented at Cape Town Acid Fast club. UCT, Cape Town, 3rd July 2018.

Table of Contents

Declaration.....IV
Abstract.....V
Acknowledgements.....VI
Presentations and publications.....VII
List of figures.....XI
List of tables.....XIII
Frequently used abbreviations.....XIV
Chapter 1: General Introduction.....1
1.1 Pathogenesis of Tuberculosis.....**2**
1.2 Epidemiology of Tuberculosis.....**4**
1.3 TB treatment.....**5**
1.4 Mechanisms of drug resistance in MTB.....**7**
1.4.1. The granuloma niche.....9
1.4.2. The cell wall barrier.....10
1.4.3. Efflux and influx mechanisms.....10
1.4.4. Biotransformation reactions and drug inactivation.....12
1.4.5. Activation of a transcriptional regulator.....13
1.5. Acquired (genetic) drug resistance in Mycobacteria.....**13**
1.5.2. Antibiotic tolerance and persistence.....14
1.5.3. Mutator genotypes.....17
1.5.4 MTB drug resistance and antibiotics.....18
1.6. The mycobacterial DNA damage response.....**19**
1.6.1. The SOS response.....19
1.6.2. RecA-independent response.....20
1.7. DNA damage repair and tolerance mechanisms.....**22**
1.7.1 DSB repair systems in MTB.....22
1.7.2. Base excision repair (BER).....24
1.7.3. Non- canonical mismatch repair.....25
1.7.4. Mycobacteria NER pathway: UvrB enzyme.....25
1.7.5. DNA damage tolerance: TLS and DnaE2.....29
Balance in Mycobacterial error Free and error prone repair.....**33**
1.8. Introduction to transposons.....**33**
1.8.1. Transposon Insertion Sequencing (TnSeq).....34
1.8.2. Tn mutagenesis and gene interaction studies.....36
1.9. Using TnSeq to elucidate DNA repair hierarchies in mycobacteria.....**37**
Chapter 2: Materials and methods.....39

2.1 General materials and methods **40**
2.1.1 Bacterial strains, media and culture conditions 40
2.1.2 Bacterial transformations 41
2.1.3 DNA manipulations 42
2.1.4 Phenotypic characterization 46
2.1.5 Generation of transposon libraries in *MSM* 48
2.1.6 Data processing: statistical analysis of essentiality 52
2.1.7 TnSeq conditional essential genes validation using CRISPR interference platform 56
Chapter 3: 60
Predicting gene essentiality in *Mycobacterium smegmatis*: Using TnSeq to derive essential gene lists for wild-type, Δ*dnaE2* and Δ*uvrB* strains 60
3.1 Introduction **61**
3.2 Aims and objectives. **63**
3.3 Results **64**
3.3.1 TnSeq library generation 64
3.3.2 TRANSIT pre-processing (TPP) results 65
3.3.3 Genome Essentiality Analysis 66
3.3.4 Comparison of essentiality data from different laboratories 67
3.3.5 Comparing TnSeq essentiality data with CRISPRi data set 68
3.3.6 Functional classification of the 'core' MSM essential gene set 71
3.3.7 Essentiality of genes in the duplicated region of the MSM chromosome ... 74
3.3.8 Comparison of MSM essential gene list with published essential genes of MTB 76
3.3.9 Validation of MSM essential genes using CRISPR- Cas9 technology 78
3.3.10 Comparison of essential genes between WT MSM, *uvrB* and *dnaE2* mutant strains 79
3.3.11 Confirmation of the genotypes of *dnaE2* and *uvrB* deletion mutants 80
3.3.11 Whole genome sequencing (WGS) 83
3.3.12 Tn library generation in Δ*dnaE2* and Δ*uvrB* strains of MSM 86
3.4 Discussion **90**
Chapter 4: 96
Investigating conditional gene essentialities in DNA repair-deficient *Mycobacterium smegmatis* under genotoxic stress 96
4.1. Introduction **97**
4.2. Aims and objectives **98**
4.3.1. Phenotypic characterization of *dnaE2* and *uvrB* deletion mutants 99
4.3.2. Phenotypic characterization: MMC DNA damage tolerance assay 100

4.3.3. Tn-Seq library selection 102
4.3.4. Analysis of conditionally essential genes during MMC treatment 103
4.3.5. CRISPRi validation of MMC conditionally essential genes 116
4.3.6. Genetic interaction analysis (GI): MMC 119
4.3.7. Analysis of genes that are conditionally essential during MOX treatment 129
4.3.8. Genetic interactions under MOX treatment 139
4.3.9. Comparison of aggravating GI genes in Δ*uvrB* and Δ*dnaE2* for both MMC and MOX treatments 145
4.3.10. Genes which are conditionally essential in MMC and MOX treatment independent of bacterial genotype 148
4.4 Discussion **150**
Chapter 5 Concluding remarks 162
Supplementary information online 168
References 169

List of figures

Figure 1. 1 Estimated incidence of MDR-TB worldwide. 5

Figure 2. 1 Determination of minimum inhibitory concentration. 47
Figure 2. 2 A schematic diagram showing steps involved in TnSeq library generation 51
Figure 2. 3 TRR data normalization. 53
Figure 2. 4 QQ (quantile-quantile)-plot of non-zero read counts for dataset.. 54
Figure 2. 5 Illustration of resampling permutation method.. 55

Figure 3. 1 Essentiality analysis of MSM WT genes using the HMM model.. 67
Figure 3. 2 Comparison of Gobe (this study) and Dragset et al WT TnSeq data.. 68
Figure 3. 3 Comparison of TnSeq and CRISPRi predictions of essential genes in WT MSM. 69
Figure 3. 4 Functional classification of core MSM essential genes (both ES and GD). 72
Figure 3. 5 Classification of essential genes that belong to the *Intermediary Metabolism and Respiration* group 73
Figure 3. 6 Classification of essential genes belonging to the *Information Pathways* group.. 73
Figure 3. 7 A Venn diagram of MSM mc^2155 and MTB H37rv ES. 625 MTB essential genes. 77
Figure 3. 8 CRISPRi validation of TnSeq essential genes. 79
Figure 3. 9 Confirmation of Δ*dnaE2* deletion mutant by PCR 81
Figure 3. 10 Confirmation of the genotype of the Δ*uvrB* deletion mutant by PCR.. 82
Figure 3. 11 Read coverage along the genomic coordinates of *dnaE2* in MSM.. 84
Figure 3. 12 A Read coverage along the genomic coordinates of *uvrB* in MSM.. 85
Figure 3. 13 Genes more essential in a Δ*dnaE2* background when compared to the WT..... 87
Figure 3. 14 Read coverage along the genomic coordinates of the duplicated region. 89

Figure 4. 1 Determination of MIC_{90} values for MMC and MOX..... 99
Figure 4. 2 MMC damage survival assay. 101
Figure 4. 3 MOX damage survival assay.. 102
Figure 4. 4 A comparison of gene sets predicted to be conditionally essential during MMC treatment in WT, Δ*dnaE2* and Δ*uvrB* libraries.. 104
Figure 4. 5 Tn insertions counts in *dnaE2* and *uvrB* genes in the WT_MMC Tn libraries.. 105
Figure 4. 6 Differential essentialities of *uvrB* and *dnaE2* in the different mutant libraries... 106
Figure 4. 7 Functional classification of "MMC essential" genes in WT, Δ*uvrB* and Δ*dnaE2* Tn libraries.. 113
Figure 4. 8 Validation of conditional essentiality of MSMEG_4259 (*dnaQ-uvrC*) under MMC treatment by CRISPRi..... 117
Figure 4. 9 Validation of the essentiality of *pstA* during MMC by CRISPRi.. 118
Figure 4. 10 Tn insertions in some of the genes which have aggravating genetic interactions with *uvrB* during MMC treatment.. 121
Figure 4. 11 Tn insertions in selected genes which have aggravating genetic interactions with *dnaE2* during MMC treatment..... 124
Figure 4. 12 Insertions in MSMEG_2868 in WT-MMC, Δ*uvrB*-MMC and Δ*dnaE2*-MMC libraries.. 125
Figure 4. 13 Validation of MSMEG_2868 conditional essentiality using CRISPRi..... 126

Figure 4. 14 CRISPRi confirmation of conditional essentiality of MSMEG_2868 in MOX, GRY and NOV.....128
Figure 4. 15 Venn diagram comparing gene sets predicted to be conditionally essential in WT, Δ*dnaE2* and Δ*uvrB* libraries under MOX treatment.....134
Figure 4. 16 Distribution of Tns insertions in *uvrB* (upper panel) and *dnaE2* (bottom panel) before after treatment of WT Tn library with MOX......135
Figure 4. 17 Distribution of transposon insertions in *uvrB* before and after treatment of the Δ*dnaE2* Tn library with MOX.. 136
Figure 4. 18 Distribution of transposon insertions in *dnaE2* before and after treatment of Δ*uvrB* Tn library with MOX......136
Figure 4. 19 Functional classification of conditionally essential genes identified in WT, Δ*uvrB* and Δ*dnaE2* libraries under MOX treatment.....137
Figure 4. 20 Validation of the essentiality of *pimE* during MOX treatment by CRISPRi. ...139
Figure 4. 21 Distribution of transposon insertions in selected genes with aggravating interactions with *dnaE2* during MOX treatment.. 141
Figure 4. 22 Functional classification 163 genes that were identified as having aggravating genetic interactions with *uvrB* during MOX treatment.....143
Figure 4. 23 Distribution of Tn insertions in selected genes with aggravating genetic interactions with *uvrB* during MOX treatment.. 145
Figure 4. 24 Venn diagram comparing gene sets predicted to have aggravating genetic interactions (GI) with *dnaE2* and *uvrB* during MMC treatment.....146
Figure 4. 25 Venn diagram comparing gene sets predicted to have aggravating genetic interactions with *dnaE2* and *uvrB* during MOX treatment.....146
Figure 4. 26 Distribution of Tn insertions in MSMEG_6235.....147
Figure 4. 27 A Venn diagram comparing gene sets predicted to be conditionally essential in MMC and MOX treatments independent of the strain used.....148
Figure 4. 28 Distribution of Tn insertions in MSMEG_4269 (*asnB*) under MOX (upper panel) and MMC (lower panel) treatments in WT, Δ*dnaE2* and Δ*uvrB* strains.....149

List of tables

Table 2. 1 General bacterial strains used in this study. 40
Table 2. 2 CRISPRI sgRNAs used in this study 58

Table 3. 1 Tn library generation 64
Table 3. 2 A summary of TPP results for WT Tn libraries 66
Table 3. 3 Essential genes that are unique to the "Gobe" WT TnSeq data. 69
Table 3. 4 Essential genes unique to Dragset et al WT TnSeq data. 70
Table 3. 5 Essential genes that are unique to CRISPRi data. 70
Table 3. 6 Essentiality calls for selected genes (only those with gene names) within the MSM mc^2155 duplicated region. 74
Table 3. 7 A summary of TPP data analysis for ΔDRKIN TnSeq libraries 75
Table 3. 8 Comparison between WT and ΔDRKIN untreated libraries by resampling method. 76
Table 3. 9 Genes which are unique to MSM during comparison with MTB essential gene set. 78
Table 3. 10 A summary of TPP sequence analysis of Δ*dnaE2* and Δ*uvrB* Tn libraries 86

Table 4. 1 A summary of TPP results for MMC-treated Tn libraries. 103
Table 4. 2 Conditionally essential genes in WT_MMC. 107
Table 4. 3 Conditionally essential genes in Δ*dnaE2*_MMC. 109
Table 4. 4 Conditionally essential genes in Δ*uvrB*_MMC. 111
Table 4. 5 Genes that belong to the information pathway functional group in MMC-treated libraries. 114
Table 4. 6 TnSeq WT _MMC conditionally essential genes that were also identified as upregulated in RNA seq MMC data by Muller et al., 2018. 115
Table 4. 7 Genes exhibiting aggravating genetic interactions with *uvrB* during MMC treatment. 120
Table 4. 8 Genes exhibiting aggravating genetic interactions with *dnaE2* during MMC treatment. 122
Table 4. 9 Conditionally essential genes in WT_MOX. 129
Table 4. 10 Conditionally essential genes in *dnaE2*_MOX 130
Table 4. 11 Conditionally essential genes in *uvrB*-MOX. 132
Table 4. 12 Genes that have aggravating genetic interactions with *dnaE2* during MOX treatment. 140
Table 4. 13 Genes that have aggravating genetic interactions with *uvrB* during MOX treatment. 142

Frequently used abbreviations

ABC	ATP- binding cassette
ATc	anhydrotetracycline
BER	Base excision repair
CAF	Central Analytical Sequencing facility
CFU	Colony forming units
CRISPR	Clustered regularly interspaced short palindromic repeats
DBS	Double stranded DNA breaks
EMB	Ethambutol
ES	Essential
GA	Growth advantage
GD	Growth defect
GGR	Global genomic repair
GI	Genetic interactions
HITS	High throughput insertion tracking by deed sequencing
HMM	Hidden Markov Model
HR	Homologous recombination
INH	Isoniazid
iNOS	Inducible nitric oxide synthase
INSeq	Insertion sequencing
KAN	kanamycin
MDR-TB	Multidrug Resistance TB
MFS	Major facilitator superfamily
MIC	Minimum Inhibitory Concentration
MMC	Mitomycin c
MMR	Mismatch repair
MOX	Moxifloxacin
MSM	Mycobacterium Smegmatics
MTB	Mycobacterium tuberculosis
NE	Non essential
NER	Nucleotide excision repair
NERiM	NER induced mutagenesis
NHEJ	Non homologous end joining
NZ mean	Non zero sites mean
OG	8-oxo-G
PafBC	Proteasome accessory factors B and C
PAM	Protospacer adjacent motif
PCR	Polymerase chain reaction
PE	Paired end
phoX2	Phagocyte oxidase
PMF	Proton motive force
PPS	Pup- proteasome system
PZA	Pyrazinamide
RecA-NDP	RecA independent promoter
RIF	Rifampicin
RND	Resistance- nodulation cell division
RNI	Reactive nitrogen intermediates
ROS	Reactive oxygen species

SMR	Small multidrug resistance
SNPs	Single nucleotide polymorphisms
SSA	Single strand annealing
ssDNA	Single stranded DNA
STM	Signature – tagged mutagenesis
TB	Tuberculosis
TCR	Transcription coupled repair
TLS	Trans-lesion synthesis
Tns	Transposons
TnSeq	Transposon Insertion Sequencing
TPP	TRANSIT pre- processor
TraDIS	Transposon-directed insertion site sequences
TRR	Trimmed total reads
WHO	World Health Organisation
WT	Wild type
XDR-TB	Extensively drug resistance TB

Chapter 1: General Introduction

1.1 Pathogenesis of Tuberculosis

Mycobacterium tuberculosis (MTB), the causative agent of tuberculosis (TB), was identified by Robert Koch in 1882 (Cambau and Drancourt, 2014). MTB is transmitted via the respiratory route when a naïve individual is exposed to infectious aerosols generated by a MTB-infected individual (Sgaragli and Frosini, 2016). Upon inhalation, MTB interacts with respiratory tract cells including macrophages and dendritic cells (Sgaragli and Frosini, 2016). Generally, the role of macrophages in the body is to remove invading organisms, foreign particles and cellular debris (Guirado et al., 2013). Alveolar macrophages are the major target cells for MTB and hence play a major role in immune defence against MTB (Guirado et al., 2013). Clearance of MTB by macrophages is not always efficient because MTB has developed ways of evading macrophage killing: MTB can, among other things, interfere with phago-lysosomal maturation, trafficking and acidification, hampering macrophage effector functions and so preventing the clearance of the tubercle bacilli (Khan et al., 2019). MTB-infected macrophages secrete cytokines and chemokines that facilitate recruitment of more inflammatory cells to the site of infection resulting in the formation of granulomas - hallmarks of MTB infection (Korb et al., 2016).

A classic granuloma has a necrotic centre surrounded by different cell types including layers of macrophages and lymphocytes (Cadena et al., 2017). Granulomas are heterogenous: They differ by size, morphology, shape, and number of bacteria. Morphologically, granulomas can be caseous (with a necrotic centre), dense cellular(having no distinct core), suppurative(with vast neutrophil infiltration) and fibrotic (surrounded or infiltrated by collagen) (Warsinske et al., 2017). The evolutionary role of granulomas in humans and other vertebrates is to protect against invading foreign particles. However, initial failure of granuloma to contain an inciting agent can lead to a deleterious granulomatous response (Pagán and Ramakrishnan, 2018). During MTB infection, granulomas are believed to contain MTB at the site of infection and thereby contribute to the control of the infection. Nonetheless, they are also implicated in early dissemination and proliferation of disease (Davis and Ramakrishnan, 2009).

The RD1 locus of mycobacteria encodes an ESX-1 secretion system which facilitates secretion of mycobacterial proteins, ESAT-6 and CFP10. These proteins are implicated in the virulence of mycobacteria by inducing host membrane lysis and cell-to cell spread (Guinn et al., 2004, Houben et al., 2012). It has also been shown that Intracellular mycobacteria use the ESX-

1/RD1 virulence locus to recruit new macrophages to newly formed granulomas(Davis and Ramakrishnan, 2009): Deletion of the RD1 locus resulted in attenuated granuloma formation indicating that chemotactic movement of cells during granuloma formation is dependent on the RD1 locus (Davis and Ramakrishnan, 2009). During granuloma formation, new recruited and uninfected macrophages find and phagocytose infected macrophages undergoing apoptosis, leading to rapid, iterative expansion of infected macrophages and consequently bacterial numbers. The primary granuloma then seeds secondary granulomas via migration of infected macrophages (Davis and Ramakrishnan, 2009).

Furthermore, the bacilli have evolved to persist inside granulomas for extended periods of time, using them as a hiding place until reactivation occurs (Lin and Flynn, 2018). MTB can survive stresses including hypoxia, nutrient deprivation, and oxidative and nitrosative stress, that it encounters within the granuloma. Surviving these environments requires complex transcriptional networks including the DosR regulon (Martin et al., 2016). Furthermore, MTB produces enzymes (e.g., superoxide dismutase) which ameliorate oxidative damage. MTB is also capable of scavenging for key nutrients inside the nutrient limiting granulomas and remodelling carbon source utilisation (Martin et al., 2016)

Long-term control of MTB infection depends on the immunological state of the host and the ability of the bacteria to counter the host defence (Mascart and Locht, 2015). The interaction between activated macrophages and IFN-secreting $CD4^+$ T cells is critical in controlling MTB infection. Host immune suppression, as in the case of low $CD4^+$ T cell counts in HIV, is frequently associated with disease reactivation (Gupta et al., 2012). For a long time, infection with MTB has been understood in the context of a binary model as either active disease or a latent infection. Latent MTB infection is characterized by the presence of immunological sensitivity to mycobacterial antigen (as determined by a tuberculin skin test or an interferon (IFN)-γ release assay) in the absence of the clinical symptoms of disease (Cadena et al., 2017). Even though this phenomenon was historically falsely associated with dormancy, MTB is likely to be actively replicating since so-called latent infections respond to isoniazid treatment - a mycobacterial cell wall inhibitor (Lipworth et al., 2016). Active disease on the other hand is characterized by clinical symptoms which include chronic cough, weight loss, fever and hemoptysis (Cadena et al., 2017). This binary model of infection has recently been rejected (Drain et al., 2018, Lin and Flynn, 2018). Recent reviews propose a broad classification of MTB infection which take into account the complex spectrum of MTB infection outcomes:

some people are constantly exposed but are not getting infected (resisters), some get infected and are able to completely clear the infection, some are infected but asymptomatic, some are latently infected and are likely to reactivate while some have active MTB infection which can either be chronic or severe (Lin and Flynn, 2018). In fact, another review proposed two more categories of MTB infection, incipient TB and subclinical TB, which falls between latent infection and active MTB disease (Drain et al., 2018). An incipient MTB infection is likely to progress into active TB in the absence of intervention (Drain et al., 2018). On the contrary, a subclinical infection has detectable sputum bacilli however, the patient does not show any clinical TB related symptoms (Lin and Flynn, 2018).

Definitive confirmation of MTB infection requires identification of tubercle bacilli from sputum (for pulmonary TB) or other samples (extrapulmonary TB) by microbiological culture, acid fast staining or nucleic acid testing (Cadena et al., 2017).

1.2 Epidemiology of Tuberculosis

The World health organisation (WHO) has reported that TB is the leading cause of death from a single infectious agent, surpassing HIV/AIDS (WHO, 2019). The burden of TB is especially heavy in low income countries (WHO, 2019). Alongside malaria and HIV/AIDS, TB has been classified as a "disease of poverty" (Stevens, 2017). These diseases are preventable and treatable with existing prophylactics, interventions and medications but, because of factors associated with poverty such as indoor crowding and pollution, poor nutrition, lack of access to proper sanitation, and lack of health education, they are endemic in low income settings (Stevens, 2017).

In 2018, an estimated 10.0 million people developed TB disease (WHO, 2019). Even though TB affects all genders and ages, the highest burden was found to be in men who are above 15 yrs. of age. This group accounted for 57% of all TB cases in 2018 (WHO, 2019). Close to two-thirds of people who developed TB were in the high TB burden countries: India (27%), China (9%), Indonesia (8%), the Philippines (6%), Pakistan (6%), Nigeria (4%), Bangladesh (4%) and South Africa (3%). The WHO European region and the WHO region of the Americas accounted for only 6 % of global cases (WHO, 2019). In the same year, deaths from TB were estimated to be 1.2 million among HIV-negative people and 251000 among HIV-positive people (WHO, 2019). The greatest threat to TB prevention and eradication efforts is the

development of multi-drug resistant (MDR)-TB (WHO, 2019). MDR-TB is caused by strains that are resistant to isoniazid (INH) and rifampicin (RIF), the two most potent first-line drugs for TB treatment (WHO, 2018).The success of MDR-TB treatment so far stands at about 50% (Quenard et al., 2017). Hence MDR-TB must be addressed as a crisis given its potential for hindering the achievement of current objectives. There is also extensively drug-resistant (XDR)-TB which is caused by strains that are resistant to INH and RIF plus any fluoroquinolone and at least one of the second-line injectable aminoglycosides (Quenard et al., 2017). The prevalence of drug resistant TB is increasing: in 2018, half a million new cases of RIF resistant TB were reported and 78% of these cases were MDR-TB (WHO, 2019). Almost half of these cases were in India (27%), China (14%) and the Russian Federation (9%) (WHO, 2019) (**Figure 1.1**).

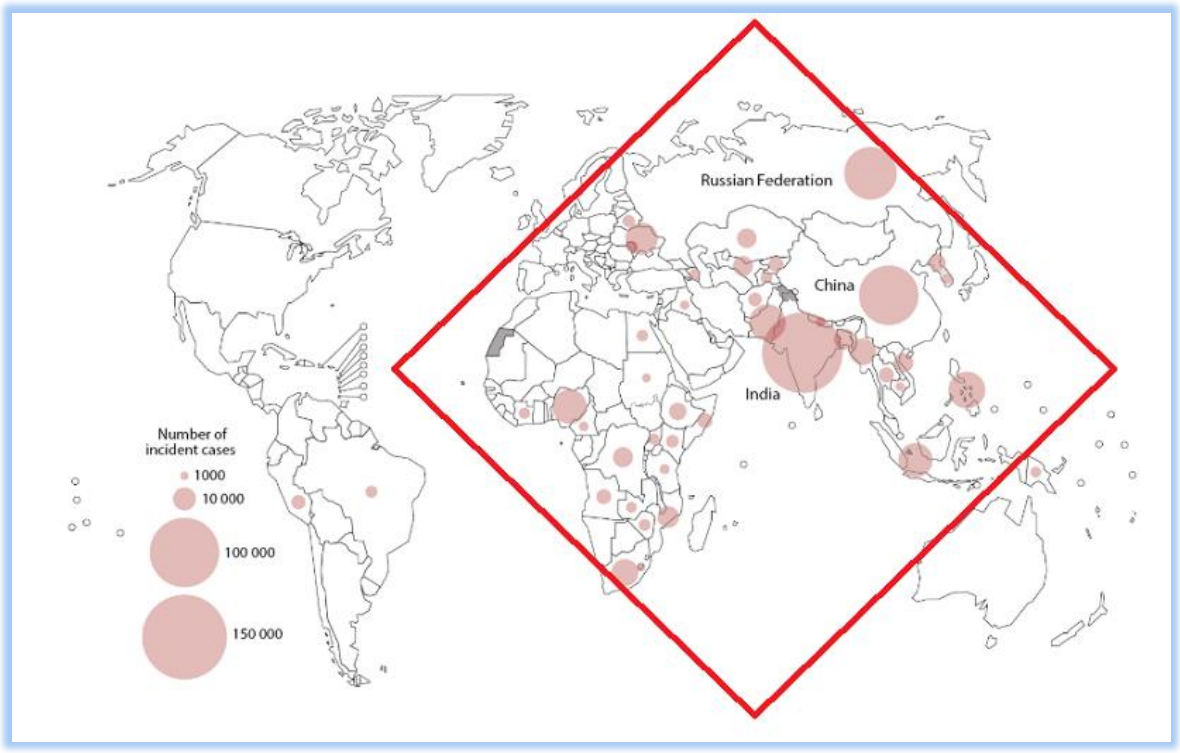


Figure 1. 1 Estimated incidence of MDR-TB worldwide: High-burden regions are highlighted in a red diamond shape. Data source: WHO Global tuberculosis report 2019.

1.3 TB treatment

Standard TB treatment takes six month (WHO, 2017). The regimen consists of 2 months of at least INH, RIF, pyrazinamide (PZA) and ethambutol (EMB) which is followed by four months of treatment with RIF and INH (WHO, 2017). Approximately 80% of people with uncomplicated pulmonary TB have no culturable bacilli in their sputum after 2 months of

standard treatment (Su et al., 2011).Nonetheless, it is essential to go through the four months additional treatment period in order to avoid relapse (Horsburgh Jr et al., 2015). Drug resistance can occur during treatment and cause relapse. In order to reduce the proportion of relapses, combination therapy with at least three drugs with different mechanisms of action is used (Horsburgh Jr et al., 2015). Compared to duration of treatment for other bacterial infections, the standard 6 months treatment regimen for TB is exceptionally long. This results in issues with adherence and toxicity (Horsburgh Jr et al., 2015). It is estimated that 16 to 49 percent of TB patients do not complete treatment (Volmink and Garner, 2007). Reasons for non-adherence include adverse side effects, stigma, cost and belief that one has been cured in the absence of symptoms (Horsburgh Jr et al., 2015). Non-adherence to treatment is one of the factors that contributes to slow progression in TB control cases. However, a recent survey has also shown that patients are lost along the pathway of care as 40% of TB patients never get diagnosed and treated (Kaplan, 2020). Socioeconomic reasons are implicated as reasons for this failure(Kaplan, 2020).Furthermore, gaps in our understanding of TB pathogenesis, host pathogen interactions and protective immunity are also factors involved in hindering TB control strategies (Kaplan, 2020).

The major barrier to eradicating TB is the emergence of drug-resistant strains of MTB (WHO, 2019, Falzon et al., 2017). Even though a large proportion of TB patients can have a relapse-free TB cure within 6 months, patients with MDR-TB require a much longer treatment duration (Falzon et al., 2017). MDR-TB treatment is also more toxic, costly, less effective and less accessible as compared to standard TB treatment (Falzon et al., 2017). A typical MDR-TB treatment regimen takes 18-20 months (WHO, 2020). This presents a formidable challenge to healthcare service providers to ensure patient adherence and cure rates. WHO recommends that MDR-TB should be treated with bedaquiline, linezolid and at least a fluroquinolone (levofloxacin or moxifloxacin (MOX))(WHO, 2019, WHO, 2020). This treatment is completed with other TB drugs which are likely to be effective in the first six months. The standardized, shorter all oral bedaquiline containing MDR-TB regimen (9-12 months) may also be offered to eligible patients (WHO, 2020).

1.4 Mechanisms of drug resistance in MTB

Drug resistance in mycobacteria is generally understood to be caused by a synergy between intrinsic and acquired (genetic) resistance (Pasipanodya and Gumbo, 2011). Intrinsic drug resistance is a natural phenomenon present in all bacterial species which predates antibiotic chemotherapy. Permeability barriers play a crucial role in intrinsic resistance (Cox and Wright, 2013). The classic example of intrinsic resistance is found in gram negative bacteria which are resistant to most clinically effective drugs used to treat gram positive bacteria (Cox and Wright, 2013). The basis for this resistance is the presence of an outer membrane which is impermeable to many molecules (Cox and Wright, 2013). In mycobacteria, the cell wall plays a crucial role as a barrier which results in limited penetration of drugs (Nguyen and Thompson, 2006). Furthermore, mycobacterial species possess various antibiotic neutralising mechanisms which work in synergy to the cell wall to bring about intrinsic resistance (Nguyen and Thompson, 2006, Awasthi and Freundlich, 2017, Nasiri et al., 2017). These mechanisms of drug resistance in MTB are summarised in **Figure 1.2** below.

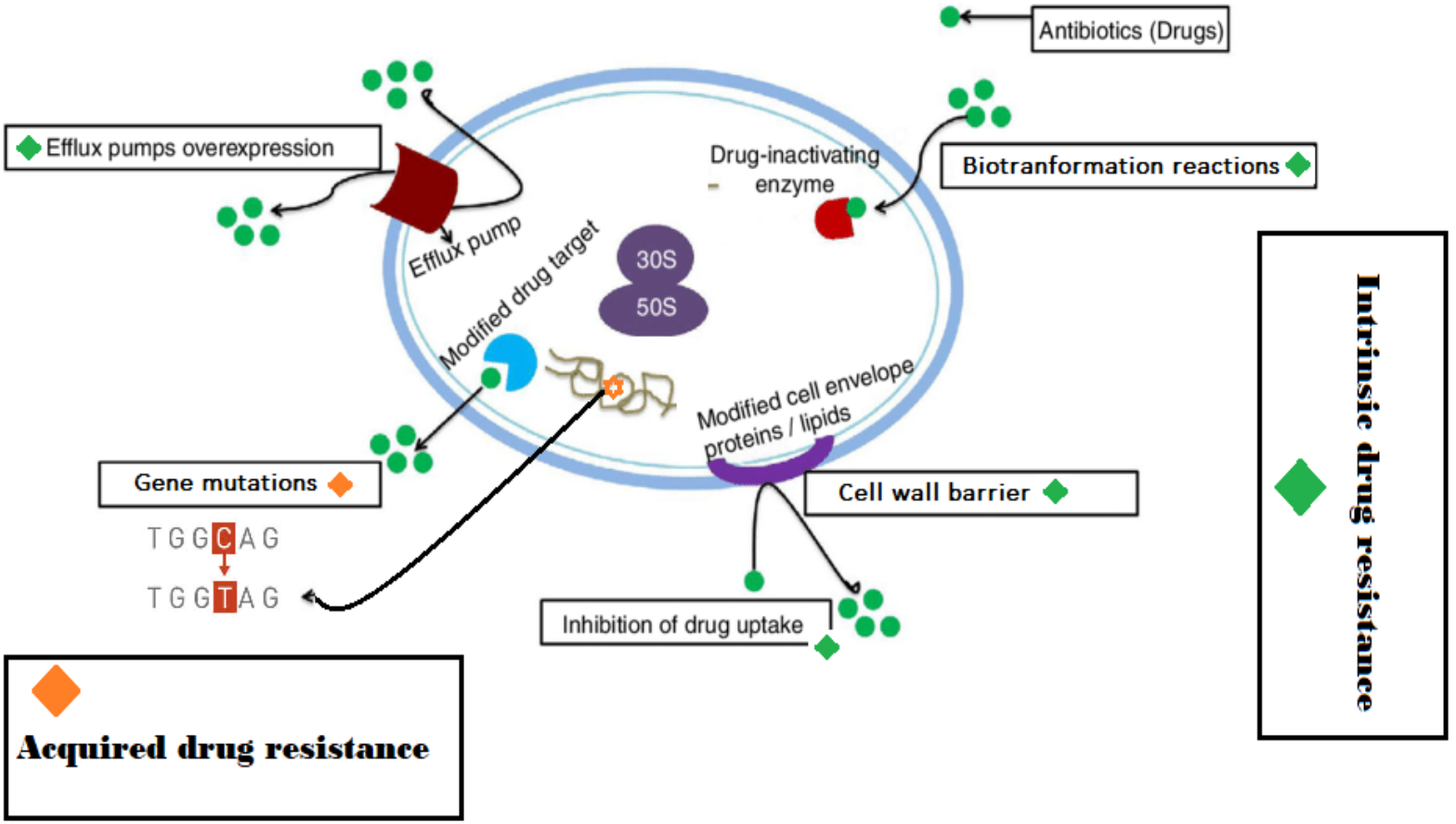


Figure 1.2: A schematic diagram showing mechanisms of drug resistance in MTB. Adapted from (Singh et al., 2020). Intrinsic drug resistance mechanisms are indicated in green diamonds and acquired drug resistance mechanisms are indicated in orange diamonds.

1.4.1. The granuloma niche

Upon MTB infection, a granuloma forms (Pagán and Ramakrishnan, 2018). The ultimate purpose of a granuloma is to contain and stop the infection from spreading (Cadena et al., 2017, Ramakrishnan, 2012). In order to reach their intended target, anti-TB drugs have to be transported through blood vessels to MTB lesions/granulomas (Dartois, 2014). However, granulomas have an impaired blood flow which makes it difficult for drugs and oxygen to be delivered to lesions (Datta et al., 2015). It has been shown that the resulting hypoxia is ultimately associated with reduction in drug susceptibility (Rustad et al., 2008). Furthermore, different studies have highlighted evidence of the existence of a heterogeneous population of MTB lesions (Dartois, 2014, Prideaux et al., 2015, Lenaerts et al., 2015). These can be cellular or non-cellular, calcified, fibrotic, necrotic or open cavities (Lenaerts et al., 2015). The interplay of factors that result in a heterogeneous architecture of lesions is not known. However, these differences affect micro-environments (including hypoxia, oxidative stress, pH, nutrient starvation) within lesions that impact the replication and metabolic properties of bacteria within them (Lenaerts et al., 2015). Lesion-specific properties can also affect drug penetration and drug access (Lenaerts et al., 2015, Dartois, 2014). For example, it has been shown that different drugs exhibit different patterns of distribution across TB lesion compartments (Prideaux et al., 2015). These authors demonstrated that INH and PZA are homogenously distributed across different lesions and that RIF only accumulated on caseum after multiple doses (Prideaux et al., 2015). Even though the distribution of MOX unlike RIF (and like INH and PZA) was dose-independent, its distribution into caseum was variable. MOX did not diffuse well into acellular regions (Prideaux et al., 2015). These patterns of drug distribution across lesions may create windows of monotherapy which lead to multiplication of mono-resistant mutants and subsequently, to the development of MDR (Prideaux et al., 2015). Subinhibitory drug concentrations are generally understood to facilitate the evolution of drug resistance, therefore heterogeneity within and between patients means that some TB cases are being unintentionally exposed to

subinhibitory drug concentrations (Andersson and Hughes, 2014). Together these findings highlight the importance of the granuloma niche in MTB resistance.

1.4.2. The cell wall barrier

The cell wall is thought to provide a non-selective barrier to anti-TB drugs owing to the impermeability of the mycobacterial outer membrane (Warner and Mizrahi, 2006).The mycobacterial cell envelope is made up of a network of peptidoglycan, the arabinogalactan polysaccharide and mycolic acids (Alderwick et al., 2015). Together, these components create a barrier that shields the organism from environmental stress and contributes to resistance of mycobacteria to many antibiotics (Alderwick et al., 2015, Abrahams and Besra, 2018). Arabinogalactan and peptidoglycan layers are thought to limit the entry of hydrophobic molecules, whereas the mycolic acid-containing outer membrane limits entry of both hydrophobic and hydrophilic molecules (Viveiros et al., 2012). Drug resistance owing to cell wall impermeability has been supported by research that shows that alterations in proteins that are involved in cell wall integrity result in increased susceptibility to multiple drugs (Nasiri et al., 2017). Biosynthesis of the peptidoglycan layer is facilitated in part by formation of UDP-MurNAc, which is synthesized by MurA and MurB enzymes (Castañeda-García et al., 2013). MTB is intrinsically resistant to fosfomycin, a cell wall synthesis inhibitor that targets MurA. However, a mutation of the wild-type aspartate residue in the MurA active site to a cysteine renders the bacteria sensitive to fosfomycin (Castañeda-García et al., 2013). Furthermore, deletion of ldt_{Mt1} and ldt_{Mt2}, which are MTB L,D-transpeptidases responsible for non-classical 3→3 linkages between opposing stem peptides in peptidoglycan, have been found to make the bacteria more susceptible to beta lactam antibiotics (Schoonmaker et al., 2014). The cell wall does not work solitarily. Mycobacterial species poses various antibiotic neutralising and efflux mechanisms which work in synergy to the cell wall to bring about intrinsic resistance (Nguyen and Thompson, 2006, Awasthi and Freundlich, 2017, Nasiri et al., 2017). These are discussed below.

1.4.3. Efflux and influx mechanisms

Intrinsic resistance owing to the impermeability of the cell wall usually works in synergy with active efflux of drugs (Viveiros et al., 2012). Efflux pumps are transporters, mediating the increased efflux of antibiotics that contributes to drug tolerance and resistance. Drug resistance

due to efflux pumps can occur in several ways: intrinsic resistance results from low-level, constitutive expression of efflux pumps; phenotypic resistance results from increased expression of efflux pumps under antibiotic stress; acquired resistance may be a result of horizontal gene transfer or mutations that allow increased expression of efflux pumps. (Hernando-Amado et al., 2016). Efflux of antibiotics can generate subinhibitory concentrations of drugs which may allow bacteria to survive long enough for chromosomal mutations conferring high-level resistance to occur (Viveiros et al., 2012). There are 5 super families of efflux pumps in bacteria: the ATP-binding cassette (ABC) superfamily, the major facilitator superfamily (MFS), the small multidrug resistance (SMR) superfamily, the resistance–nodulation-cell division (RND) superfamily and the multidrug and toxic compound extrusion (MATE) superfamily (Pule et al., 2015). Whereas the ABC efflux family uses ATP to extrude drugs, all the other efflux families use the transmembrane proton motive force (PMF) (Pule et al., 2015). The role of efflux proteins in maintaining intrinsic multi-drug resistance of MTB is supported by the fact that mycobacterial species contain a large number of known drug efflux systems (Nash, 2016). In clinical isolates of MTB, antibiotic stress seems to trigger expression of efflux pumps: efflux pumps *jefA, drrA, drrB, efpA, mmr* and Rv1217-Rv1218 were overexpressed under INH and RIF stress (Li et al., 2015). This implies that efflux pumps might play a role in the emergence MDR strains (Li et al., 2015). This notion is further supported by a study that showed that some clinical MTB strains that are resistant to RIF and INH have no mutations in genes associated with INH and RIF showing that resistance can occur through mechanisms which are not mutation based (Yamchi et al., 2015).

Furthermore, it has emerged that there is a relationship between antibiotic susceptibility induced by drug target mutations and efflux pumps (Fange et al., 2009): when efflux pumps are inhibited, high affinity reducing targets mutations do not have an effect on drug susceptibility(Fange et al., 2009). This phenomenon termed ‘target resistance masking’ seems to occur mainly on pathogens with minimal intracellular drug degradation, low cell-membrane drug permeability and those with high affinity drug binding sites (Fange et al., 2009). Inhibition of efflux pumps is also implicated in slowing down drug resistance evolution hence suggesting that combination therapy which include efflux pumps inhibition together with drug target therapy is an effective way of treating bacterial infections while minimising drug resistance evolution (Fange et al., 2009). Efflux inhibitors also have a potential role in shortening treatment duration (Viveiros et al., 2012)

1.4.4. Biotransformation reactions and drug inactivation

MTB has the ability to perform biotransformation reactions on antibiotics (Awasthi and Freundlich, 2017). These reactions can either be beneficial or detrimental to the bacterium. For example, most first line antitubercular drugs including INH, PZA and ETH, are activated within the bacterium (Awasthi and Freundlich, 2017). INH is activated with in the bacterium by a catalase peroxidase encoded by katG (Awasthi and Freundlich, 2017).Deactivation of antibacterial agents through biotransformation is also well documented (Awasthi and Freundlich, 2017). Enzymatic strategies of antibiotic inactivation include hydrolysis, N-alkylation, amidation and nitro reduction (Awasthi and Freundlich, 2017, Wright, 2005). Hydrolysis affects antibiotics that have hydrolytic chemical bonds; *e.g.*, amidases target the β-lactam ring of penicillin and esterases hydrolyses ester bonds in macrolides antibiotics (Wright, 2005). Group transferases (*e.g.*, acyltransferases, phosphotransferases, ADP-ribosyl transferases) are enzymes that modify antibiotics such that binding to target is impaired. These enzymes are only active in the cytosol because they usually require co-substrates such as ATP, UDP glucose, glutathione, acyl CoA, for activity (Wright, 2005).

Most potential antitubercular drug are inactivated by MTB (Awasthi and Freunlich, 2017): pyridobenzimidazole compound 14, a cidal antitubercular agent, is inactivated by Rv0560c through the biotransformation mechanism of N-methylation, sulphanilamide is inactivated through amidation inside MTB, aminoglycosides are generally deactivated by MTB through amidation, esterification, phosphorylation (Awasthi and Freundlich, 2017, Nasiri et al., 2017). Furthermore, the non-pathogenic *Mycobacterium smegmatis* (MSM) is intrinsically resistant to RIF because it encodes a RIF ADP-ribosyl transferase, Arr, which ADP-ribosylates RIF preventing it from binding to the Rna polymerase (RNAP) β-subunit (Baysarowich et al., 2008). RIF inhibit bacterial transcription by targeting the β-subunit of RNAP. The ADP-ribosylation of RIF at the C23 hydroxyl group by Arr, abolishes a critical hydrogen bond between RIF and RNAP and also prevents inter and intra molecular interactions between the two molecules (Harbottle et al., 2021). This inactivation of RIF leads to higher MICs in MSM compared to those reported for MTB (Baysarowich et al., 2008, Dabbs et al., 1995). Methylation of mycobacterial 23S rRNA prevents binding of macrolides (Nasiri et al., 2017). Pentapeptide repeat proteins encoded by mycobacterial *mfpA* gene have been shown to bind and protect mycobacterial DNA gyrase from being targeted by quinolones (Nasiri et al., 2017).

1.4.5. Activation of a transcriptional regulator

Intrinsic drug resistance in mycobacteria may occur due to an interactive network of regulatory proteins, effector proteins and inducers (Nasiri et al., 2017). Actinomycetes harbour WhiB family transcriptional regulators. In mycobacteria, this family of transcriptional regulators functions in fundamental cell processes such as virulence, redox homeostasis, cell division and antibiotic resistance (Morris et al., 2005). WhiB7 is a transcriptional regulator whose expression is induced by a variety of factors including antibiotic treatment, iron starvation, heat shock and entry into stationary phase (Geiman et al., 2006). WhiB7 has been shown to activate resistance mechanisms for macrolides, aminoglycosides, tetracyclines and lincosamides (Morris et al., 2005). Several other MTB control genes have also been implicated in drug tolerance, persistence and dormancy. These include *mbtB*, *dosR*, and *hspX* (Nasiri et al., 2017). DosR is a well-studied mycobacterial transcriptional regulator which is induced by hypoxia and multiple stresses. It is implicated in survival of MTB in granulomas and to antibiotic tolerance (Nasiri et al., 2017).

1.5.1 Acquired (genetic) drug resistance in Mycobacteria

In MTB, clinically relevant drug resistance is a genetically encoded ability to tolerate a specific drug (Gandhi et al., 2010). Most bacteria listed as antimicrobial resistance threats acquire resistance traits through exchange of genetic material via horizontal gene transfer (Eldholm and Balloux, 2016). MTB is an exception: in MTB, chromosomal mutations, mainly single-nucleotide polymorphisms, play a major role in genetically encoded drug resistance. These errors arise during DNA replication and DNA repair (McGrath et al., 2014). Mutations that result in drug resistance are found in genes that encode enzymes that are drug targets. Mutations can also occur in gene products that activate pro-drugs and in regulatory regions of target genes (Müller et al., 2013). RIF resistance results from mutations in the *rpo*B gene, fluoroquinolone resistance occurs from a mutation in the *gyr*A/B genes and, for INH resistance, prodrug activation is compromised by a mutation in the *kat*G gene while mutations in the *inhA* promoter upregulate *inhA* expression, increasing the number of InhA target molecules (Nasiri et al., 2017). Nonetheless, drug resistance of some clinical MTB isolates cannot be explained by these mutations indicating that there are other unknown mutations that confer resistance to these drugs.

In their review, Fonseca *et al*., 2015 highlight low-level resistance as another strategy of developing drug resistance. Here, mutations in other genes which are not known resistance genes lead to an increase in minimum inhibitory concentration (MIC) of drug above that of the average susceptible bacterial population but below the threshold for clinically relevant resistance. These mutations could be a stepping stone to drug resistance (Fonseca et al., 2015). In fact, Hicks et al. have showed that common mutations in PrPR, a transcriptional regulator of propionate metabolism confer conditional drug tolerance to multiple mechanistically unrelated drug including RIF, INH and fluroquinolones (Hicks et al., 2018).These mutations are not captured by standard *in vitro* drug susceptibility tests which measures drug resistance only. However, the *prpR* tolerance conferring mutations have been found to be enriched in drug resistant strains of MTB (Hicks et al., 2018) indicating that that these mutations are the stepping stones towards drug resistance. In fact, studies in *E.coli* have shown that resistance is preceded by tolerance and that tolerance increases the chances for resistance mutations to spread in a population (Levin-Reisman et al., 2017).

While drug resistance describes the genetically-encoded, inherited ability of bacteria to grow at high concentrations of an antibiotic irrespective of treatment duration, antibiotic tolerance allows a population of bacteria to survive a transient exposure to antibiotic at concentrations that exceed the MIC (Brauner et al., 2016). Drug tolerance is not captured by standard susceptibility testing but it has a great potential in causing treatment failure (Hicks et al., 2018).

1.5.2. Antibiotic tolerance and persistence

There are many bacterial factors that contribute to the acquisition of drug resistance mutations. Development of persister cells and the possibility of mutator genotypes are some of the factors implicated in other bacteria (Müller et al., 2013). Persisters are "a sub-population of transiently antibiotic-tolerant bacteria" (Fisher et al., 2017). Bacterial persisters were originally described in *Staphylococcus* (Bigger, 1944) where it was observed that removal of the antibiotic stress allowed normal growth and restored sensitivity to the applied antibiotic comparable to the parental culture (Bigger, 1944). It is now recognised that in the presence of drugs, bacteria can behave in varied ways. While some bacteria can resist drugs, some can tolerate or even persist in the presence of antibiotics (Brauner et al., 2016). The term "resistance" refers to the heritable ability of bacteria to replicate in the presence of a drug as measured by an increase in minimum inhibitory concentration (MIC) (Balaban et al., 2019) whereas tolerance and persistence are

characterised by increase in survival without an increase in MIC (Balaban et al., 2019). However, while tolerance is the ability of a general bacterial population to survive antibiotics, persistence only occurs in a subpopulation of bacterial cells (Balaban et al., 2019). Antibiotic persistence and tolerance necessitate prolonged duration of TB treatment and are both associated with treatment failures and consequently, drug resistance (Brauner et al., 2016, Hicks et al., 2018). For instance, treatment of MTB with INH generates INH persisters and INH resistance (Vilchèze and Jacobs Jr, 2019).

Persister formation mechanisms are not well known, however ongoing research in *E. coli* shows that not all persisters are created the same. Some studies have observed an increase in the levels of toxin/antitoxin (TA) transcripts in persisters cells (Dörr et al., 2010). Induction of TA modules can occur during the SOS response: specifically, an SOS-regulated TA locus, *tisAB/istR*, has been linked to ciprofloxacin persistence in *E. coli* (Dörr et al., 2010). The ppGpp alarmones and signalling molecules which play a role during stress response in many bacteria (Zhang et al., 2018) were shown to facilitate biofilm persister formation in *E. coli* and were also observed to be involved in a pathway that stimulates persistence formation during carbon source transition (Amato and Brynildsen, 2014). Still in *E. coli*, enhancement of genetic resistance in ofloxacin (OFX) persisters was observed to be dependent on RecA-mediated SOS induction through the error-prone polymerase UmuDC (Barrett et al., 2019). What was particularly striking in that study was the observation that the resistant mutants were not only resistant to ofloxacin but also to other antibiotics with different mechanisms of action (Barrett et al., 2019).

Other factors that are implicated in the formation of tolerant and persistent bacteria include drug efflux pumps, starvation, antibiotics, reduced metabolism, and reduced ATP levels (Brauner et al., 2016, Balaban et al., 2019). While persistence is usually triggered by some of these factors, spontaneous persistence which occurs without triggers during balanced growth can also occur (Balaban et al., 2019). More importantly, some mutations has been shown to result in slow killing by drugs without an increase in MIC (Van den Bergh et al., 2016). *hipA7* mutations in *E. coli* have been shown to increase persistence from 0.1% in wildtype (WT) to about 20% in mutant strains (Balaban et al., 2019). A recent study in MTB has shown that mutations in the propionate metabolism transcriptional factor *prpR* confer conditional multi-drug tolerance both *in vitro* and *ex vivo* in macrophages a phenomenon which altered sensitivity to multiple drugs (Hicks et al., 2018).

Persistent bacterial cells are not genetically resistant, hence the phenotypic resistance cannot be passed to daughter cells (Brauner et al., 2016). However, persistence and tolerance increase the survival time of bacteria exposed to antibiotics, which in turn increases the chances of subsequent acquisition of drug resistance mutations (Müller et al., 2013). Therefore, eliminating drug persisters is expected to shorten TB chemotherapy (Müller et al., 2013).

1.5.2.1 Heteroresistance

Population heterogeneity generates phenotypes specialized in surviving environmental stress. Population heterogeneity can result in antibiotic persistence/tolerance and heteroresistance (Dewachter et al., 2019). Heteroresistance occurs when a subpopulation of bacterial cells are able to replicate at antibiotic concentrations that otherwise inhibit cell growth in the rest of the population (Andersson et al., 2019). Heteroresistance can occur due to infection with more than one genetically distinct populations or through the emergence of rare spontaneous resistant mutants (Andersson et al., 2019).The heterogenous population could either be in one location or different locations within the infected organ. Mutations plays a major role in the generation of heteroresistance (Andersson et al., 2019). One striking characteristic of heteroresistance is that the frequency of the resistant subpopulations is higher than that typically observed for mutations that cause antibiotic resistance in susceptible cells (Andersson et al., 2019). It then logically follows that this phenomenon can lead to undesirable clinical outcomes. In fact some studies have linked heteroresistance with increased mortality, long treatment durations, increased clinical complication and ultimately treatment failure (Park et al., 2012, Andersson et al., 2019). However, there is contradicting evidence that shows no correlation between treatment failure and heteroresistance (van Hal et al., 2011). A recent study in South Africa showed that the frequency of genetic heteroresistance in MTB was dependent on the type of antibiotic with fluoroquinolones and Bedaquiline being the most common drug resulting in heteroresistance (Nimmo et al., 2020). However this study did not find any association between bacterial genetic diversity and clinical outcomes at least for patients infected with a single strain (Nimmo et al., 2020).

1.5.3. Mutator genotypes

Hypermutators have been observed in various bacterial species (Jolivet-Gougeon et al., 2011). MTB isolates of the Beijing genotype have been repeatedly associated with MDR-TB. This led to the notion that these isolates might exhibit a hypermutator phenotype (Rad et al., 2003). However, measurement of mutation rates across MTB lineages resulted in contradictory results (Müller et al., 2013) with some studies concluding that MTB Beijing genotype does not develop resistance mutations at an elevated rate compared with other non-Beijing genotypes (Werngren and Hoffner, 2003). However, others have found that MTB strains from East Asian and Beijing sub-lineage (lineage 2) acquire in-vitro drug resistance more rapidly when compared to Euro-American Lineage (lineage 4), a finding which suggest a high mutation rate in lineage 2 (Ford et al., 2013). This in vitro mutation rate was found to correlate with in vivo mutation rates as determined through whole genome sequencing (Ford et al., 2013).

Despite development of MDR-TB and, most recently, total drug resistance (Velayati et al., 2009), studies have shown that drug resistance mutations in MTB are a rare occurrence; that is, MTB does not necessarily have an excessive mutation rate compared to other organisms *in vitro* (McGrath et al., 2014). Until recently, this observation was considered puzzling given that MTB, unlike other bacteria, does not appear to possess a mismatch repair (MMR) system which is generally conserved throughout evolution and acts to correct replication errors (Mizrahi and Andersen, 1998). Recent work has shown that another mechanism compensates for the absence of MMR in MTB: mutation avoidance through *nucS* (Castaneda-Garcia et al., 2017). The DNA repair protein NucS has no evolutionary or structural homology to the classical MMR enzymes, MutS and Mut (Castaneda-Garcia et al., 2017).When *nucS* is inactivated in MSM the phenotype mimics that of MMR-null mutants which include increased homologous recombination and increased mutation rates (Castaneda-Garcia et al., 2017). Therefore, NucS has a role in maintaining genome stability in mycobacteria (Castaneda-Garcia et al., 2017). A functional evolutionary convergence between MutS/L and NucS based systems has specifically been shown by their similar specificity on transition mutations (Castañeda-García et al., 2020). However, unlike the canonical MMR, NucS based system has low activity of transversions and is unable to correct indels (Castañeda-García et al., 2020).

1.5.4 MTB drug resistance and antibiotics

Antibiotics are a major selective force for genetic resistance. In their review, McGrath et al. argue that numerous single nucleotide polymorphisms (SNPs) might be occurring during infection, but that allelic fixation of these mutations depends on the ability of the bacilli to overcome host defence mechanisms. However, during drug therapy, drug resistance-conferring mutations will be readily fixed (McGrath et al., 2014). These mutations can impair bacterial fitness even in the absence of antibiotic therapy (McGrath et al., 2014). Interestingly, clinical isolates conferring these mutations are usually more fit when compared to laboratory-generated mutant strains suggesting that mutations that correct a loss of fitness (compensatory mutations) are also readily fixed in a population (McGrath et al., 2014). All these speak to a compensatory evolution that eliminates the fitness cost associated with antibiotic resistance (Fonseca et al., 2015); hence, there is a need to study these compensatory mechanisms.

Drug resistance is facilitated where antibiotics are present at sub-inhibitory concentrations (McGrath et al., 2014) The presence of sub-inhibitory concentrations can result from non-adherence, poor penetration and poor activity due to interaction with other inhibitory substances in cells. Treatment of other respiratory infections with broad spectrum antibiotics can also expose MTB to sub-inhibitory concentrations which can drive resistance mutations (McGrath et al., 2014). In fact, *in-vitro* mutagenesis has been observed to be increased during exposure of both MSM and MTB to sub-inhibitory concentrations of ciprofloxacin, a fluoroquinolone which acts by inducing the mycobacterial SOS response (Malik et al., 2012).

Evidence suggest that there are complex metabolic events that occur subsequent to the interaction of antibiotics with their molecular targets which suggest that the initial interactions of drugs with their targets cannot fully account for the antibiotic lethality (Dwyer et al., 2015). Studies show that, even though their mechanisms of action do not directly target DNA metabolic processes, many bactericidal antibiotics ultimately result in DNA damage due to the generation of hydroxyl radicals (Kohanski et al., 2010, Grant et al., 2012). Formation of reactive oxidants due to antibiotics is implicated in the elevation of mutation rates (Kohanski et al., 2010). This occurs as a result of damage to DNA and the activation of error-prone components of the SOS response whose function is to repair or tolerate DNA damage (Kohanski et al., 2010). Recent studies on this subject provide further evidence that bactericidal antibiotics induce redox stress through the production of reactive oxygen species (ROS) (Dwyer et al., 2014). The mechanism behind this phenomenon is not well understood (Dwyer

et al., 2015). However, bactericidal antibiotics have been shown to induce elevated basal cellular respiration which is directly related to their efficacy (Lobritz et al., 2015). This elevated cellular respiration is a potential source of ROS which consequently leads to DNA damage (Lobritz et al., 2015). Elucidating the role of DNA repair in the emergence of antibiotic resistance is therefore critical: any mutations in these pathways will affect the ability of the bacteria to repair damage and consequently result in elevated mutation rates which could result in drug resistance. In turn, by revealing hierarchies of DNA damage repair and tolerance, the potential exists to develop new drugs that can potentiate existing anti-TB drugs(Reiche et al., 2017)

1.6. The mycobacterial DNA damage response

Host cells have developed mechanisms of killing MTB which include production of ROS by phagocyte oxidase (phOX2) and production of reactive nitrogen intermediates (RNI) by inducible nitric oxide synthase (iNOS). ROS and RNI react with cellular molecules including proteins, carbohydrates, lipids and nucleic acids resulting in a bactericidal environment (Ehrt and Schnappinger, 2009). Nonetheless, MTB has a formidable intrinsic resistome which enables it to persist in the face of all these stresses. MTB counteracts these unfavourable effects through detoxifying ROS and RNI (Ehrt and Schnappinger, 2009). In addition, MTB has the ability to overcome DNA damage through DNA damage repair and tolerance pathways (Davis and Forse, 2009).

1.6.1. The SOS response

The DNA damage response is critical for bacterial survival (Janion, 2008). The SOS response is the main mechanism for the regulation of DNA repair in many bacteria (Janion, 2008). The products of SOS regulon genes are involved in specific DNA repair mechanisms, tolerance of DNA damage, and induced delay of the cell cycle. (Kreuzer, 2013). In the SOS response, RecA coordinates expression of repair genes while LexA represses transcription of the genes during normal growth by binding to an upstream DNA sequence called the SOS box (**Figure 1.3A)** (Janion, 2008). The ultimate trigger of induction of the SOS response is the formation of single-stranded DNA (ssDNA)(Kreuzer, 2013). Recruitment of RecA on ssDNA is facilitated by RecBCD or RecFOR enzymes (Baharoglu and Mazel, 2014). Helicase and nuclease activity of RecBCD on double-strand DNA breaks (DSB) results in the formation of ssDNA which is a substrate for RecA (Baharoglu and Mazel, 2014). RecFOR recruits RecA to nicks and gaps

occurring on individual strands of DNA. RecA binds ssDNA and form nucleoprotein filaments that then acquires co-protease activity to facilitate the auto-proteolysis of the repressor, LexA (Baharoglu and Mazel, 2014). Apart from DNA damage-inducible genes, LexA also regulates transcription of both its own gene and the *recA* gene **(Kreuzer, 2013).** Hence, increased RecA protein levels correlate with a highly induced DNA damage response and cleavage of LexA, whereas increased LexA shuts down the DNA damage response when the inducing signal of ssDNA fades (Kreuzer, 2013).

1.6.2. RecA-independent response

The DNA damage response in mycobacteria is divided into two broad pathways according to mechanisms of initiation: the classical SOS response is RecA/LexA-dependent but there also exists a RecA-independent mechanism (Smollett et al., 2012). Most bacteria use RecA/LexA dependent mechanisms during DNA damage response (Wang et al., 2011). In MTB, it has been found that a larger number of DNA damage-inducible genes are regulated independently of LexA and RecA (Rand et al., 2003). This RecA-independent (RecA-ND) (**Figure 1.3A**) mechanism was initially revealed when it was observed that LexA did not bind to upstream regions of the DNA damage-inducible genes, *uvrA* and *ssb* (Brooks et al., 2001). The RecA-ND mechanism was subsequently shown to predominate in MTB (Rand et al., 2003). In their study, Rand *et al* found that most DNA damage-inducible genes which are involved in DNA repair or recombination were inducible in an MTB mutant deficient in *recA*. Genes inducible through this RecA-independent mechanism included those involved in base excision repair (e.g., *xthA*, *nei*), nucleotide excision repair (NER) (e.g., *uvrB*), damage reversal (e.g., *ogt*) and recombination (e.g., *recA*, *radA*) (Rand et al., 2003). The presence of this mechanism has been supported by characterization of a common motif in these genes which has been identified as a RecA/LexA independent promoter (RecA-NDP) (Gamulin et al., 2004). A regulator that binds to the conserved motif in this promoter was initially proposed to be a *clp* gene regulator, ClpR, corresponding to Rv2745c and MSMEG_2694 in MTB and MSM respectively (Wang et al., 2011). However, this was contrary to results obtained in *Corynebacterium glutamicum* in which it was observed that deletion of *clgR* did not affect the transcriptional response to mitomycin-C (MMC) treatment and UV exposure (Engels et al., 2005).

Subsequent work has established that proteasome accessory factors B and C (PafBC), which are encoded in the Pup-proteasome system (PPS), serve as transcriptional activator of the

LexA/RecA-independent mycobacterial DNA damage response (Müller et al., 2018). Hence the *recA/lexA* -independent DNA response is now referred to as the PafBC- mediated pathway (**Figure 1.3B**). The PPS system is a protein degradation machinery which uses a ubiquitin-like tagging system (Olivencia et al., 2017). PafBC has been shown to bind to the RecA-NDp motif, upregulating at least 150 proteins involved in DNA damage repair. Most genes expressed this way function in recombination, replication and repair (Müller et al., 2018). These include among others, genes involved in the nucleotide excision repair, base excision repair and recombination (Müller et al., 2018).While the classic *recA/lexA* SOS response is induced by single-stranded DNA molecules and upregulation of *recA*, the signal used to activate induction of the PafBC-mediated response is unknown (Müller et al., 2018). PafBC proteins have been shown to be at constant levels in bacterial cells. The unknown signal which is triggered by DNA damage is hypothesized to bind to PafBC and consequently enable it to drive expression of DNA damage repair genes from the RecA-NDp (Müller et al., 2018). In the PafBC-mediated pathway, return to baseline after DNA stress is facilitated by the PPS which degrades the DNA repair proteins that are no longer required (Müller et al., 2018).

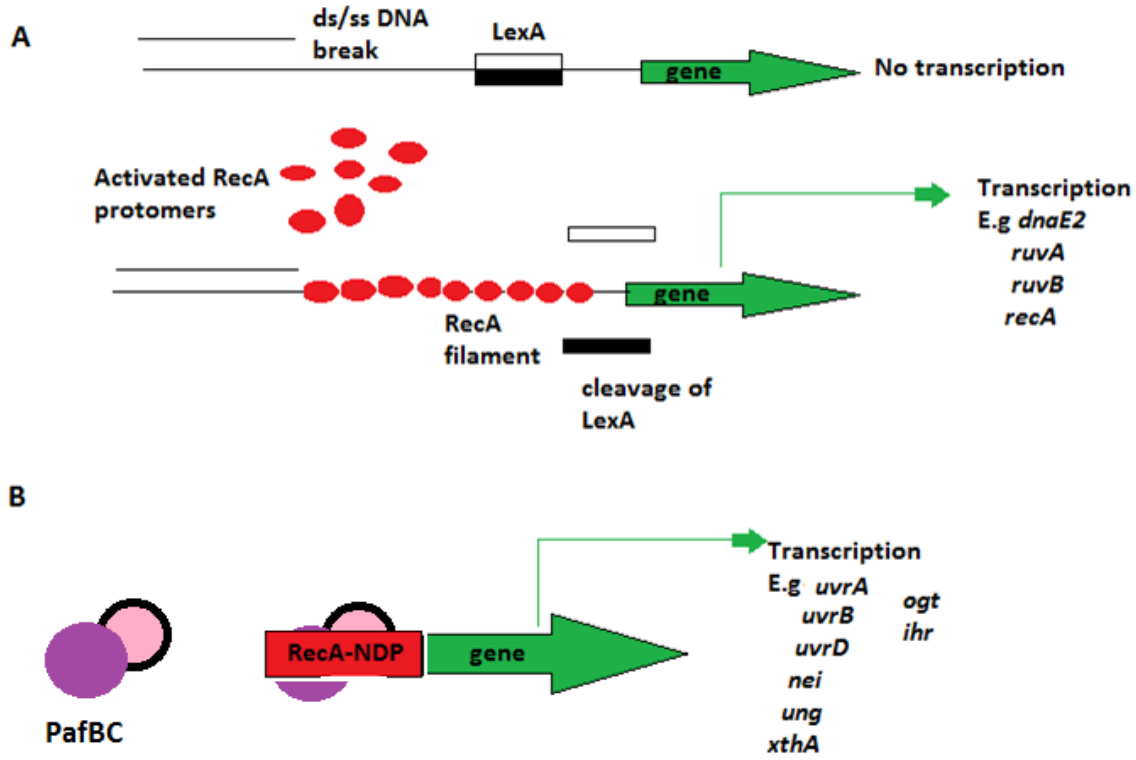


Figure 1. 3 DNA damage response pathways of mycobacteria A) The classical SOS response. During normal growth LexA acts as a repressor by binding to the SOS box of DNA damage-inducible genes, thereby preventing their transcription. Upon genotoxic stress, RecA is recruited to ssDNA, forming nucleoprotein filaments and acquiring a co-protease activity

which facilitates cleavage of LexA. After LexA has been cleaved, transcription of damage-inducible genes can occur. Schematic adapted from (Andersson and Hughes, 2014). B) RecA/LexA-independent DNA damage response. The presence of this mechanism has been supported by characterization of a common motif in these genes which has been identified as a RecA/LexA-independent promoter (RecA-NDP) (Gamulin et al., 2004). The regulator that binds to the conserved motif in this promoter is PafBC (Müller et al., 2018).

1.7. DNA damage repair and tolerance mechanisms

There are many forms of DNA damage: a base can be eliminated or modified by oxidation or alkylation, strand crosslinking can occur if two bases are covalently joined, single-strand or double-strand breaks can occur when one or both DNA strands are cleaved. Therefore, the repair mechanism(s) employed depends on the kind of damage that has occurred (Davis and Forse, 2009).

1.7.1 DSB repair systems in MTB

The most cytotoxic form of DNA damage are DSBs because their repair is critical for chromosomal replication to proceed (Gupta et al., 2011). MTB contains three genetically distinct DSB repair systems: homologous recombination (HR), nonhomologous end-joining (NHEJ), and single-strand annealing (SSA) (Gupta et al., 2011).

1.7.1.1. Nonhomologous end joining

NHEJ repair is facilitated by Ku and LigD proteins (Gupta et al., 2011). NHEJ does not require a homologous template for repair. It is therefore mutagenic since it can add additional nucleotides or result in deletions in the DNA surrounding the break (Gupta et al., 2011). NHEJ has also been shown to function mostly in non-replicating cells where dNTPs have been depleted and ribonucleotides (rNTPs) are present in abundance. Even though the exact role for the preference of ribonucleotides is not known, it has been suggested that NHEJ may be critical for repair of DSBs during MTB persistence. The rNTPs incorporated into the DNA are probably processed by RNase enzymes/PolA (Singh, 2017). In their study using MSM, Gupta et al.,2011 observed that mycobacteria preferentially avoid mutagenic repair by NHEJ because a deletion mutant lacking the capacity for HR did not result in increased use of the NHEJ pathway. In fact, they observed that HR, an error-free repair pathway, is a dominant repair pathway of DSB repair and accounted for 40% of DSB surviving g cells (Gupta et al., 2011).

1.7.1.2. Homologous recombination

HR is a RecA-dependent error-free DNA repair mechanism (Gupta et al., 2016). It is usually preferred over NHEJ for DSB repair (Gupta et al., 2011). HR starts with resection of DSB by helicase-nuclease which creates a 3' tailed ssDNA that is then bound by single-strand binding proteins. RecA then forms nucleoprotein filaments on the ssDNA which facilitates homology search and strand invasion in the homologous sister strand (Gupta et al., 2016). Steps following strand invasion include: DNA synthesis, resolution of Holliday junction intermediates and, finally, gap filling and nick ligation (Gupta et al., 2016). Resolution of Holliday junction intermediates is facilitated by among others, RuvABC, RuvC and RecG. RuvA and RuvC have been shown to be upregulated during UV damage of MTB and RecG was found to be upregulated in macrophages in a mouse model of infection (Singh, 2017). Strand exchange during HR repair is facilitated by RecA, SSBa and SSBb proteins (Singh, 2017). Unlike other bacteria, RecBCD helicase-nuclease is not used for DSB resection and it does not function in mycobacterial HR (Gupta et al., 2016). Instead, AdnAB is a helicase-nuclease that is responsible for HR in mycobacteria (Gupta et al., 2016). Deletion mutants of AdnAB in MSM were hypersensitive to DNA cross-linkers, alkylating agents and ionising radiation (Gupta et al., 2016).

MTB strains deficient in NHEJ, HR or both were tested in different animal models of infection including guinea pigs, mouse hollow-fibre model, C57BL/6J mice and C3HeB/FeL mice and showed no difference in mortality, bacterial load and histopathology between WT and the mutants (Heaton et al., 2014). Several explanations have been suggested for these results including that DSB repair is not required for MTB pathogenesis or that the animal models do not inflict DSBs; alternatively, there could be other repair pathways that are compensating the lack of NHEJ and HR (Heaton et al., 2014).

1.7.1.3. Single strand annealing pathway

A mycobacterial RecBCD-dependent SSA pathway has also been found to repair DSBs in wild type (WT) cells (Gupta et al., 2011) This pathway is believed to be particularly important in repairing DSBs that arise in repetitive areas of the mycobacterial chromosome (Gupta et al., 2011). The mechanism is initiated when DSB are flanked by repeats on both sides. Complementary single-strands are unfolded by resection of DNA strands (Gupta et al., 2011). The repeats on each DNA strand then anneal by complementary base-pairing to form a duplex and the segments which were between the repeats are lost. Therefore, SSA is a potent source

of deletions (Gupta et al., 2011). In mycobacteria, SSA is mediated by RecBCD. However, RecO has also been linked to the SSA pathway(Singh, 2017).

1.7.2. Base excision repair (BER)

DNA helix non-distorting lesions are generally repaired by BER. Here, bases that have been alkylated, deaminated and oxidized are repaired (Singh, 2017, Davis and Forse, 2009). Oxidative damage due to antibiotics is implicated in the elevation of mutation rates (Kohanski et al., 2010). Oxidative stress causes 8-oxo-G (OG), which are oxidised guanine nucleotides. OG causes mutations due to mispairing with adenine (Singh, 2017). In most bacteria, OG damage is countered by the GO system which employ glycosylases, MutM, MutY, MutY and MutT (van Loon et al., 2010). These DNA glycosylases remove damaged bases from the DNA, creating abasic sites (van Loon et al., 2010). These sites are then processed by apurinic/apyrimidinic (AP) endonucleases which generate nicked DNA by incising the DNA backbone. DNA ligase and DNA polymerase I (PolA) are then recruited to fill the nick site (van Loon et al., 2010, Singh, 2017).

The importance of BER in MTB is highlighted by the presence of numerous glycosylases which show functional redundancy in this pathway (Minias et al., 2019). Mycobacteria encodes four MutT, two MutM (Fpg) and one MutY enzymes as a defence against OG (Cole et al., 1998, Mizrahi and Andersen, 1998). Fpg/MutM removes OG lesions (Singh, 2017). If an adenine does get incorporated because of the presence of OG, then adenine is removed by MutY (Singh, 2017).

A recent study suggest that mycobacteria has a complex multi-layered defence system against OG damage (Dupuy et al., 2020). This study show that MutY/MutM1 are critical for protecting against OG-mediated DNA damage. It also shows that MutT enzymes of mycobacteria are involved in nucleotide sanitation which facilitates resistance to antibiotic lethality (Dupuy et al., 2020). However, a critical relationship exist whereby in the absence of MutT enzymes, MutY/M and accessory DNA polymerases contribute to antibiotic- induced lethality (Dupuy et al., 2020)

Several BER genes have been associated with pathogenesis of MTB; *e.g*., a transposon mutant of *fpg* was attenuated in a primate model of infection, while *mutY* and *mutT* mutations have been observed in drug-resistant strains of MTB (Minias et al., 2019).

1.7.3. Non- canonical mismatch repair

Errors during DNA replication can lead to mismatched bases (Singh, 2017). MMR pathway detects mismatched bases in newly synthesised DNA strands. This repair pathway plays a major role in avoiding mutations and preventing recombination between homeologous DNA strands (Singh, 2017). The classic MMR pathway in most bacteria is mediated by MutS and MutL. MTB however does not have this canonical MMR system (Mizrahi and Andersen, 1998). Nevertheless, the absence of MMR does not affect the mutation rates of MTB when compared to other MMR bearing bacteria (McGrath et al., 2014). An active mutation avoidance mechanism which is facilitated by NucS, an endonuclease with no structural similarity to the classical MMR proteins protein was recently identified in mycobacteria (Castaneda-Garcia et al., 2017). Deletion mutants of *nucS* in MSM show a phenotype that is consistent with MMR inactivation in other bacteria; *i.e.*, increased mutation and homeologous recombination rates (Castaneda-Garcia et al., 2017). Whereas the functions of NucS are similar to MutSL-based MMR on repairing transitions, it is however unable to correct small indels and has low activity on transversions when compared to the canonical MMR enzymes(Castañeda-García et al., 2020)

The potential for redundant or overlapping functions implies the potential existence of DNA repair hierarchies. It is important, therefore, to understand how these repair mechanisms are coordinated and what happens when certain DNA repair components are disabled. This is especially so given that these mechanisms, together with their complementary pathways, ensure mycobacterial survival in the face of otherwise lethal genotoxic stress provided by both the host and anti-TB drug, a phenomenon which facilitates drug resistance. Inhibition of these mechanisms together with their complementary pathways might therefore be used to target MDR-TB. The work presented in this thesis will focus on NER and trans lesion DNA synthesis (TLS), two pathways which are involved in DNA damage repair and tolerance, respectively.

1.7.4. Mycobacteria NER pathway: UvrB enzyme

NER occurs in both prokaryotes and eukaryotes, and involves a very similar processes, however, while UvrABC proteins are used in bacteria, eukaryotes use a larger number of proteins (Truglio et al., 2006). Three key steps are involved in the NER process: damage detection and verification, incision, and excision of damaged oligomer. These processes are carried out by four proteins, namely: UvrA, UvrB, UvrC and UvrD (Kisker et al., 2013).

Initially, NER was thought to have evolved for repairing lesions caused by UV irradiation, *e.g*. thymine dimers. But it has since been recognized that, unlike other repair mechanisms (*e.g*., BER), NER can recognize and repair a wide range of DNA damage which may be chemically or structurally unrelated (Truglio et al., 2006). NER facilitates repair of DNA damage such as deamination of bases, abasic sites, intra-strand crosslinks (*e.g*., cyclobutane pyrimidine caused by UV irradiation) and inter-strand crosslinks (caused by chemicals such as mitomycin C) (Kurthkoti and Varshney, 2012, Davis and Forse, 2009). DNA distortion seems to be a major component of the recognition process by NER with gross distortions being incised more rapidly than those that distort DNA less (Truglio et al., 2006).

Even though the different proteins of the NER pathway have specific functions, they work in a cascade to remove toxic lesions through a multistep process (**Figure 1.4**). The initial damage sensing is carried out by UvrA, which exists as a dimer and is a member of the ATP-binding cassette (ABC) superfamily of ATPases (Kisker et al., 2013). After initial damage has been recognized, UvrB is recruited to verify the damage. This occurs through the function of a UvrB:DNA complex (Truglio et al., 2006). UvrB alone does not have strong binding affinity to dsDNA. Therefore, UvrA functions to load UvrB onto damaged sites and, consequently, induces a conformational change that allows UvrB to bind tightly to damaged DNA (Truglio et al., 2006). Initially, a UvrA2:UvrB2:DNA complex that contains two molecules of each of the proteins is formed. UvrB then verifies that the damage recognized by UvrA exists (Truglio et al., 2006). In fact, it has been shown that alone, UvrA performs a 3-dimensional search for damage by binding to DNA for approximately 7 seconds and is also able jump from one molecule to the other (Kad et al., 2010). However, in the presence of UvrB, a one dimensional sliding motion is assumed by some UvrA:UvrB complexes (Kad et al., 2010). Furthermore, there is six-fold increase in residence time (Kad et al., 2010).

If damage is absent (which is possible because UvrA can bind to undamaged DNA), UvrB dissociates and facilitates dissociation of UvrA from DNA. If DNA damage is present, a UvrB:DNA pre-incision complex forms which signals departure of UvrA and is recognized by UvrC (Truglio et al., 2006). UvrB plays a central role in the NER pathway because it interacts not only with UvrA but also with other NER proteins (Kisker et al., 2013). UvrB is composed of five domains: 1a, 1b, 2, 3, and 4 (Kisker et al., 2013). Domain 1b facilitates interaction with DNA, domain 2 interacts with UvrA and domain 4 interacts with both UvrA and UvrC (Kisker et al., 2013).Domains 1a and 3 harbours six helicase motifs making UvrB a member of the

helicase superfamily (Truglio et al., 2006). These domains share high structural similarity to the DNA helicases PcrA, Rep and NS3. Furthermore, an ATP binding site is located at the interface of domains 1a and 3 (Truglio et al., 2006). Unlike other helicases, UvrB is associated with limited DNA unwinding with only up to 22 nucleotides displaced depending on the melting temperature of the DNA (Truglio et al., 2006). It is proposed that the helicase domains of UvrB are used to distort DNA at lesions in order to facilitate recognition and incision by UvrC (Truglio et al., 2006).

The pre-incision UvrB:DNA complex signals recruitment of UvrC which catalyses the incision of the DNA lesion (Kisker et al., 2013). This protein contains two catalytic sites that are responsible for 3' and 5' incision reactions (Kisker et al., 2013). Post-incision events include removal of the incised oligomer and UvrC (possibly by UvrD), the removal of UvrB possibly by Pol I, and the subsequent filling of the gap. Finally, the repair patch is sealed by DNA ligase (Kisker et al., 2013).

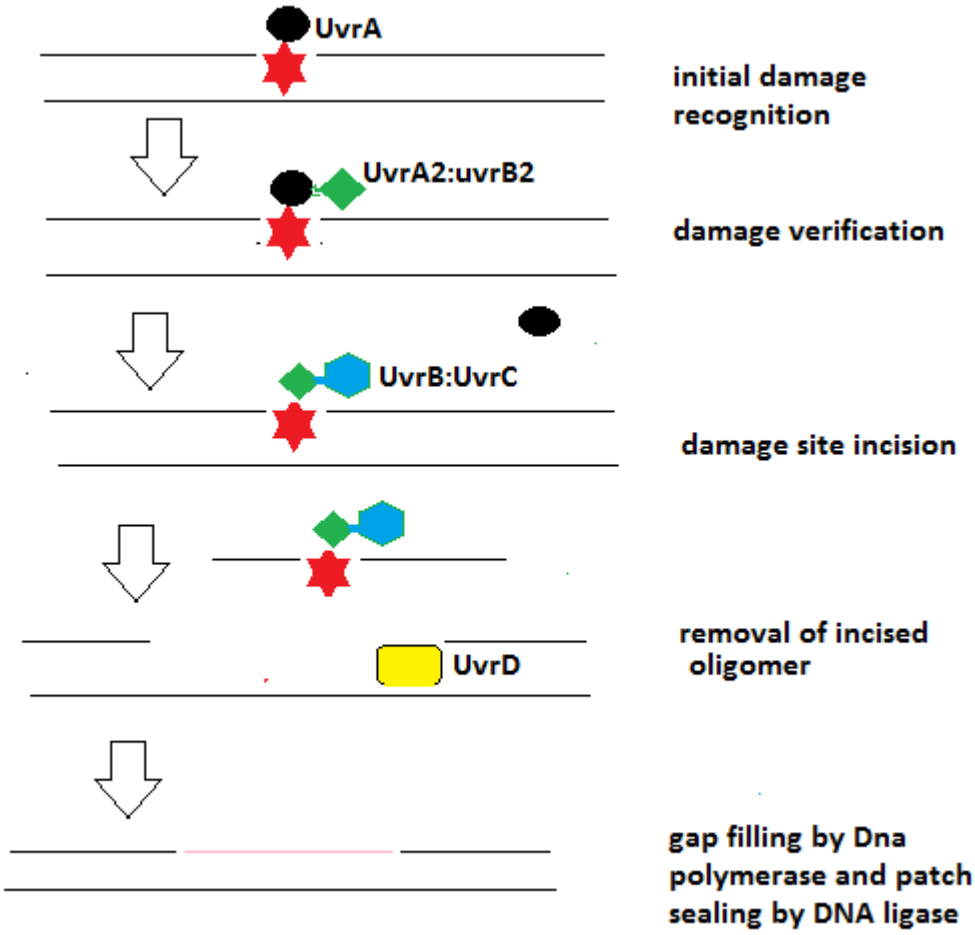


Figure 1. 4 NER pathway. The key enzymes of the NER pathway work in a cascade to remove toxic lesions through a multistep process. The initial damage sensoring is carried out by UvrA.

After initial damage has been recognized, UvrB is recruited to verify the damage. Initially an UvrA2:UvrB2:DNA complex that contains two molecules each of the proteins is formed, If DNA damage is present, a UvrB:DNA preincision complex forms which signals for departure of UvrA and is recognized by UvrC (Truglio et al., 2006). UvrC catalyses the incision of the DNA lesion. Post incision events include removal of the incised oligomer and UvrC by UvrD, the removal of UvrB possibly by Pol I and the subsequent filling of the gap. Finally, the repair patch is sealed by DNA ligase(Kisker et al., 2013). Schematic adapted from (Kurthkoti and Varshney, 2012).

All mycobacterial genomes (including MTB, MSM and BCG) have homologues of the key NER proteins (Singh, 2017, Güthlein et al., 2009). In MSM, UvrB and UvrD proteins have been shown to be essential for survival *in vitro* following UV irradiation, and ROI and RNI exposure (Güthlein et al., 2009). Mutants lacking these genes were hypersensitive to UV, MMC, and ROI/RNI-generating substances (Güthlein et al., 2009). In a study which used transposon mutagenesis to identify genes involved in ionizing radiation in *E. coli*, it was observed that nonirradiated populations contained numerous transposon insertions in the *uvrB* gene, indicating that mutations in this gene are dispensable for normal growth (Byrne et al., 2014). However, after passage 5 of irradiation no insertions were observed in this gene indicating that any cells with mutations in *uvrB* were eliminated from the population (Byrne et al., 2014). Studies in MTB show that expression of UvrA and UvrB is increased during exposure to hydrogen peroxide (Boshoff et al., 2003) and also during phagocytosis by macrophages (Graham and Clark-Curtiss, 1999). Furthermore, UvrA and UvrC expression was found to be elevated in MTB harvested from granulomas of TB patients (Rachman et al., 2006). In another study, *uvrC* mutants of MTB were shown to be more sensitive to UV irradiation when compare to the control but showed no difference in susceptibility to ROS generated by H_2O_2 (Prammananan et al., 2012).

The mycobacterial proteasome serves as a defence against oxidative and nitrosative stress (Darwin et al., 2003). In a study by Nathan and colleagues (Darwin et al., 2003), MTB genes required for defence against RNI were identified through genome-wide transposon (Tn) mutagenesis. Insertions in the proteasome-associated genes rendered the mutants sensitive to RNI. Notably, mutations in *uvrB* rendered MTB hypersensitive to nitrite (Darwin et al., 2003). The *uvrB* mutant was tested under different conditions and the results supported the function of this enzyme in protecting MTB against oxidative and nitrosative stress (Darwin and Nathan, 2005): WT mice infected with the *uvrB* mutant lived longer, as did mice lacking the inducible

nitric oxide synthase enzyme, indicating that NER-*uvrB* is required not only for nitrosative stress tolerance. In addition, MTB *uvrB* deletion mutants were highly susceptible to UV treatment *in vitro* (Darwin and Nathan, 2005). These studies show that the NER pathway plays an important role in the maintenance of mycobacterial genome integrity.

7.4.3. Unconventional functions of the NER pathway

Apart from the classical function of removing bulky lesions from DNA, the NER pathway is also implicated in acting as a backup for glycosylases in removing methylated and oxidised bases (Tark et al., 2008). Furthermore, apurinic sites seem to serve as substrates for the UvrABC proteins hence bringing together the BER and NER pathways (Tark et al., 2008).

It has also been hypothesised that the NER pathway might be a source of spontaneous mutations especially in resting cells: evidence suggests that both *E. coli* and human NER proteins cluster on undamaged DNA (Branum et al., 2001). Association of the NER pathway with mutagenesis in bacteria was first established by Nishioka and Doudney who observed that UV mutagenesis occurs in *E. coli* WT-strains but not in *uvrA*-inactivated strains at early time points after UV-irradiation (Nishioka and Doudney, 1969). Tark et al., 2008 also observed that lack of functional *uvrB* and *uvrC* genes reduces the frequency of stationary phase mutations in *Pseudomonas putida* (Tark et al., 2008). A recent study has tried to explore the precise mechanistic and genetic links between NER and mutagenesis. In this study, which was carried out in *E. coli*, the authors propose that NER induced mutagenesis (called NERiM) requires both DNA PolIV and PolV damage inducible DNA polymerases (Janel-Bintz et al., 2017). They also confirm that NERiM exhibits a quadratic dose-response curve suggesting closely spaced lesions located on opposing strands. Furthermore, the mutations are replication independent and occur in stationary phase cells, providing an intriguing possibility for ongoing evolution in the absence of replication (Janel-Bintz et al., 2017).

1.7.5. DNA damage tolerance: TLS and DnaE2

DNA lesions that escape the repair process can result in arrested DNA replication, inhibiting cell cycle completion (Friedberg et al., 2006). To avoid this, cells have evolved mechanisms which allow them to relieve stalled replication. These mechanisms are collectively referred to as DNA damage tolerance mechanisms (Friedberg, 2005). These include TLS, post-replicative gap filling and replication-fork regression (Friedberg et al., 2006). During post-replication gap filling and replication fork regression, cells avoid continuing DNA replication in the immediate

region of DNA damage where replication has stalled. This avoids accumulation of mutations; therefore, the two mechanisms are generally error-free (Friedberg, 2005). The most widely studied DNA damage tolerance mechanism in bacteria is TLS.

TLS promotes survival against endogenous insults by allowing cells to replicate past different DNA lesions and distortions (Davis and Forse, 2009). The specialized TLS polymerases are able to copy across lesions following which replicative polymerases resume synthesis (Davis and Forse, 2009). These polymerases often lead to mutations and hence are referred to as error-prone (EP) DNA polymerases (Goodman, 2002). Pol II, Pol IV (encoded by *dinB*) and Pol V (encoded by *umuD* and *umuC*) are TLS polymerases harboured by *E. coli* (Goodman, 2002). Pol V and Pol IV are members of the EP Y-family, while Pol II is a member of B-family polymerases (Goodman, 2002). In *E. coli*, these polymerases are induced as part of the SOS response (Goodman, 2002). Most mycobacterial species do not possess homologues of PolII but do encode between 3-5 Y-family polymerase homologues. In MTB, these polymerases were originally annotated as DinX and DinP (Davis and Forse, 2009), but have subsequently been referred to as DinB1 (Rv1537) and DinB2 (Rv3056), respectively (Kana et al., 2010). Neither DinB1 nor DinB2 showed inducible expression during DNA damage in MTB (Boshoff et al., 2003). Furthermore, deletion of these genes did not affect the rate of spontaneous mutations (as judged by RIF resistance) and did not have an impact on the growth of MTB *in vivo* in a mouse model or *ex vivo* in macrophages (Kana et al., 2010). This indicates that DinB homologs from MTB might not be carrying out the same functions as their counterparts from other organisms (Kana et al., 2010).

In the *E. coli* model, the replisome is divided into three catalytic centres: the core complex, the clamp loader complex and the helicase primase complex (Timinskas et al., 2014). A large and stable complex termed DNA polymerase III is formed by the assembly of the core complex and clamp loader complex. The core complex is composed of Ɛ-exonuclease (encoded by *dnaQ*) which offers proofreading activity, protein Θ (encoded by *holE*) which stabilizes the exonuclease, and the Pol IIIα subunit (encoded by *dnaE*) which is the replicative polymerase (Ditse et al., 2017b, Timinskas et al., 2014). Pol IIIα subunits belong to C-family DNA polymerases which are further divided into PolC-type and DnaE-type polymerases. The DnaE-type group is subdivided into three groups being DnaE1, DnaE2 and DnaE3 (Timinskas et al., 2014). In mycobacteria, dnaE1 encodes the essential Pol IIIα subunit, while a second *dnaE2*-encoded C-family subunit appears to function solely in TLS (Ditse et al., 2017b). Expression

of both *dnaE1* and *dnaE2* is upregulated in the DNA damage response in MTB (Boshoff et al., 2003); therefore, despite possessing multiple putative Y-family TLS polymerases, the C-family members appear to be the only polymerases which are induced following genotoxic stress in mycobacteria (Warner et al., 2010, Boshoff et al., 2003).

In mycobacteria, replicative fidelity is ensured by DnaE1(Rock et al., 2015). The primary source of proofreading, the replicative exonuclease, is encoded in the polymerase and histidinol phosphatase (PHP) domain of DnaE1: mutants carrying targeted substitutions of amino acids essential for PHP exonuclease function exhibited severe growth defects and massively elevated mutation rates. Therefore, the intrinsic polymerase fidelity and/or proofreading are able to preserve the mycobacterial error rate (Rock et al., 2015). While both DnaE2 and DnaE1 share many major structural features, DnaE2 lacks the C-terminal τ-interacting domain. The absence of this region has been proposed to account for the inability of DnaE2 to substitute essential replicative function (DITSE et al., 2017a).

1.7.5.1. DnaE2 in Mycobacteria DNA repair and mutagenesis

Induction of the SOS response has been associated with an increase in targeted and untargeted mutations in intracellular organisms (Schlosser-Silverman et al., 2000). In MTB, DnaE2 has been implicated in damage-induced mutagenesis (Boshoff et al., 2003). This study also showed that *dnaE2* expression is RecA/LexA-dependent (Boshoff et al., 2003). To confirm the function of DnaE2, allelic exchange mutants of MSM and MTB were created. UV-induced resistance to RIF was completely lost in the *dnaE2* mutants. Also, the mutants failed to persist *in vivo* therefore, DnaE2 in MTB has an important role in DNA damage tolerance and also in generating mutations that increase fitness (Boshoff et al., 2003).

In most bacteria, DnaE2 is present in a DNA-damage inducible, LexA-regulated *imuA-imuB-dnaE2* gene cassette (Erill et al., 2006). In MTB, *dnaE2* is in a "split cassette", *imuA'-imuB*/*dnaE2*, in which *dnaE2* is separated from its essential "accessory proteins" by 24.7 kb (Warner et al., 2010). The *imuA-imuB-dnaE2* cassette has been shown to be absent in organisms that have *imuDC* (Pol V)-facilitated TLS (Erill et al., 2006). DnaE2 does not have the C-terminal region. This contrasts with Pol IIIα subunits of C-family DNA polymerases in *E. coli* (Warner et al., 2010) In *E. coli*, the C-terminal region facilitates interaction of the α subunit with the clamp loader subunit τ. The interaction of α and τ subunits enables

simultaneous synthesis of the lagging and leading strands during DNA replication by Pol III. Therefore, the absence of the C-terminal in mycobacterial DnaE2 implies that it does not substitute essential replicative function (Warner et al., 2010). DnaE2 has also been shown to lack a β-clamp binding motif. Instead, a study of mycobacterial *imuA'-imuB/dnaE2* split cassette shows that ImuB interact with the β-clamp, DnaE2 and ImuA through a β-clamp binding motif and a C-terminal domain respectively (Warner et al., 2010).

1.7.5.3.TLS and pathogenesis other bacteria

Combined evidence from metagenomics, taxonomic and phylogenetic analysis has been proposed to implicate DnaE2 in adaptation of marine progenitors to land colonisation (Wu et al., 2014). The study also suggested that acquisition of DnaE2 was followed by GC content increase and genome expansion (Wu et al., 2014). The authors further show that high GC, DnaE2-containing bacteria are characterized by enhanced metabolic flexibility; for example, *Symbiobacterium thermophilum* a DnaE2-containing bacterium possesses a variety of respiratory systems found only in gram negative bacteria (Wu et al., 2014). In addition to this evolutionary role, DnaE2 has also been found to facilitate sporulation in *Myxococcus xanthus*, a gram negative soil bacterium; DnaE2 was dispensable for cellular growth and social motility but indispensable for sporulation (a stress-induced phenomenon which occurs during nutrient starvation) (Peng et al., 2017). The study also show that overexpression of DnaE2 improves growth development and sporulation but also raised mutation rates (Peng et al., 2017). However, compared to the essential *dnaE1*, expression of *dnaE2* was always low in the cell. For this reason, it was argued that DnaE2 ensures low mutations during cell adaptation to new environments in order to avoid cell damage from high mutation rates (Peng et al., 2017).

Even though DnaE2 is clearly involved in mutagenic TLS in mycobacteria (Boshoff et al., 2003), this function appears not to be universal: In *Pseudomonas putida*, for example, abrogation of DnaE2 function enhanced the appearance of base substitution mutations in starving cells. The same effect was also observed under UV irradiation(Koorits et al., 2007). Together, these results indicate that DnaE2 in *P. putida* might act as an anti-mutator, a phenotype contrary to that which has been observed in other bacteria (Koorits et al., 2007).

Balance in Mycobacterial error Free and error prone repair

As discussed in sections above, DNA repair can either be error free of error prone. In MTB, error free pathways include NER, BER, HR and NucS-MMR (Davis and Forse, 2009, Castaneda-Garcia et al., 2017) and error prone repair include TLS, SSA and NHEJ (Davis and Forse, 2009, Singh, 2017). The vast majority of DNA repair mechanisms are error-free (Davis and Forse, 2009). Replicative polymerases also have proofreading exonucleases which remove misincorporated nucleotides (Tippin et al., 2004).This shows that the survival of organisms is dependent on maintenance of genome integrity. Nonetheless, organisms have to adapt to changing environments which in the context of pathogenic mycobacteria include antibiotic stress and host immune responses (Warner, 2010). Genomic adaptation occur through selection and maintenance of chromosomal mutations in MTB (Warner, 2010). These purposeful mutations result in genetic diversity (Tippin et al., 2004). Mutations can also occur where DNA damage is repaired through error-prone mechanisms e.g. stalled replication can only be repaired through mutagenic TLS polymerases (Tippin et al., 2004). Genomic adaptation in MTB is evidenced by the emergence of MDR and XDR (Warner, 2010). Most mutations occur at a cost to the organism hence it is crucial to maintain a balance between stability and mutagenesis. In general, the relative fitness of an organism is impacted by the interplay between error-free repair and error-prone repair (Warner, 2010).

The importance of DNA repair mechanisms in MTB pathogenesis shows their potential as drug targets (Reiche et al., 2017, Warner, 2010). Mechanisms that maintain genome integrity through error-free repair and those that facilitates adaptation through mutagenesis can be targeted (Warner, 2010). One strategy of inhibiting DNA repair proteins requires targeting those that are conditionally essential. These are proteins that are non-essential *in vitro* but are essential for pathogenesis *in vivo*. Two such proteins are UvrB, an NER protein and DnaE2, a TLS protein (Warner, 2010)

1.8. Introduction to transposons

The PhD work described in this thesis was performed by applying genome-wide Tn mutagenesis in the non-pathogenic MSM mc^2155 model. A technique which is based on the use of transposable elements (Chao et al., 2016).

Transposable elements also known as "jumping genes" were discovered by Barbara McClintock in the 1950s during an analysis of maize chromosomes (Fedoroff, 2001). Tns have

since been discovered in almost all organisms including eukaryotes, archaea and prokaryotes. Tns have similar replication strategies to that of viruses, hence they are believed to have been transmitted horizontally (Friedli and Trono, 2015). Initially, Tns were termed "junk DNA" because they were thought not to have any active function in the genome. However, it is now well established that Tns are important facilitators of evolution in organisms (Friedli and Trono, 2015). This occurs through numerous processes such as gene mutations, transcriptional modulation, gene shuffling, genomic recombination, *etc*.(Friedli and Trono, 2015).

Although Tns are generally understood as random jumping genes, Tn excision and insertion is not entirely random (Craig, 1997). Some Tns have been found to possess considerable site selectivity. Site selectivity can depend on communication between the transposase and other accessory genes or it could be through interaction between target site DNA and the transposase (Craig, 1997). In fact, some Tns have a preference for targets that share a common consensus sequence, hence interaction of the transposase with these nucleotide sequences is a determinant of target selectivity (Craig, 1997). Nevertheless, the near random and hence the almost indiscriminate way of inserting into target DNA is a property of Tns that make them important molecular tools for single gene and genome-wide studies in micro-organisms (Hayes, 2003). The discovery of Tns made it easier to delete genes and characterize their phenotypes (Kleckner et al., 1975). Tns are therefore effective mutagens and numerous Tn-based molecular methods have been developed to harness this property (van Opijnen and Camilli, 2013). Currently, the construction of a library of bacterial strains each containing single randomly located Tns (insertional mutagenesis) is the most commonly used application of transposons (van Opijnen and Camilli, 2013).

1.8.1. Transposon Insertion Sequencing (TnSeq)

Next generation sequencing has revolutionised Tn mutagenesis methods. Through this technology, a new generation of Tn mutagenesis methods has been enabled which are collectively called Tn insertion sequencing, or TnSeq (Gray et al., 2015).TnSeq enables all strains in a pooled library to be screened with accurate quantification. Furthermore, through this method, the initial library can provide useful information as it can predict essential genes (Gray et al., 2015). In 2009, several different approaches of TnSeq were developed. These include: high-throughput insertion tracking by deep sequencing (HITS), transposon-directed

insertion site sequencing (TraDIS), Insertion sequencing (INSeq) and Tn sequencing (van Opijnen and Camilli, 2013).

These methods share the same basic principle which entails creating a Tn mutagenesis library (which by definition does not contain insertions in genes essential for growth), exposure of the library to different test conditions in order to select for conditional essential genes, and quantification of Tn junctions through deep sequencing (Chao et al., 2016). The relative abundance of each mutant corresponds to the fitness of that mutant under the particular condition (Chao et al., 2016). If Tn insertions are more abundant in a gene, then it is not required for fitness in that condition; if they are less abundant, it implies the gene is needed by the bacterium to survive or be fit in that environment (van Opijnen and Camilli, 2013). TnSeq has been used to identify essential genes under optimal *in vitro* conditions, to identify conditionally essential genes (*i.e*., genes that are required only under a specific condition), and also to identify and study genetic interactions (Long et al., 2015).

In this thesis, the ϴMycomarT7 Tn donor phasmid was employed. This is a temperature sensitive phasmid which contains a transposase gene (C9 Himar1) and a MycoMarT7 transposon (Murry et al., 2008). The phasmid produces phage at 30^{o}C, and transduction of bacterial cells is carried out at 37^{o}C, a temperature which does not allow replication of the phage. The MycoMarT7 transposon encodes a kanamycin resistance gene which enables selection of Tn mutants (Murry et al., 2008). An assumption underlying this approach is that the Tn inserts randomly and with equal probability at -TA- dinucleotides; therefore, absence of Tn insertions in a gene signifies essentiality(Chao et al., 2016). These are regions where mutations cannot be sustained biologically, hence are inferred as essential (Chao et al., 2016). This assumption has recently been challenged by DeJesus et al. who showed that not all -TA- dinucleotides are equally permissive for insertions. Specifically, regions which contain the (GC)GNTANC(GC) sequence are non-permissive for *Himar1* insertions (DeJesus et al., 2017a) hence highlighting some of the limitations of the TnSeq method.

Essential genes or conditionally essential genes are those that lack of have significantly lower than average Tn insertions (DeJesus et al., 2015). However, TnSeq data can be affected by many factors: Library saturation indicates the percentage of -TA- dinucleotides that have been covered in the genome. Ideally, a Tn library should reach a saturation of 50% to ensure that conclusions based on relative abundance are valid (DeJesus et al., 2015). This is because library saturation influences statistical analysis and statistical power (Chao et al., 2016). Analyses of

highly saturated Tn libraries can precisely distinguish essential regions and non-essential regions (DeJesus et al., 2015). In less well saturated libraries, there is an overlap between essential and non-essential region i.e. there is a high probability that non-essential regions will lack Tn insertions owing to chance and will therefore be incorrectly classified as essential (Chao et al., 2016).

One of the main assumptions of TnSeq is that a decrease in frequency of a specific mutant results from attenuated survival in that condition. However, Tn mutants are sometimes lost due to fitness-independent processes which may be resulting from experimental bottlenecks (Chao et al., 2016). Experimental bottlenecks are common in infection studies where only a few inoculated bacteria can establish infection in the host (Chao et al., 2016). Furthermore, the quality of TnSeq data can be affected by experimental differences during amplification and sequencing of Tn sites (Chao et al., 2016). A good example is the use of PCR to add sequencing adaptors: The PCR method can be affected by unequal amplification of DNA junctions due to differences in fragment sizes or uneven annealing efficiencies at Tn sites (Chao et al., 2016).

1.8.2. Tn mutagenesis and gene interaction studies

Through the use of Tn mutagenesis, a more efficient method of studying genetic interactions is possible (Joshi et al., 2006). Genes interact with each other, leading to novel phenotypes; therefore, genes cannot be studied in isolation (Phillips, 1998). The study of genetic interactions identifies mutations that alter each other's phenotypes. These interactions are called 'epistasis' (Joshi et al., 2006). Epistatic interactions can be negative, positive and suppressive (Joshi et al., 2006). Negative interactions are those that produce a more severe attenuation than expected. Positive interaction are mutations that produce a less severe fitness defect than expected and, in suppressor mutations, the consequence of gene deletions is countered (Costanzo et al., 2010).

A more severe fitness cost than expected is usually found in genes of separate pathways that perform redundant functions (Joshi et al., 2006). A less severe fitness cost is observed in genes of the same pathway that compensate each other and suppressor mutations usually induce compensatory pathways that enable the specific mutant strain to survive in an otherwise toxic environment (Joshi et al., 2006). Genetic interaction studies are useful as they elucidate pathways or processes that act in parallel but supporting similar biological processes towards

cell viability (Joshi et al., 2006). Genome-wide Tn mutagenesis can be carried on strains that lack genes of interest (query genes) in order to explore genetic interaction pathways in these backgrounds (van Opijnen and Camilli, 2013).

1.9. Using TnSeq to elucidate DNA repair hierarchies in mycobacteria

MDR-TB is a major stumbling block to TB control. This doctoral research project focuses specifically on elucidating mycobacterial DNA repair mechanisms that enable tolerance and repair of genotoxic stress when the key NER protein, UvrB, or the trans-lesion polymerase, DnaE2, are disabled. These two enzymes were selected since they represent dominant components of two separate but complementary arms of the mycobacterial DNA damage response: the DNA repair protein, UvrB, is part of the PafBC- mediated regulon whereas the DNA damage-tolerance polymerase, DnaE2, is part of the RecA/LexA SOS response. Most importantly, the NER/BER mutants are the only major DNA repair components which consistently drop out of studies of in vivo essentials and DnaE2 is the only gene so far implicated in induced mutagenesis (Warner, 2010). Furthermore, both NER and DnaE2 mutants have been associated with attenuated infection in vivo (Warner, 2010).

Genome-wide Tn mutagenesis was applied in the respective gene deletion backgrounds to identify conditionally essential repair pathways under DNA-damaging conditions, including exposure to antibiotics. This strategy aimed to identify targets for co-drugs that will shorten treatment duration and reduce the risk of drug resistance by severely limiting (or eliminating) the intrinsic capacity of MTB to tolerate lethal drugs for extended periods. A summary of the approach is shown in **Figure 1.5**. Tns allow random mutagenesis, effectively enabling all genes to be assayed (other than essentials) without bias. The approach adopted here enables "negative" screens – that is, where the intention is to identify genes which are lost from a pool. These two features are critical to this study which aims to identify genes which are conditionally essential in the defined genetic backgrounds.

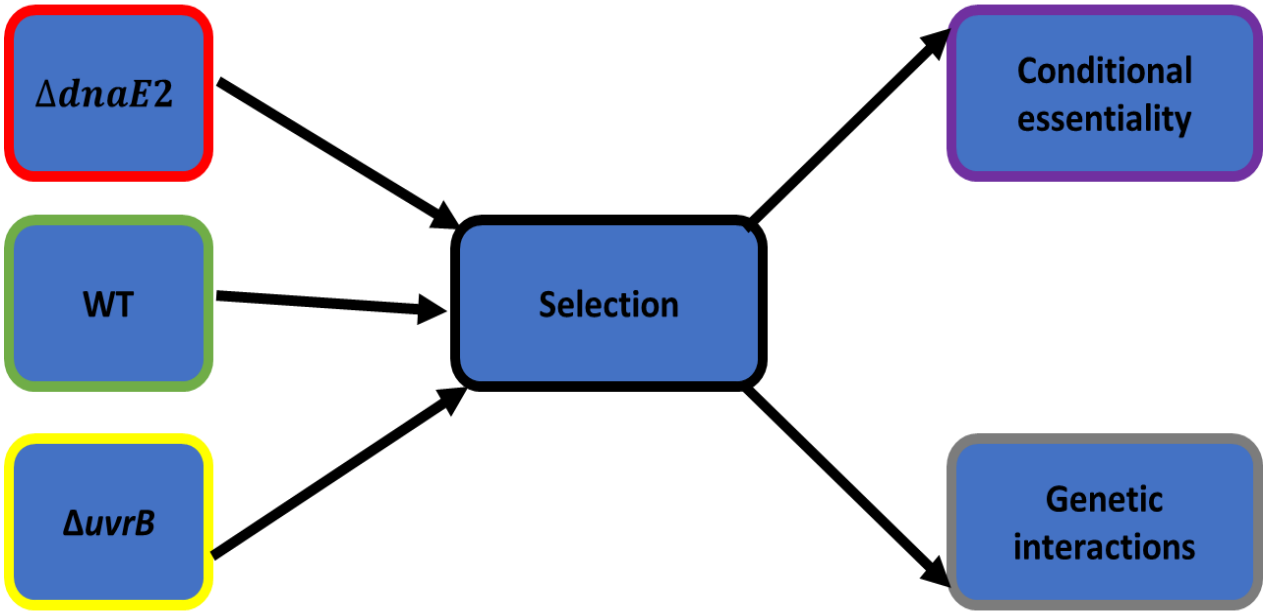


Figure 1. 5 General overview of objectives. Tn libraries will be generated in WT as well as *uvrB* and *dnaE2* KO mutants of MSM libraries will be selected on MMC and MOX treatments. After Illumina sequencing, conditionally essential genes will be identified by comparing un-treated and treated libraries for each strain. Genetic interactions will be identified by comparing four different libraries: WT untreated, WT treated, KO untreated and KO treated libraries.

Chapter 2: Materials and methods

2.1 General materials and methods

2.1.1 Bacterial strains, media and culture conditions

MSM was cultured on Middlebrook 7H10 solid agar supplemented with 0.5% glycerol and 10% (v/v) Middlebrook Oleic Acid Dextrose Catalase (OADC) enrichment (Difco^{TM}) (7H10/OADC), or in Middlebrook 7H9 broth supplemented with 0.2% glycerol, 10% (v/v) Middlebrook OADC, and 0.05% Tween80 (7H9/OADC). Cultures were incubated at 37^{0}C in shaking incubators (IncoShake Incubator, Labotec). Manipulations of MSM were carried out in a biosafety level 2 laboratory (BSL 2).

Cultivation of *E. coli* carrying plasmids was done on Luria-Bertani broth or agar (LB) (MacWilliams and Liao, 2013) Both solid and liquid cultures were incubated at 37^{o}C. When required, the final concentration of antibiotics was; 50µg/ml kanamycin (KAN). *E. coli* stocks were stored in nutrient broth containing 15% glycerol at -80^{o}C.

Table 2. 1 General bacterial strains used in this study.

Strain	Description/ Genotype	Reference/ Source
E. coli DH5α	*supE44 ΔlacU169* (*F80 lacZΔM15*) *hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	Promega
Mycobacterium smegmatis mc^{2}155	High-frequency transformation mutant of *MSM* ATCC 607	(Snapper et al., 1990)
Mycobacterium Smegmatis dnaE2 knockout (ΔdnaE2)	An unmarked MSM derivative in which *dnaE2* has been deleted.	Kindly provided by Prof. Digby Warner
Mycobacterium Smegmatis uvrB knockout (ΔuvrB)	An unmarked MSM derivative in which *uvrB* has been deleted	Kindly provided by Prof. Digby Warner
M.Smegmatis-mc^{2}155 ΔDRKIM	A derivative of MSM mc^{2}155 which lacks a 56kb duplicated region.	(Warner et al., 2006)

2.1.2 Bacterial transformations

2.1.2.1 Heat shock transformation of E. coli

Transformation of ligations DNA was carried out on *E. coli* DH5α. Competent cells were prepared by washing the *E. coli* cells with 100mM $Cacl_2$ and then resuspending in 100mM $Cacl_3$ and 80% glycerol and finally storing 100µl aliquots at -80^{0}C (University of Cape Town, MMRU unit Laboratory protocol). For transformation, *E. coli* DH5α cells were thawed on ice and 40µl cells were added to 1µl of plasmid DNA. After incubating for 30 minutes on ice, the cells were heat shocked at 42^{o}C for 45 seconds and then immediately placed on ice for 2 minutes. 200ml of 2x tryptone-yeast extract was added to rescue the cells at 37^{o}C while shaking at 250 RPM for one hour. These were then plated on LB plates containing the appropriate antibiotic and incubated at 37^{o}C overnight.

2.1.2.2 Electroporation into MSM

50 ml logarithmic liquid cultures of MSM mc^{2}155 (OD_{600} = 0.8- 1.0) were transferred to conical tubes and incubated on ice for 2hrs. After this incubation the cells were harvested by centrifugation at 5000 rpm for 10 minutes at 4^{0}C. The pellet was resuspended and washed in ½ volume (25ml) of ice cold 10% (v/v) glycerol. The washing procedure was repeated halving the volume of 10% glycerol each time to get a maximum of 3 washes. The competent cells were finally re-suspended in 2ml of 10% glycerol. To electroporate cells, 100ng of plasmid DNA was mixed with 100 ul competent cells. The mixture was then transferred to a 100 ul Bio-Rad gene pulser cuvette and a time constant protocol of: 1200 Voltage and a time constant of 5MS was used. The cells were then recovered in 600 ul of middle brook 7H9/OADC broth and incubated for 18hrs at 37^{o}C. The following day, cells were spread on 7H10 plates containing appropriate antibiotics and plates were incubated at 37^{o}C for 3-5 days (University of Cape Town, MMRU unit Laboratory protocol).

2.1.3 DNA manipulations

2.1.3.1 Chromosomal DNA extraction

MSM DNA was extracted using CTAB method adapted from Majumdar *et al*: 20 mL culture at an OD_{600} of 2-3 were pelleted by centrifugation and the pellet re-suspended in 500 µ1 TE buffer in a screw-capped tube. 50 µL of 10 mg/mL lysozyme was then added and cells incubated at 37°C overnight in a shaking incubator. The following day RnaseA was added to a final concentration of 25 µg/ml and cells incubated for 1 hr at 37^{o}C. 70 µ1 of 10% sodium dodecyl sulfate (SDS) and 50 µL of proteinase K (10 mg/mL stock solution) were then added and cells incubated for 1 h at 60°C with shaking in a thermomixer. Meanwhile, 5M NaCl and 10% CTAB were heated at 60^{o}C. 100 µl of 5M NaCl and 100 ul of 10% SDS were then added and mixed by inverting. Tubes were then incubated for 15 mins at 60^{o}C in a thermocycler. After the tubes have cooled, 700µl of chloroform/isoamyl alcohol (24:1) was added and tubes inverted until a homogenous white solution formed. Centrifugation of tubes for 10 mins at 4^{o}C (16000xg) was carried out. The aqueous phase was then transferred to tubes containing 700 µl ice cold isopropanol. The tubes were incubated in ice for 30 mins. Then tubes were centrifuged for 10 mins. The supernatant was drained, and the pellet washed with 70% ethanol. After another 10 mins of centrifugation the ethanol was drained, and the pellet dried in a speed vacuum concentrator for 10 mins. Finally, the pellet was dissolved in 50 ul of
TE buffer or nuclease free water (Majumdar et al., 2017)
For samples that had a lot of RNA contamination we did RNase treatment and then ethanol precipitation of DNA.

RNase treatment: In order to clean RNA contamination, we added 1/10 volume of RNase (at 10 mg/ml), then incubated at 37^{o}C for 1 hr. The enzyme was then heat inactivated at 70^{o}C for 15 mins and placed on ice to cool.

Ethanol precipitation: To precipitate DNA 1/10 volume of 3M cold sodium acetate was added to the sample together with 2.5 volume of 100 percent ethanol. The mixture was gently mixed by inverting a few times and then placed at -80^{o}C for 30 mins. The DNA was then pelleted by centrifuging for 20 mins at 4^{o}C. The supernatant was discarded, and the pellet was then washed with 70 % cold ethanol. The sample was centrifuged for 5 mins at 4^{o}C and then

all the ethanol was pulled off with a pipette tip. The pellet was vacuum dried and then resuspended in 50 – 100 ul of TE buffer or nuclease free water.

2.1.3.2 Plasmid Extractions

Plasmid extractions from *E. coli* DH5α cells were done using the QIAprep spin miniprep kit (QIAGEN). The procedure was carried out as per the manufacturer's instructions. In cases where we did not have the kit, we used the alkaline lysis method. To do this, 1.5 ml of *E. coli* culture was centrifuged at 14 000 RPM for 1 min (Eppendorf 5415D). After discarding the supernatant, the cells were resuspended in solution I (0.5 M glucose, 50 mM Tris-HCl pH 8 and 10 mM EDTA). The mixture was vortexed and then 200ul of solution II (0.2 M NaOH, 1% SDS) was added and mixed by inverting 3-5 times. After this, 300 ul of solution III (3 M potassium acetate, pH 5.5) was added and mixed by inverting 3-5 times. The sample was then centrifuged at 14000 RPM for 5 minutes at room temperature. After centrifugation the supernatant was RNase treated (1 µl of 10 µg/ml stock, Sigma Aldrich) for 10 min at room temperature. Then an equal volume of isopropanol was added, and the tube inverted 3-5 times then incubated for 30 minutes at -80ºC. After this incubation, the sample was centrifuges at 14 000 RPM for 5 mins and after decanting the supernatant, the pellet was washed with 70% ethanol after which it was dried at 35ºC in vacuum centrifuge (MiVac DNA concentrator, GeneVac). The pellet (plasmid DNA) was then resuspended in 30µl TE buffer.

2.1.3.3 DNA quantification

Quantification and purity of DNA were assessed using Thermoscientific Nanodrop 2000c spectrophotometer and for a more accurate measurement we used Quibit 30 flurometer from Invitrogen (Life technologies).

2.1.3.4 Plasmid and whole genome DNA sequencing

Sequencing of plasmids was performed by the Central Analytical Sequencing Facility (CAF) at the University of Stellenbosch- Cape Town. Sequencing reads were visualised using genome compiler software (https://designer.genomecompiler.com). WGS of isolates was done through Genohub service providers. Analysis of WGS was done by Dr Anastasia Koch at the University of Cape Town- MMRU research unit.

2.1.3.5 Polymerase chain reaction (PCR)

Polymerase chain reaction was used to amplify DNA fragments and as a screening method to reveal the presence of a DNA fragment. Roche's Fast start kit was used for screening PCRs. Further amplification of fragments for cloning purposes was carried out using either Phusion high fidelity DNA polymerase (Bio labs) or KOD hot start DNA polymerase (Novagen).

For fast start Taq DNA polymerase reactions, reactions were set up containing: 0.5 U Taq DNA, 200uM DNTPs, 1× buffer, 1× GC mix, 0.2 µM primers and approximately 1ng of DNA. The following cycling parameters were used: initial denaturation at 94^{o}C for 3 minutes, denaturation at 94^{o}C for 0.5 minutes, annealing for 0.5 minutes at a temperature which is dependent on the melting temperature of primers, extension at 72^{o}C for 1 minute depending on the size of fragment (1 minute per Kb DNA). Thirty cycles were carried out and the final extension was at 72^{o}C for 5 minutes.

For reactions with KOD hot start DNA polymerase, 50µl reactions were set up as follows; 5µl of 10× buffer, 3ul of 25mM MgSO4, 5ul of 10uM dNTPs, 1.5ul of primers at 10µM, 1µl of template DNA (up to 200ng), 1µl of KOD hot start DNA polymerase and finally 0.6% of DMSO. Thermocycling conditions were as follows; initial denaturation at 95^{o}C for 2minutes, denaturation at 95^{o}C for 20 seconds, annealing for 30 seconds at the temperature which was dependent on the melting temperature of primers (Rule: Primer Tm – (5–10) °C) extension at 70^{o}C for 15 seconds per kilo base pair, thirty cycles were carried out and the final extension was at 72^{o}C for 5 minutes.

Reactions with Phusion high fidelity DNA polymerase were carried out by mixing primers at a final concentration of 0.5 uM with template DNA at a final concentration between 30-100ng. A 2X Phusion master mix provided in the kit was then added at a final concentration of 1×. Finally, DMSO at a final concentration of 3% was added to the reaction mixture. Thermocycling conditions were as follows: preheat at 98^{o}C for 1.5 minutes, denaturation at 98^{o}C for 10 seconds, annealing for 20 seconds at 3 ^{o}C above the lowest primer melting temperature, and extension at 72^{o}C for 20 sec per kilo base pair and a final extension at 72^{o}C for 5minutes. Thirty cycles were carried out.

2.1.3.6 Restriction enzyme digestions

Restriction enzyme digestions were carried out in 30 lμ reaction volumes as follows; 10U of enzyme, 0.1 mg/ml of BSA, 1× recommended buffer and 1 μg DNA. Reactions were incubated for 2 hours in a 37ºC water bath.

Ligation reactions

Ligations of linear DNA were accomplished using T4 DNA ligase or Rapid DNA ligase kit. The amount of insert added in the reaction was dependent on the size of the vector and was calculated using the formula:

$$\frac{ng\ of\ vector \times kb\ size\ of\ insert}{kb\ size\ of\ vector} \times \frac{3}{1} = ng\ of\ insert$$

Note:3/1 is the molecular proportion of insert over vector.

Ligation was through sticky ends. When using T4 DNA ligase, a 10ul reaction volume was set up by adding 1 unit of ligase enzyme to 1× ligase buffer. The amount of vector and insert added were calculated and if necessary sterile water was used to make up the volume to 10ul. The reaction was then incubated at 16ºC overnight and then transformed.

Reactions with the Rapid DNA ligase kit were carried out in a final volume of 20ul by mixing 4 units of T4 DNA rapid ligase buffer with 1x T4 DNA ligase. Vector and insert amounts were calculated as above and the reaction volume were made up to 20ul with sterile water. The reaction was then incubated at room temperature for 5 minutes prior to transformation.

2.1.3.7 Electrophoresis

Agarose gel electrophoresis was used to separate DNA fragments, for plasmid size confirmation, restriction digestion screening and for assessment of PCR products. In this study 0.8% agarose gels were made in 1× Tris-acetate-EDTA buffer (TAE –40 mM Tris, 20 mM acetic acid and 1 mM EDTA, pH 8.3). 5ul of nucleic acid staining (Red safe nucleic acid stain) was used per 100ml of gel. Various sizes DNA ladders from ROCHE were used depending on the fragment size expected. The gel was electrophoresed between 80-100V and visualized under UV light.

2.1.4 Phenotypic characterization

2.1.4.1 Determination of minimum inhibitory concentration of drugs

MIC were be determined using resazurin microtiter assay (Palomino et al., 2002). This is a colorimetric assay which uses resazurin an oxidation-reduction indicator for drug susceptibility testing(Palomino et al., 2002).Resazurin is blue in color when in the oxidized state and changes to pink when reduced. This provide a simple measure of viability which can be differentiated visually (Palomino and Portaels, 1999).

MSM strains of interest were cultured to an OD_{600} of 0.6. Antibiotics of interest were prepared and diluted in the appropriate diluent to 2× the top concentration desired in the test. 50 µl of 7H9 media was dispensed into all the wells of the clear round bottom 96 well microtiter plate. 50 ul of the 2X antibiotic of interest was pipetted into wells of row 2 (see microtiter plate layout in **Figure 2.1**). Using a multichannel pipette, the antibiotic was mixed 3 times and 50 ul was transferred to row 3. This serial two-fold dilution was carried out until row 11. The 50 ul from row 11 was discarded. Bacterial cells at OD_{600} of 0.6 were diluted 1:1000 in 7H9 media to achieve an inoculum size of approximately 10^5 CFU/ml. 50 µl of this diluted inoculum was added to all the wells except wells in row 12 which received 50 ul of 7H9 media instead. Growth controls containing no antibiotic (row 1 which had bacterial cells with no antibiotics) and sterility controls without inoculation (Rows 12 which received media only) were also included (**Figure 2.1**). The microtiter plate was incubated for 2 days at 37ºC. After 2 days of incubation, 10µl of 0.01% resazurin solution was added to each well, incubated overnight at 37°C. The following day the plate was assessed for colour development. A change from blue to pink indicates reduction of resazurin and therefore bacterial growth. The MIC_{90} was defined as the lowest drug concentration that prevented this colour change (Palomino et al., 2002).

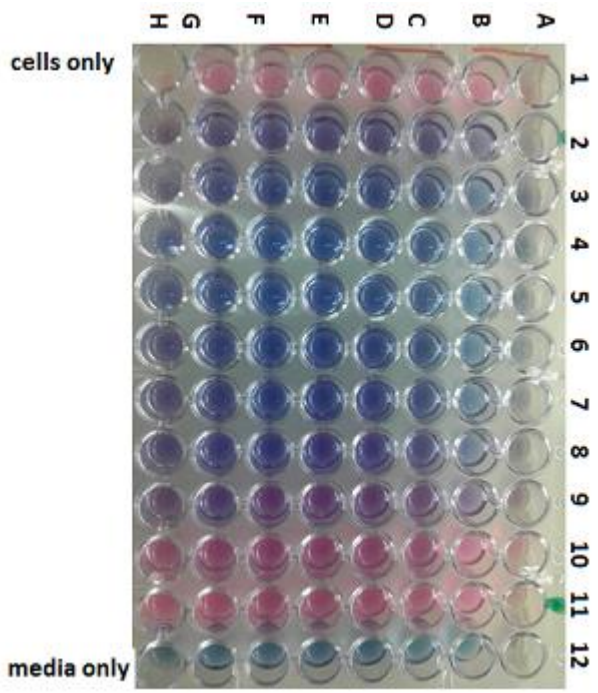


Figure 2. 1 Determination of minimum inhibitory concentration. Cultures were grown to mid-log phase (OD_{600} 0.6), diluted 1000× and 50 µl was added to wells containing two-fold dilution series of drug. After 3 days of incubation, 10 µl of 0.01% resazurin solution was added to each well, incubated overnight at 37°C, and assessed for color development. A change from blue to pink indicates reduction of resazurin and therefore bacterial growth. The MIC_{90} was defined as the lowest drug concentration that prevented this color change.

2.1.4.2 Damage Tolerance essays/Survival assays

This assay was carried out to find a concentration of an antibiotic which is high enough to induce genotoxic stress but also low enough to support growth at a very high inoculum. 10 ml cultures of MSM strain of interest were grown to an OD_{600} of 0.6 after which 10-fold serial dilutions were prepared and 10 µl of each dilution was spotted onto solid 7H10/OADC plates which were either untreated or contained mitomycin c or moxifloxacin. Treatment was carried out at 2×, 1×, ½ and a ¼ MIC specific for each strain for each antibiotic. Plates were incubated for 3–6 days at 37°C.

2.1.5 Generation of transposon libraries in *MSM*

2.1.5.1 Transposon library generation

MycoMarT7 phage stock was prepared according to a method described by Majumdar *et.al.2017*. The procedure for transducing *MSM* with MycoMarT7 was also adapted from Majumdar *et.al.2017*: 200 ml of *MSM* strain of interest was cultured to an OD_{600} of 0.8 - 1.0 in 7H9 media. The bacterial cultures were pelleted by spinning at 4000 x g for 10 minutes at 37ºC. The pellet was washed three times with an equal volume of prewarmed (37 ºC) MP buffer. Washing with MP buffer is critical because it removes Tween 80 which can prevent adsorption of phage to the bacterial cells if present (Groman and Bobb, 1955). The bacterial cells were then resuspended in 4 ml of MP buffer. 100 µl was taken from this 4ml suspension to be used for plating as negative control. The negative control bacterial cells were spread in 7H10 plates supplemented with 20 µg/ml KAN. The negative control had passed if we did not see any bacterial growth after 3 days of incubation.

For transduction, 2 ml of high titer phage stock (1× 10^11 pfu/ml) was added to the bacterial cells and the culture was incubated at 37ºC for 18 hours (overnight). After transduction, cells were pelleted and washed 3 times with PBS buffer supplemented with 0.1% tween and were then resuspended in 6ml of PBS-Tween 80. 100 µl of the suspension was plated out in serial dilutions on 7H10 medium with KAN to determine the titer of the library. The rest of the suspension was plated on 15 cm pre-warmed (37ºC) 7H10-OADC-Tween80 plates supplemented with 20 µg/ml KAN. 500 µl of cells were used per plate giving a total of 12 plates. Bacterial cells were spread using glass beads. For untreated libraries, plates were incubated for 2- 3 days.

After this incubation cells were harvested by scraping the colonies from plates and resuspending in 7H9-OADC-Tween 80 broth supplemented with 15% glycerol. The library was then frozen at -80 in 1 ml aliquots (Majumdar et al., 2017).

2.1.5.2 Determination of library titre

100 µl of culture was taken out after MSM was transduced with MycoMarT7 bacteriophage. Ten-fold serial dilutions were carried out in 7H9 liquid medium. Dilutions; 10^{-1}, 10^{-2}, 10^{-3} and 10^{-4} were done. 100 µL of each dilution was then plated in 7H10 medium supplemented with KAN. After 3 days of incubation, a plate which has the fewest number of colonies was chosen and the individual distinct colonies were counted.

To calculate library titres, we used the following formula:

(No of bacteria × dilution factor/ volume plaited in ml) × 10 ml),

where 10 ml was the total volume of the bacterial suspension after transduction. Only libraries with at least 100, 000 CFU were subjected to selection and analysis.

2.1.5.3 Tn library selection

The Tn libraries were stored in 1ml aliquots at -80^{o}C. Library selection was carried out by treating two million bacterial cells (for each library strain) with MMC and MOX. The antibiotics were used at ¼ MIC specific for each strain. Library selection/treatment was done by plating the bacteria on 7H10 media supplemented with MMC/MOX and 20 µg/ml KAN. For this experiment we first determined the number of bacteria in 1 ml aliquots by determining the titre per ml. An aliquot which is equivalent to a million cells was then taken out and resuspended in 2ml of 7H9 medium and then spread in a 245 × 245 mm plate of 7H10-OADC-Tween80 plates supplemented with 20 µg/ml KAN and an antibiotic of interest. Spreading was done with the help of beads. Treated libraries were incubated for 4-6 days because colonies took longer to appear on plates. After this incubation cells were harvested by scraping the colonies from plates and resuspending in 7H9-OADC-Tween 80 broth supplemented with 15% glycerol. The library was then frozen at -80 in 1 ml aliquots.

2.1.5.4 Library/sample preparation for sequencing

Only Libraries with titers of greater than 100,000 CFU were prepared for sequencing to ensure comprehensive coverage of the possible TA dinucleotide insertion sites in the MSM genome. The method for Library preparation for sequencing was adapted from Majumdar et al. DNA was extracted from the 1mL freezer stocks using the CTAB method described above. But since we were dealing with a very dense inoculum, we divided the 1ml inoculum in 250 aliquots for

easy extraction. The processes involved in library preparation for sequencing are explained below and summarised in a schematic shown in **Figure 2.2.**

2.1.5.4.1 Random DNA shearing

Random DNA sharing was carried out using dsFragmentase enzyme. In order to maximise yield, three reactions for each sample were carried out. For each reaction, 5 ug of library DNA was mixed with 3 µL of 10× "NEB next dsFragmentase" reaction buffer, 3 µL of "next dsFragmentase" enzyme and water up to final volume of 30 µL. The sample was then mixed by pipetting up and down before incubating for 35 minutes at 37 °C to generate fragments in the range of 50-200 bp. The reaction was stopped by adding 5 µl of 0.5 M EDTA. Complete shearing of DNA was confirmed by gel electrophoresis where a smear was observed. Purification of the reaction was carried out using QI quick PCR Purification Kit from QIAGEN (Majumdar et al., 2017). DNA quantification was done using Quibit 30 flurometer from Invitrogen (Life technologies).

2.1.5.4.2 End repair and A-tailing

1.8 ug of sheared DNA was used per end repair reaction. The DNA was mixed with 5 µL of 10× "End-it" DNA repair buffer, 5 µL dNTPs, 5 µL ATP, 1 µL of "End-it" enzyme and sterile distilled water up to a final volume of 50 µL. After mixing by pipetting up and down, the reaction was incubated at 37 °C for 45 min. QI quick PCR purification columns were then used to clean up the reaction. The reaction was eluted with 65 µl of sterile distilled water and was then directly used for A-tailing reaction. For A-tailing, 10 µL 10× Taq DNA polymerase buffer, 20 µL 10 mM dATP, and 5 µL Taq DNA polymerase were added to the 65 µl end-repaired DNA and incubated at 72 °C in a thermocycler. Purification was carried out with QIAquick PCR purification columns and eluted with 50 µl of sterile distilled water (Majumdar et al., 2017).

2.1.5.4.3 Adapter ligation

Sequencing adaptors were prepared by mixing 24 µL each of adapter 1 and 2 and then adding 2 µL of 50 mM MgCl2 and heating the mixture to 95 °C for 10 min in a thermocycler and lowly reducing the temperature to 20 °C over a period of 2 h with ramping at 1 %. 4 µL of this ligated barcode were then added to the 50 µL A-tailed DNA mixture together with 8 µL of T4 DNA ligase, 10 µL of the 10× ligase buffer, and sterile distilled water up to 100 µL final volume. The reaction was Incubated at 16 °C overnight yielding adapter ligated DNA. Reaction purification was carried out using QIAquick PCR purification kit as above but the columns were washed at least 5 times before elution with100 µL sterile distilled water (Majumdar et al., 2017). DNA concentration was determined using Quibit 30 flurometer from Invitrogen (Life technologies)

2.1.5.4.4 PCR Amplification of Tn-DNA Junction

To amplify Tn insertion junctions on the chromosome, 8 PCR reactions per samples were set up. 100ng of the adapter ligated DNA from the above reaction was used per reaction. 15 µl of 2× dream Taq master mix, adapter ligated DNA, 1.5 µl Tn primer, 1.5 µl adaptor primer and water up to 30 µl were mixed together and amplification was carried out using one cycle of 95 °C for 10 min, 20 cycles of 95 °C for 30 s, 58 °C for 30 s, 72 °C for 45s and a final extension cycle of 72 °C for 5 minutes. The 8 reactions were pooled and electrophoresed on a 2% gel. After which we cut out a band corresponding to fragments sizes of 300 to 500 bp. DNA was extracted from the gel using Qiagen gel extraction kit. After six washes, DNA was eluted using 50 µl of sterile distilled water (Majumdar et al., 2017).

2.1.5.4.5 Addition of adapters for multiplexing

This is the last PCR to be carried out before sequencing. Soapy adapter which is used to enable multiplexing and a Sol_MAR adapter which recognise the Tn sequence are integrated into 100 ng of gel extracted adapter ligated DNA from the above step. To do this we added 4 µL of the gel extracted DNA, 2 µL of 1 µM Sol-mar mix, 2 µL of 1 µM Sol-ap-tag mix, 10 µL of 2× dream Taq master mix and water up to 20 µl. Amplification was carried out using one cycle of 95 °C for 5 min and ten cycles of 95 °C for 30 s, 58 °C for 30 s, 72 °C for 45 s, and one final extension cycle of 72 °C for 5 min. DNA purification was carried out using QIAquick PCR

purification kit and the DNA concentration was determined using Quibit 30 flurometer from Invitrogen (Life technologies). 40 µL of at least 10ng/µL DNA was sent for illuminar sequencing (Majumdar et al., 2017).

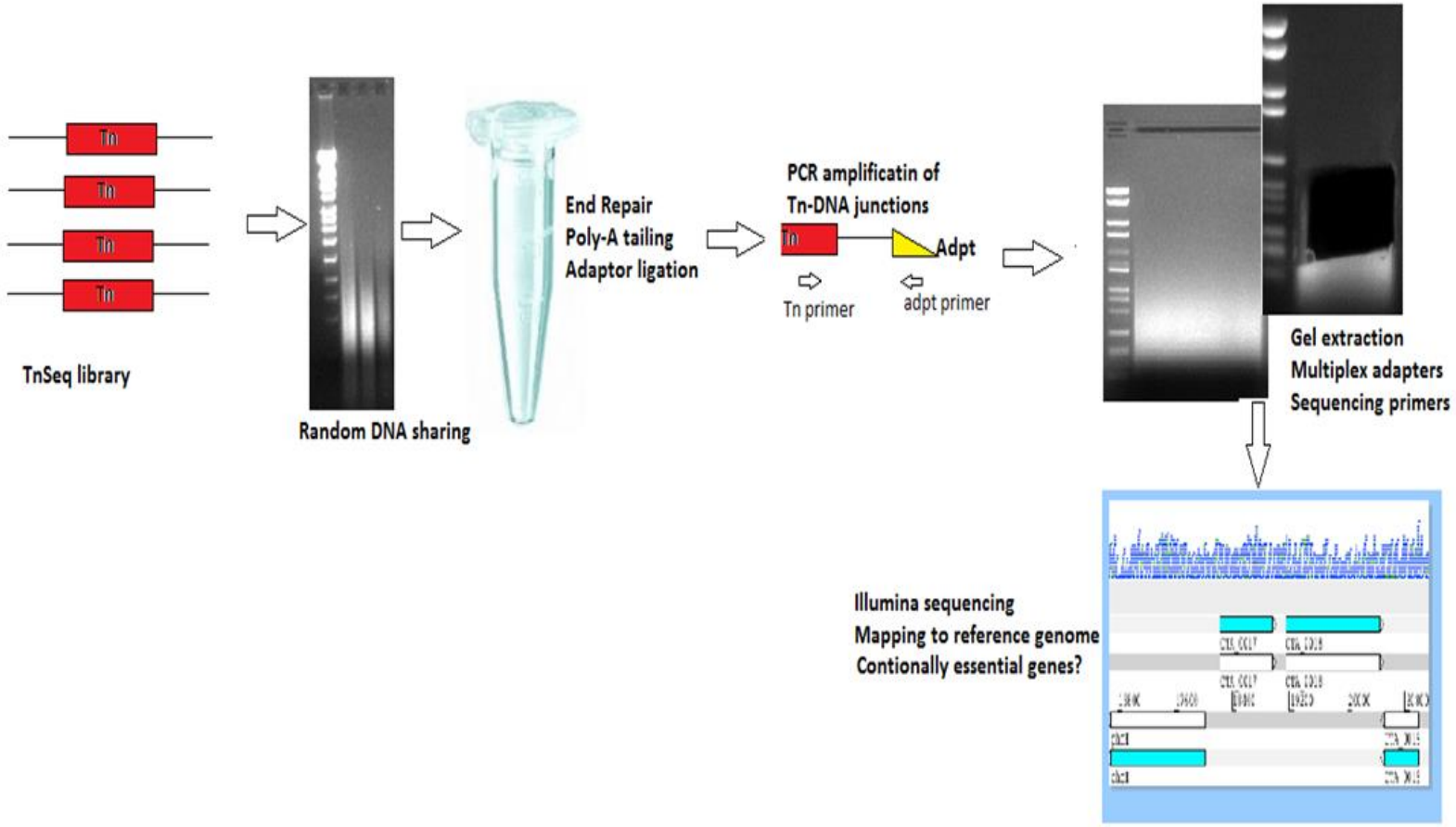


Figure 2. 2 A schematic diagram showing steps involved in TnSeq library generation

2.1.6 Data processing: statistical analysis of essentiality

2.1.6.1 Initial data processing: TRANSIT pre-processor

Prepared library replicates were sequenced on an Illumina MiSeq. At least five million 2 x 300 PE (Paired End) reads per sample were obtained. Raw sequence data were exported to fastq files for further analysis. Preliminary data analysis was carried out using TRANSIT pre-processor software (DeJesus et al., 2015). TRANSIT is a computer software used for analysing Himar1 TnSeq data. Analysis was carried out as recommended by instructions at; http://github.com/mad-lab/transit. In summary the tool extract read counts from raw sequences, maps them to the reference genome strain MSM mc^2155 and outputs them in wig format for loading into TRANSIT software. TRANSIT pre-processor (TPP) generates statistics for assessing library quality which helps in diagnosing potential problems in library preparation. Four key parameters are used to assess the quality of libraries: mapped reads, insertion density, no-zero sites mean (NZ- mean) and primer/vector matches (DeJesus, 2019).

There should be several million reads that map to the reference genome. If mapped reads are drastically low, firstly one should check if they loaded the correct reference genome. Secondly, this could indicate presence of non-bacterial DNA e.g. Tn vector contamination and primer dimers (DeJesus, 2019). Vector contamination can be limited by washing the library several times with PBS buffer supplemented with 0.1 % Tween 80 (Majumdar et al., 2017). Primer dimers can be reduced by improving fragment size selection during sample preparation. Good libraries should also have an NZ mean of at least >10 in regions with Tn insertions. But most importantly, the insertions density which represent the % of TA sites in no-essential sites should be at least >35% (DeJesus, 2019). In This study, data sets with insertion density of less than 30% were not analysed further.

2.1.6.2 Data analysis with TRANSIT: Quality control

Before data can be analysed with TRANSIT software, normalisation was carried out. Normalization ensures that sources of variability are not mistakenly treated as real differences in datasets. Trimmed total reads (TRR) is a recommended method of normalisation for most data sets. In this method, data is normalized by the total read-counts which ensures that datasets have the same mean over all counts. It also trims the top and bottom 5% of reads counts which

represent poorly sequences portions of the data. This method also normalises for differences in insertion density for replicates data sets and is especially important when comparing libraries with different insertion density (DeJesus, 2019). Example of TRR normalised and non-normalised data is shown in **Figure 2.3.**

Normalization: nonorm

File	Density	Mean Read	NZMean Read	NZMedian Read	Max Read	Total Reads	Skew
dnaE2_MMC_A.wig	38.8	20.3	52.3	26.0	29068	1575415	54.1

Normalization: TTR

File	Density	Mean Read	NZMean Read	NZMedian Read	Max Read	Total Reads	Skew
dnaE2_MMC_A.wig	38.8	150.9	389.3	193.6	216460	11731617	54.1

Figure 2. 3 TRR data normalization. Data is normalised by the total read counts which ensures that datasets have approximately the same mean count over all sites. TOP: non normalised data. BOTTOM: TRR normalised data. Figures were generated using TRANSIT software.

Normalisation of data through TRR does not correct skewed data sets. Skewed data sets do not fit closely to a geometric distribution. But rather have a few highly over-represented regions dominating the distribution (DeJesus and Ioerger, 2015) e.g. **Figure 2.4**. A beta geometric distribution equation can be used to correct for skewedness. This method can adjust the observed read counts so that they fit the geometric distribution line closely (e. g **Figure 2.4**). Since read counts are not truly geometric, a curvature is expected in the data, however strong curvatures from the geometric distribution line gives an indication of data skewedness (DeJesus and Ioerger, 2015).

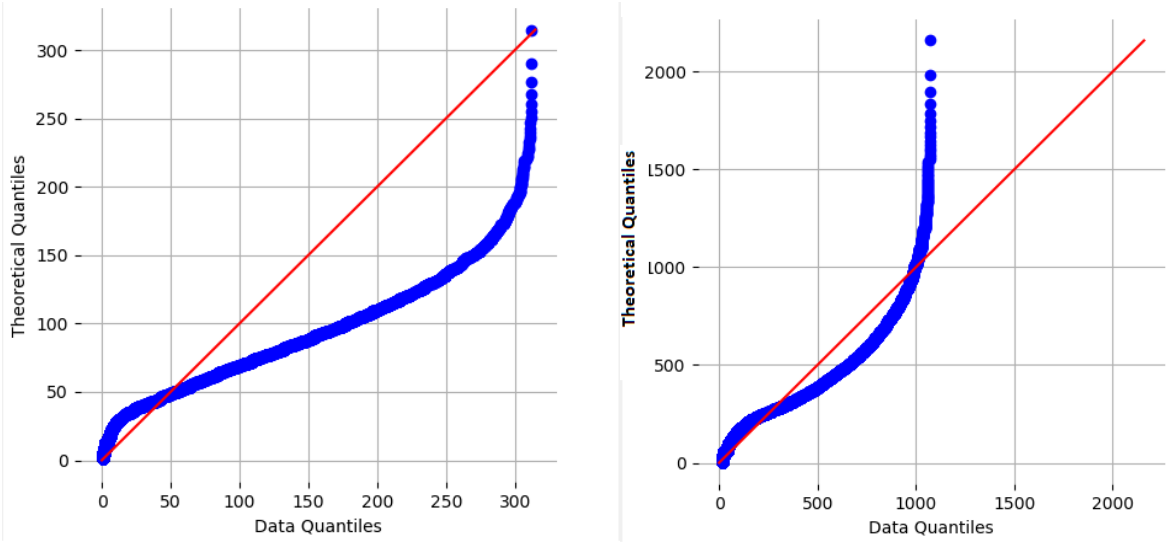


Figure 2. 4 QQ (quantile-quantile)-plot of non-zero read counts for dataset. (blue curvy line) versus the ideal geometric fit (Red diagonal line). LEFT: shows a skew which indicate lack of fit in the data represented. RIGHT: shows better fit after a beta geometric normalisation has been applied to the data Figures were generated using TRANSIT software.

2.1.6.3 Data analysis with TRANSIT: Hidden Markov Model (HMM).

Essential genes were determined by the Hidden Markov Model (HMM) tool on TRANSIT v2.1.2(DeJesus and Ioerger, 2013). This method analyses for Essential regions (ES), non-essential regions (NE), growth defect regions (GD) and growth advantaged regions (GA) in a single growth condition. The HMM is applied to the sequence of TA sites to obtain essentiality assignment based on the read count at the site and the distribution over the surrounding sites. A geometric distribution is used to determine the likelihood of read count in each state where the mean is near 0 for essential regions, near the global mean for non-essential regions, intermediate or particularly low for growth deficient regions and lastly regions with higher than average read counts are classified as growth advantaged regions (DeJesus and Ioerger, 2013). To do the analysis, all replicates Tn libraries for a single strain in one condition were loaded in a 'control' window and the HMM method was run using default settings.

2.1.6.4 Data analysis with TRANSIT: Resampling

Resampling is also a method embedded in the TRANSIT software (DeJesus et al., 2015). This method evaluates essentiality of genes between two conditions hence it analyses for conditional essentiality. In this method, read counts are summed over all sites to estimate mean read count

of every gene in each condition. The difference in mean counts between the two conditions is compared to a 'null' resampling distribution. Finally, a P-value is calculated. The 'null' resampling distribution comes from random reshuffling of counts at TA sites in the region of interest in all the data sets. True conditional essentiality is present when insertions occur in one condition but not the other, resulting in a difference that is much higher than the 'null' resampling distribution. Apart from identifying conditionally essential genes, this method also identifies genes whose disruption leads to reduced fitness (**Figure 2.5**). These are genes which have reads counts that are much lower (but are present) in one condition compared with the other (DeJesus et al., 2015). We used this method to compare untreated libraries with treated libraries for the same strain. In this analysis, untreated libraries were uploaded on TRANSIT interface as control samples while treated libraries were uploaded as experimental samples. We also used this method to compare WT untreated libraries with each untreated mutant strain library.

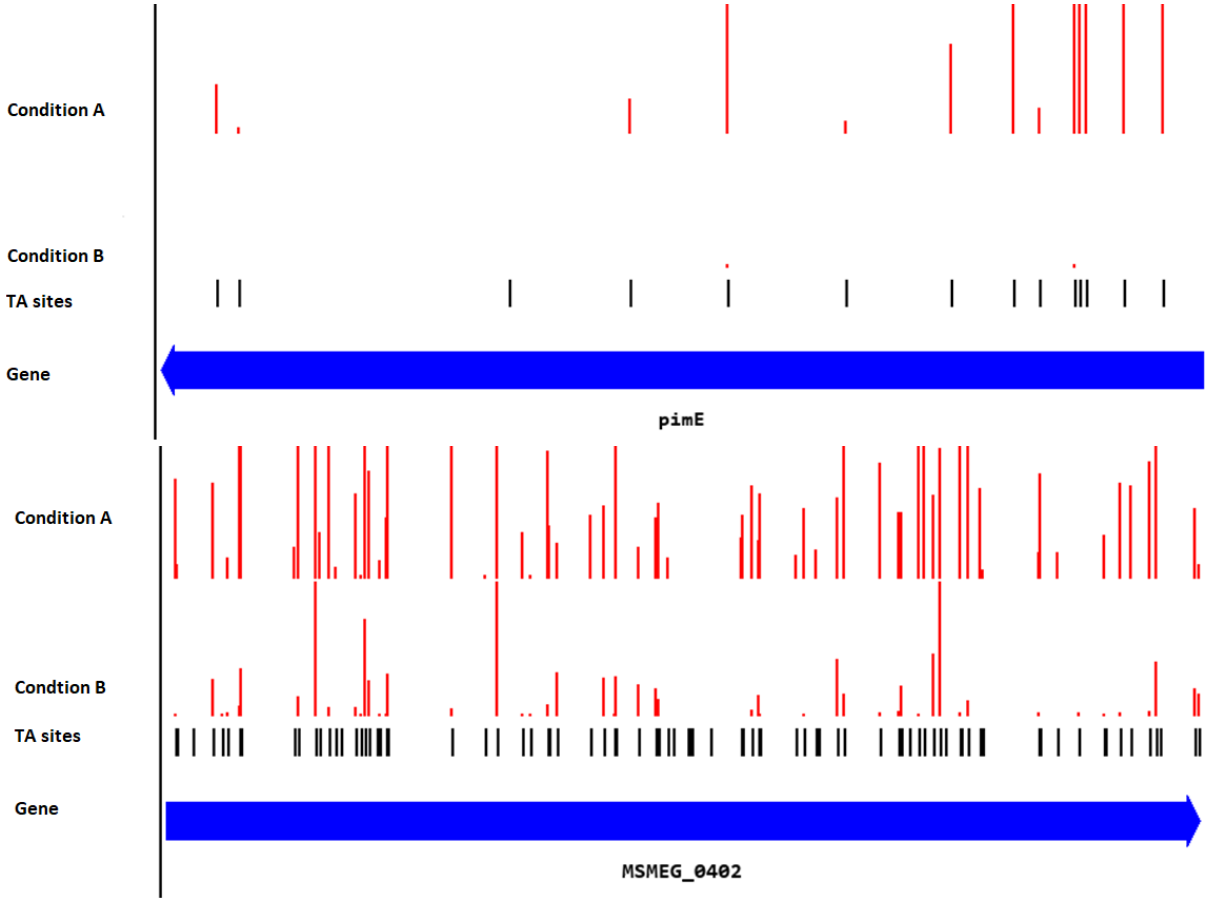


Figure 2. 5 Illustration of resampling permutation method. This method identifies true conditionally essential genes (TOP) and genes which show reduced fitness in one condition compared to the other (BOTTOM). The blue arrow depicts a gene of interest and the red horizontal lines represent Tn insertions and their magnitude in this gene**.**

2.1.6.5 Data analysis with TRANSIT: Genetic interactions (GI) analysis

Studying GI is important as it reveals networks of functionally related genes that work in redundant pathways and related pathways. TRANSIT software also has a built in bayesian statistical method for analysing genetic interactions. This method can compare Tn libraries generated in different genetic backgrounds. The analysis is a four way comparison of Tn insertion counts in order to identify Tn mutations that differentially affected bacterial fitness depending on the genetic background hence, genes of interest are those that show a statistically significant change in enrichment that does not only depend on the treatment condition but also depends on the difference in genetic background (DeJesus et al., 2017b). This analysis requires four-way library design. In this study, WT and mutant library in untreated conditions were compared with WT and mutant library in a treated condition. This analysis also generates a P-value to show statistical significance. For the analysis, the cut-off was a P value <=0.05. The software also categorises the GI into aggravating, alleviating and suppressive interactions. Aggravating interactions result from insertions that decrease fitness in the mutant background. Alleviating insertions do not affect the fitness of the bacteria in the mutant background even though the genes both individually decrease fitness of the bacterium. Lastly, suppressive insertions are those that reverses the fitness defect in the mutant background (DeJesus et al., 2017b). In this study, we identified aggravating GI.

2.1.7 TnSeq conditional essential genes validation using CRISPR interference platform

CRISPR (clustered regularly interspaced short palindromic repeats) Cas (CRISPR associated) system is part of the bacterial adaptive immunity. A widely used CRISPR/Cas system from *Streptococcus pyogenes* requires CRISPR associated protein 9 ($Cas9_{spy}$) and a single guide RNA (sgRNA). sgRNA consist of two parts being a CRISPR RNA (crRNA) and a trans-activating crRNA (tracr RNA). Target specificity is facilitated by base pairing of the sgRNA to DNA target as well as a protospacer adjacent motif (PAM) within the gene of interest. In the presence of sgRNA, Cas9 endonuclease homes in on the target sequence and makes a double-stranded cut directly upstream of the PAM sequence (Jinek et al., 2012). The Cas9spy/CRISPR has since been repurposed by inactivating the Cas9 nuclease activity. The inactivated Cas9 is now called dCas9. This system now termed CRISPRi for CRISPR interference, enables gene silencing of specific genes by blocking transcription (Qi et al., 2013).

Streptococcus pyogenes Cas9 based CRISPR system leads to poor knockdown efficiency in mycobacteria. Therefore, a plasmid based CRISPRi system that utilises a dCas9 from *Streptococcus thermophiles* has been developed(Rock et al., 2017). The single plasmid platform expresses the dCas9sth1 under the control of an anhydrotetracycline (ATc)- inducible promoter and the sgRNA under the control of a strong constitutive promoter (Rock et al., 2017).

2.1.7.1 sgRNA cloning protocol

A list of all possible sgRNAs in the MSM mc^2155 genome was obtained from Dr Jeremy Rock. We chose the sgRNA that corresponded with the lowest PAM score as suggested by De Wet et al., 2018. The lower the Pam score the more efficient is the knockdown generation (unpublished data, de Wet et al., 2018). Two oligos that make up the sgRNA were selected for each gene of interest. The oligo sequences were sent for synthesis with Inqaba biotec™
. To anneal the oligos, 4 µL of each oligo (100 uM each) was added to 42 µL of annealing buffer. The oligos were annealed using the following program: 95ºC for 2minutes, 0.1 ºC/sec to 25ºC. Ligation of the sgRNA to the MSM CRISPR vector was carried out by mixing together: 1 µl annealed oligos, 0.5 µl Bombil digested CRISPR vector, 1 µL T4 DNA ligase, 0.5 µL T4 DNA ligase buffer and 7 µL of dH2O. The ligation was carried out at 55ºC for 3hrs in a thermocycler.
Heat shock transformation of *E. coli* DH5alpha cells was then carried out by mixing 1 ul of the ligation mixture with 5 µL of competent cells. After transformation, cells were plated on LB agar supplemented with 50 µg/ml KAN. After overnight incubation a single clone was cultured overnight, and plasmid extracted and sent for sequencing. Sequencing was carried out by Stellenbosch Cas facility using primer 1834 (Rock et al., 2017). Plasmids that contained inserts were then used to electroporate MSM strains.

2.1.7.2 Electroporation of the CRISPRi system

Electroporation was carried out using Bio Rad gene pulser Celt time constant protocol. 1 mm brown cap cuvettes were used. 100ng of plasmid DNA was mixed with 100 µL of bacterial competent cells. Voltage was set to 1200 and time constant was set to 5 Ms. After electroporation the cells were rescued overnight in 600 ml of 7H9 broth. MSM CRISPRi plasmid, PLJR962 containing an *mmpL3* gene sgRNA was used as a positive control.

After overnight rescue in 7H9 broth, the cells were spun for 1 min and the pellet resuspended in 100 µL 7H9 medium and then cultured in 7H10/KAN agar plates and incubated for 3 days. After 3 days one colony was picked and sub-cultured in 7H9-KAN broth to an OD_{600} of 0.6-0.8. We initially did a ten-fold serial dilution of this 0.6 – 0.8 culture and plated the dilutions in 7H10 plates with or without ATc but in most of these plates the phenotype was not profound because the density of cells was high. We then decided to do a 1/1000 dilution of the 0.6 – 0.8 culture first then subjected it to 10-fold serial dilutions. 5 µL of the dilution series was then spotted on two plates of 7H10_KAN and 7H10_KAN_ATc in order to validate genes that were identified as essential in untreated conditions. To validate genes that were conditionally essential in treated conditions the drug being investigated was added in both plates while ATc was added in one plate as above. Plates were then incubated for 2-3 days for untreated libraries and 4-6 days for treated libraries. True conditional essentiality was validated when growth was observed in a plate which does not have ATc while little or no growth is observed in a plate with ATc.

Table 2. 2 CRISPRI sgRNAs used in this study

Target gene	Gene Name	Top Oligo	Bottom Oligo
Msmeg_1670	sdhA	GGGAGCGCTCGTTGCCGTGCTCGG	AAACCCGAGCACGGCAACGAGCGC
Msmeg_1672	sdhC	GGGAGCTGGTAGCGCGGGCCCTTGGC	AAACGCCAAGGGCCCGCGCTACCAGC
Msmeg_5780	-	GGGAGACACCGCGAAGCCGGACCG	AAACCGGTCCGGCTTCGCGGTGTC
Msmeg_0840	-	GGGAGCTCGAATCCATCAGGTGTGA	AAACTCACACCTGATGGATTCGAGC
Msmeg_0866	-	GGGAAACAGGTTCGGCACCGACGG	AAACCCGTCGGTGCCGAACCTGTT
Msmeg_2868	-	GGGAATCTGGCTGTACGCGGGGCTGGA	AAACTCCAGCCCCGCGTACAGCCAGAT
Msmeg_5149	pimE	GGGAGCGGCCCAGCGGCGCGCCCC	AAACGGGGCGCGCCGCTGGGCCGC
Msmeg_1522	rpsK	GGGAGCCCAGGCGATGACGTTGCCC	AAACGGGCAACGTCATCGCCTGGGC
Msmeg_0402	-	GGGAGTCTTGAGCACCGAGAGGATCG	AAACCGATCCTCTCGGTGCTCAAGAC
Msmeg_0389		GGGAGCGAGCCGCATACGCCGGAT	AAACATCCGGCGTATGCGGCTCGC
Msmeg_0400	-	GGGAAGTTCCTCGCCCGCACCGAT	AAACATCGGTGCGGGCGAGGAACT
Msmeg_0688		GGGAACGTCGTCGACGTCGAACTT	AAACAAGTTCGACGTCGACGACGT
Msmeg_4259	-	GGGAACGACGGGCGGCGACCAGTT	AAACAACTGGTCGCCGCCCGTCGT
Msmeg_5392	kdpA	GGGAATCCCCGCGGTGGTGGGGGACA	AAACTGTCCCCCACCACCGCGGGGAT
Msmeg_0031	-	GGGAGCCGCGCGGGTCGCGGCCGGT	AAACACCGGCCGCGACCCGCGCGGC

Msmeg_4853	-	GGGAGCCCCCGGCCGCCTCGGCTT	AAACAAGCCGAGGCGGCCGGGGGC
Msneg_3743	soJ	GGGAAGCAGCGCCAGGCCGCGTAG	AAACCTACGCGGCCTGGCGCTGCT
Msmeg_2683	thyX	GGGAGCGCTGCGACAGCTGCGAGTA	AAACTACTCGCAGCTGTCGCAGCGC

Chapter 3:

Predicting gene essentiality in *Mycobacterium smegmatis*: Using TnSeq to derive essential gene lists for wild-type, Δ*dnaE2* and Δ*uvrB* strains

3.1 Introduction

MSM, a non-pathogenic saprophyte, (Bercovier and Vincent, 2001) is commonly used as a mycobacterial model (Shiloh and Champion, 2010). The characteristics of MSM that makes it an attractive experimental organism include the fact that it is fast-growing, with a doubling time of approximately 3-4 hours (colonies on solid agar are observed within 2-3 days, in contrast with MTB which doubles every ~22 hours, producing discernible colonies only after 3-4 weeks); MSM is also a biosafety level 2 organism hence there is no need for extensive training and accidental exposure carries no risk(Shiloh and Champion, 2010). While there is utility in using MSM to understand MTB, the non-pathogen is a distinct mycobacterial species with a specific lifestyle and environmental adaptations. For example, succinate dehydrogenase has been found to be crucial for the survival of MSM under *in vitro* conditions while this pathway seems to be dispensable for growth of MTB (Pecsi et al., 2014). Similarly, a study of MTB antibiotic susceptibility, growth and gene expression during phosphate limitation (Glover et al.,2007) discovered that RegX3, part of the two-component SenX3-RegX3 which regulates phosphate dependent gene expression, is essential for *in vitro* growth of MSM but not for MTB (Glover et al., 2007).

. On initiation of this doctoral research, no essential gene lists were available for MSM: the first peer-reviewed genome-wide MSM essentiality study was released in 2019, during the course of this study (Dragset et al., 2019); therefore, establishing a genome-wide essential gene list in MSM was an important first step to aid in data interpretation. In addition, inferred WT essentialities were compared with equivalent lists derived for the mutant strains, Δ*uvrB* and Δ*dnaE2* to establish differences in gene essentiality between these strains under standard, untreated *in vitro* conditions.

The NER pathway functions to protect cells against UV radiation-induced DNA damage or DNA damage induced by oxidative metabolism or other endogenous processes which cause bulky DNA distorting lesions (Kraemer et al., 2007). In both prokaryotes and eukaryotes, NER is generally an error free process involving sequential steps towards DNA damage repair (Spivak, 2016). It starts with damage recognition, followed by excision and removal of the damage containing oligonucleotide and, finally, the filling of the resulting gap (Dupuy. A, 2015). The NER has two sub-pathways; the global genomic repair (GGR) and the transcription coupled repair (TCR) sub-pathways (Spivak, 2016). The difference in the two NER sub-pathways is in the first step of recognizing lesions: lesions throughout the genome are detected

by GGR whereas TCR is activated by blockages in transcription (Spivak, 2016). These pathways are present in both bacteria and human cells (Spivak, 2016).

TLS allows for the continuation of DNA replication when replication-blocking lesions escape the cell's DNA repair processes (Goodman, 2002). During TLS, specialist non-replicative DNA polymerases are used to bypass base damaged sites. These specialized polymerases, which are present in both prokaryotes and eukaryotes (Goodman, 2002), are able to copy DNA templates across lesions, following which the replicative polymerases resume synthesis (Davis and Forse, 2009). Most TLS polymerases lack intrinsic proofreading activity and possess large active sites to accommodate bulky lesions; therefore, they often have reduced fidelity on non-cognate DNA templates (Goodman and Woodgate, 2013). Owing to their ability to cause mutations, TLS polymerases have been associated with antibiotic resistance, adaptation of microbes to different environments (Scotland et al., 2015) as well as genome instability and cancer in eukaryotes (Haynes et al., 2015).

Previous studies have consistently implicated NER and DnaE2-dependent TLS as "dominant" DNA repair and damage tolerance systems in mycobacteria (Warner, 2010, Boshoff et al., 2003, Darwin and Nathan, 2005). To uncover the potential contributions of other "cryptic" DNA repair pathways, we set to identify conditional gene essentialities in engineered MSM mutants lacking either pathway. To this end, we applied whole-genome Tn mutagenesis to *uvrB* and *dnaE2* knockout strains under standard and genotoxic conditions. Tns are discrete DNA segments with the capacity to move ("jump") from one locus in a genome to another (Fedoroff, 2012). In cells, Tns are not always "jumping". Instead, several mechanisms exist to silence transposition (Friedli and Trono, 2015). These include mutations that affect their ability to move, defense mechanisms of the host which include DNA methylation, or the use of small interfering RNAs in eukaryotes (Friedli and Trono, 2015). However, Tns can be made to jump *in vitro*. In its simplest form, *in vitro* transposition requires Tn terminal repeats, transposase, target DNA and a buffer (Hayes, 2003). Tns are therefore effective mutagens and numerous transposon-based molecular methods have been developed to harness this property.

TnSeq involves genome wide transposon mutagenesis and deep sequencing. It is a powerful tool for discovering essentiality of genomic regions in bacteria (Chao et al., 2016). The basic principle of this method entails creating a Tn mutant library which, by definition, does not contain disabling disruptions of genes required for growth (Chao et al., 2016). *Himar1* mariner Tn is generally believed to insert with equal probabilities at all TA- dinucleotides (Lampe et

al., 1998) therefore its absence in a mutant library is taken to indicate biological selection against the corresponding mutant (Chao et al., 2016). Even though the *Himar1* marine transposons are believed to insert only at TA dinucleotides, it has since been found that a small proportion of TA dinucleotides are less permissible for insertion (DeJesus et al., 2017a). A sequence bias of the *Himar1* element rendered approximately 10% TA dinucleotide insertion sites impermissible for insertion in MTB. This small proportion of less permissible sites has been observed in other bacteria also, and hence, indiscriminate inclusion of these nonpermissive sites in TnSeq analyses could slightly inflate the number of predicted essential regions (DeJesus et al., 2017a). Nonetheless, TnSeq is based on the assumption that transposons insert randomly and quantification of Tn junctions is then carried out through deep sequencing (Chao et al., 2016).The relative abundance of each mutant correspond to the fitness of the mutant in that particular condition (van Opijnen and Camilli, 2013).

3.2 Aims and objectives.

The overall aim of this PhD work was to investigate cryptic repair DNA pathways that might compensate for absence of the dominant TLS system, DnaE2-mediated damage tolerance, and NER. This chapter aimed specifically to establish baseline essential gene sets of WT MSM mc^2155 in untreated conditions and to compare the list of essential genes with equivalent lists derived for Δ*uvrB* and Δ*dnaE2* mutants.

The specific objectives were:

- To validate the genotypes of Δ*dnaE2* and Δ*uvrB* using PCR and whole-genome sequencing.
- To generate whole-genome Tn libraries in WT, Δ*uvrB* and Δ*dnaE2* backgrounds under optimal growth conditions.
- To generate comparative essential gene lists for the strains
- To validate selected Tn essentials using CRISPRi

3.3 Results

3.3.1 TnSeq library generation

To identify genes that are essential for optimal growth of MSM *in vitro*, we constructed libraries of Tn insertion mutants using a temperature sensitive ΦMycoMarT7 vector carrying the *Himar1* transposon (Long et al., 2015). The amount of phage and bacterial cells used are shown in **Table 3.1**. At the outset, It was not easy to get the recommended 100,000 CFU required according to the published protocol (Majumdar et al., 2017). In order to rectify this problem, the number of phage used to infect the cells was increased from the recommended titer of 1×10^{11} pfu/ml to 9×10^{13} pfu/ml (**Table 3.1**). Even though this enabled us to achieve the recommended CFUs, upon sequencing, very high phage contamination of this library was observed (**Table 3.2**). To address this problem, log-phase cultures were infected overnight with lower phage titers of 2×10^{9} (**Table 3.1**). This adaptation of the protocol resulted in high library titers (**Table 3.1**) which had low phage contamination (**Table 3.2**).

Table 3. 1 Tn library generation

Tn library	Phage titer (pfu/ml	Library titer (CFU/ml)	Transduction time	OD_{600} of culture	Volume of culture (ml)
WT1	9×10^{13}	1×10^{5}	7 h	2.0	100
WT2	2×10^{9}	$>1.2\times10^{5}$	overnight	0.6-0.8	200
WT3	2×10^{9}	$>1.2\times10^{5}$	overnight	0.6-0.8	200

Notes:

1. 200ml of 0.6-0.8 OD_{600} cultures were spun and infected with ΦMycoMarT7 vector carrying the *Himar1* transposon.
2. Ten-fold serial dilutions of the resulting library were made to determine the library titer. Each library titer is the total number of CFU harvested from 12 square petri dishes (120mm×120mm).
3. pfu- plaque forming units; CFU- colony forming units

3.3.2 TRANSIT pre-processing (TPP) results

Library preparation for sequencing after library generation includes isolation of DNA, fragmentation of DNA, end-repair, addition of sequencing adapters, and enrichment of transposon adjacent regions before subjecting the libraries to Illumina deep sequencing (Majumdar et al., 2017)

Prepared libraries were sequenced on the Illumina MiSeq v3 platform through Genohub Inc services. Illumina indexing barcodes were used to facilitate multiplexing of the libraries for sequencing. At least five million 2 x 300 (Paired-End) reads per sample were obtained. Statistical analysis was carried out using the TRANSIT software Tool for Himar1 TnSeq Analysis (DeJesus et al., 2015). This software facilitates mapping (aligning) of datasets generated from Illumina sequencing (that have the terminus of the Tn as a prefix and a genomic suffix) to the genome to identify which TA sites they present. To facilitate use of this software, Dr. Thomas R. Ioerger (Texas A&M University, USA) developed a TPP software which extracts read-counts from raw sequences, maps them to the MSM mc^2155 reference genome and outputs them in wig format for loading into TRANSIT (DeJesus et al., 2015). **Table 3.2** summarizes the statistics that were generated on the libraries after TPP analysis.

According to DeJesus et al. (2015), a good library should have a genome saturation of >50%. Analysis data through the TRANSIT software also requires the data to be normalized. The default normalization is the NZ mean method which normalizes datasets to have the same mean over non-zero sites. A good library should have an NZ mean of >10 (DeJesus et al., 2015). While WT2 and WT3 libraries met these criteria, library 1 had low genome coverage. It can also be noted from the results that the number of reads that mapped to the genome were very low in WT1 and WT2 libraries (**Table 3.2**). Using low phage concentrations and transducing overnight appeared beneficial as it generated libraries that were more complex and with low phage/vector contamination (**Tables 3.1 and 3.2**). While library densities rarely reach 100%, our pooled library, comprising 3 biological replicates (WT_3BIOREP), show good genome coverage of 76.2% (**Table 3.2**). This ensures statistical confidence: predictions of essentiality in high-density libraries are more reliable because TA sites missing insertions most likely arise from selection rather than chance (DeJesus et al., 2015).

Table 3. 2 A summary of TPP results for WT Tn libraries

Tn library	Total reads	Mapped reads	No. TAs hit	% genome coverage	NZ mean	Phage Contam. (%).
WT1	4 790 307	82 594	24 682	31.7	3.3	46.5
WT2	4 194 106	988 467	53 749	69.1	18.0	0.8
WT3	37 540 565	20 048 858	43 771	56.3	61.0	12.3
WT_3BIOREP	46 524 836	21 119 053	59 221	76.2	62.8	14.8
Notes: WT_3BIOREP is a pooled library of all the 3 biological replicates						

3.3.3 Genome Essentiality Analysis

Essentiality analysis was performed to determine primary functions necessary for growth under specific *in vitro* conditions. Essential genes were determined using the HMM(DeJesus and Ioerger, 2013). This model uses a geometric distribution to determine the likelihood of a read count in each state, where the mean is near 0 for ES regions, near the global mean for NE regions, and intermediate or low for GD regions, with regions with higher than average read-counts classified as GA regions (DeJesus and Ioerger, 2013). A total of 451 genes were required for optimal *in vitro* growth of MSM (Supplementary data 1). In this study, genes required for optimal growth were defined as including all those classified as ES and GD by HMM analysis. **Figure 3.1** shows the HMM results for the pooled WT library (WT_3BIOREP). Based on a predicted 6,791 open reading frames in MSM (Mohan et al., 2015), these Tn data indicated that approximately 7% and 92% of MSM genes were ES and NE respectively. This contrasts with estimates that ~15%of prokaryotic genomes comprise essential genes with the rest of the genome being non-essential (Koonin, 2003). The MSM results also differ from MTB essential gene analyses which predict ~15% genome essentiality (DeJesus et al., 2017a). The results therefore indicate that unlike other bacteria, MSM encodes a smaller proportion of essential genes under the conditions tested, where only 7% of the genome is required for optimal growth.

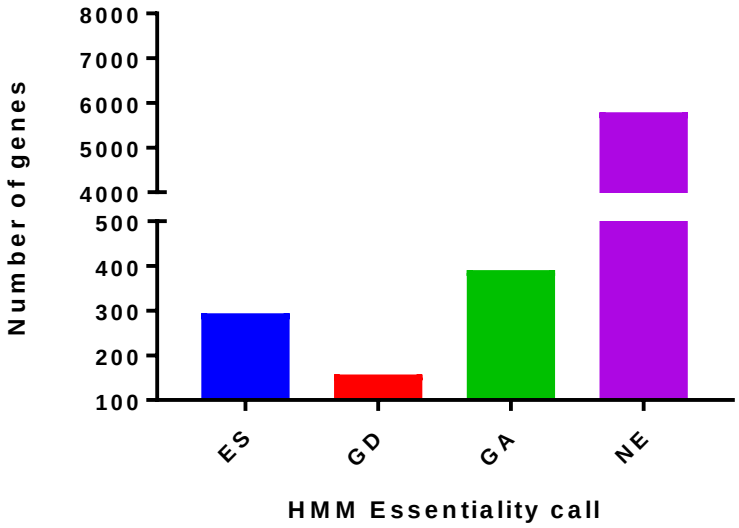
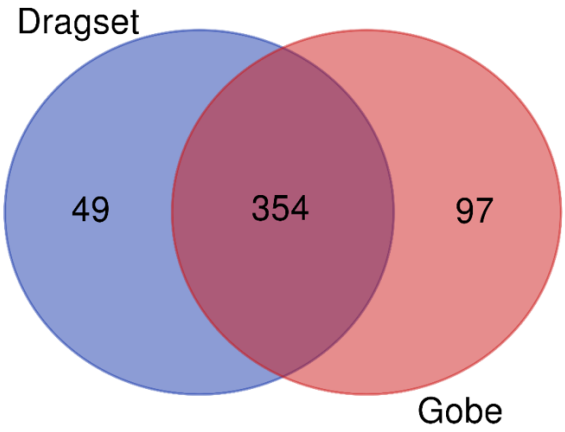


Figure 3. 1 Essentiality analysis of MSM WT genes using the HMM model. Altogether 451 (both ES and GD) genes were required for optimal growth of MSM *in-vitro*.

3.3.4 Comparison of essentiality data from different laboratories

During the course of this study, an essential gene list of MSM mc^{2}155 by TnSeq was published by another research group (Dragset et al., 2019). Our research group also carried out essentiality analysis of MSM genes using CRISPRi method (de Wet et al., 2018). This gave us an opportunity to compare our data not only with the published TnSeq data but also with CRISPRi data set. The comparison was to find out if our results are comparable with those from other laboratories and if different methodologies can yield comparable essential gene sets. We hypothesized that the majority of genes identified as essential in this study would also be essential in other previously reported libraries given that the strain used (MSM mc^{2}155) and the culture conditions (Middlebrook 7H10 agar supplemented with OADC) were the same

Indeed, we found a strong overlap between our data and Dragset et al., data (**Figure 3.2**). Almost 80% of genes essential in this study were also essential in Dragset et al., data. However, our library had more GD genes compared to the Dragset data (**Figure 3.2**). These accounted for more than half of the genes that were specific only to our data.



HMM essentiality call in Gobe and Dragset et al., 2019 Tn libraries

Call	Gobe	Dragset
ES	294	306
GD	157	97
GA	391	415
NE	5791	5815
NA	83	83

Figure 3. 2 Comparison of Gobe (this study) and Dragset et al WT TnSeq data. Genes that were identified as essential for optimal growth (both ES and GD) in the two libraries are compared to see if there are any differentially essential genes.

3.3.5 Comparing TnSeq essentiality data with CRISPRi data set

Parallel work in my host laboratory applied CRISPRi to investigate genome-wide essentiality in MSM (de Wet et al., 2018), focusing specifically on the essentialities of MSM genes with orthologs in MTB. The results of the CRISPRi analysis provided an opportunity to compare the TnSeq data generated in this study with published TnSeq data (Dragset et al., 2019) as well as the CRISPRi dataset. There were 451, 403 and 459 essential genes identified in this work ("Gobe"), the TnSeq gene essentiality paper ("Dragset") and the CRISPRi datasets, respectively. The comparison yielded 277 genes which were common to all three libraries (**Figure 3.3**), referred to hereafter as "core MSM essential genes" (Supplementary data 1). This number represents approximately 60% of all essential genes in the three libraries. We noticed that many genes were unique to the CRISPRi analysis, indicating that some genes which were not predicted to be essential by TnSeq were identified using CRISPRi. These included the F_0F_1 ATP synthase genes which are duplicated in MSM (Tran and Cook, 2005), the ribonucleotide-diphosphate reductase subunit beta (*nrdF2*) which is contained within a duplicated chromosomal region (Warner et al., 2006), and the potassium-transporting ATPase subunit A (*kdpA*) and subunit B (*kdpB*) which are located on an operon (Ali et al., 2017)(**Table 3.5**). The "Gobe" TnSeq dataset contained 84 unique genes (**Table 3.3**), including riboflavin synthase subunit alpha (*ribE*), cobalamin biosynthesis protein (*cobD*), and ATP-dependent DNA ligase (*ligB*). We also identified 28 genes which were unique to "Dragset" (**Table 3.4**), including the phosphate transport system regulatory protein *phoU* and the ribonuclease P protein component, *rnpA*.

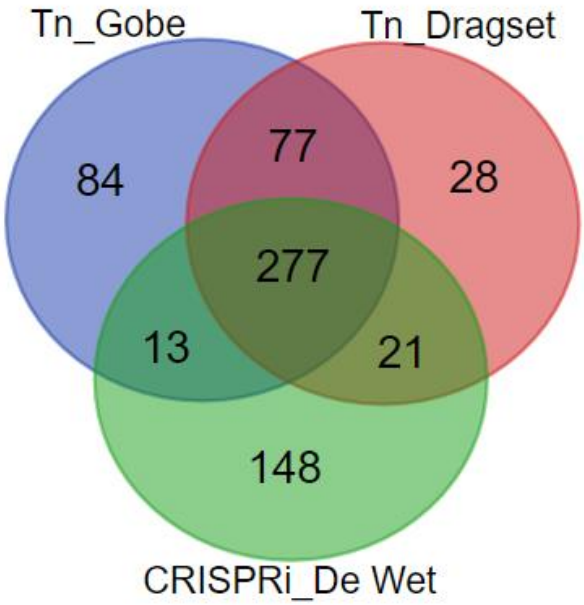


Figure 3. 3 Comparison of TnSeq and CRISPRi predictions of essential genes in WT MSM. Data are from this study (Tn_Gobe), Dragset *et al*., 2019 (Tn_Dragset) and De Wet *et al*., 2018 (CRISPRi_ De Wet).

Table 3. 3 Essential genes that are unique to the "Gobe" WT TnSeq data. The 'Gobe' data was compared to Dragset et al., 2019 TnSeq data and CRISPRi data. Only genes with annotated functions are shown out of 84 genes which were unique to this data set. The complete list is presented in Supplementary data 1.

Gene name	Description
ribE	riboflavin synthase subunit alpha
mbtD	polyketide synthetase
fabG	3-oxoacyl-(acyl-carrier-protein) reductase
grpE	co-chaperone, GrpE
thiS	sulphur carrier protein, ThiS
thiC	thiamine biosynthesis protein, ThiC
cofD	LPPG:FO 2-phospho-L-lactate transferase
nuoN	NADH dehydrogenase subunit N
hypB	hydrogenase accessory protein, HypB
hypD	hydrogenase expression/formation protein, HypD
hypE	hydrogenase expression/formation protein, HypE
ligB	ATP-dependent DNA ligase
cobD	cobalamin biosynthesis protein
speB	agmatinase

Table 3. 4 Essential genes unique to Dragset et al WT TnSeq data. The Dragset et al data was compared to Gobe TnSeq data and CRISPRi data. Only genes with annotated functions are shown out of 28 genes which were unique to this data set. The complete data set is presented in Supplementary data 1.

Gene name	Description
etfA	electron transfer flavoprotein, alpha subunit
asnB	asparagine synthase (glutamine-hydrolyzing)
rplU	50S ribosomal protein L21
alr	alanine racemase
phoU	phosphate transport system regulatory protein, PhoU
thyX	FAD-dependent thymidylate synthase
rnpA	ribonuclease P protein component

Table 3. 5 Essential genes that are unique to CRISPRi data. The CRISPRi data was compared to Dragset and Gobe TnSeq data. Only genes with annotated functions are shown out of the 148 genes which were unique to this data set. The complete data set is presented in supplementary data 1.

Gene name	Description
psd	phosphatidylserine decarboxylase
senX3	sensor histidine kinase, SenX3
ftsE	cell division ATP-binding protein, FtsE
nrdF2	ribonucleotide-diphosphate reductase subunit beta
rnc	ribonuclease III
ftsY	signal recognition particle-docking protein, FtsY
ispG	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase
rpsO	30S ribosomal protein S15
dapB	dihydrodipicolinate reductase
miaA	tRNA delta (2)-isopentenylpyrophosphate transferase
dapB	diaminopimelate epimerase
pyrB	aspartate carbamoyltransferase catalytic subunit
pyrC	dihydroorotase
secG	preprotein translocase subunit SecG
cyoE	protoheme IX farnesyltransferase
hisF	imidazole glycerol phosphate synthase subunit HisF
hisI	phosphoribosyl-AMP cyclohydrolase
rpmI	50S ribosomal protein L35
cysS	cysteinyl-tRNA synthetase
mraZ	cell division protein MraZ
dnaJ	chaperone protein DnaJ
hrcA	heat-inducible transcription repressor
atpC	F_0F_1 ATP synthase subunit epsilon
atpD	F_0F_1 ATP synthase subunit beta
atpG	F_0F_1 ATP synthase subunit gamma

atpA	F_0F_1 ATP synthase subunit alpha
atpE	F_0F_1 ATP synthase subunit C
atpB	F_0F_1 ATP synthase subunit A
cysNC	bifunctional sulfate adenylyltransferase subunit 1
cysD	sulfate adenylyltransferase subunit 2
fumC	fumarate hydratase
kdpA	potassium-transporting ATPase subunit A
ispE	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase
rpmF	50S ribosomal protein L32
sucD	succinyl-CoA synthetase subunit alpha
sucC	succinyl-CoA synthetase subunit beta
gltA	type II citrate synthase
rpsF	30S ribosomal protein S6
serA	D-3-phosphoglycerate dehydrogenase
rpmE	ribosomal protein L31
mraW	*S*-adenosyl-methyltransferase, MraW
proC	pyrroline-5-carboxylate reductase
murA	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
pgsA	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase
trxB	thioredoxin-disulfide reductase
rluB	ribosomal large subunit pseudouridine synthase B
nusB	transcription antitermination factor, NusB
ppa	inorganic pyrophosphatase
kdpB	K+-transporting ATPase, B subunit
fmt	methionyl-tRNA formyltransferase
rplL	ribosomal protein L7/L12
rpmA	ribosomal protein L27
galU	UTP-glucose-1-phosphate uridylyltransferase
ffh	signal recognition particle protein
serB	phosphoserine phosphatase
glpX	fructose-1,6-bisphosphatase, class II
gatC	glutamyl-tRNA (Gln) amidotransferase, C subunit
hot	hypoxanthine phosphoribosyltransferase
purN	phosphoribosylglycinamide formyltransferase

3.3.6 Functional classification of the ‘core’ MSM essential gene set

To understand the functions of the 277 ‘core’ MSM essential genes, functional classification analysis was performed using Mycobrowser (https://mycobrowser.epfl.ch/). The genes clustered into 3 major functional classes – intermediary metabolism/respiration, information

pathways and cell wall/cell processes, with the largest number of genes falling into the metabolism/respiration group (**Figure 3.4**).

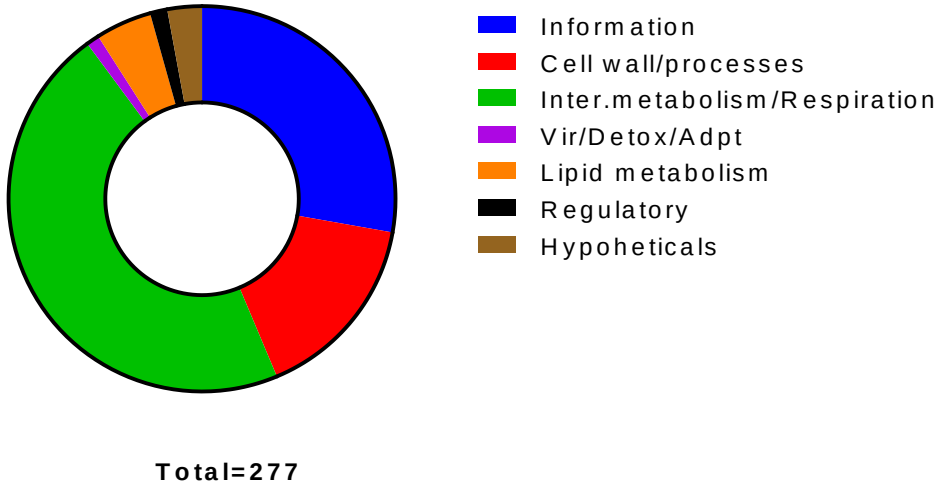


Figure 3. 4 Functional classification of core MSM essential genes (both ES and GD). MTB homologs of all the essential genes were first identified using Mycobrowser (https://mycobrowser.epfl.ch/). All 277 genes possessed MTB orthologs.

Since the intermediary metabolism/respiration together with the information pathways categories represent a large proportion of the essential gene list, we sought to subdivide these groups further to obtain greater insight into the biological processes represented (**Figures 3.5 and 3.6**). This analysis showed that genes involved in amino acid biosynthesis were abundant. These were followed (in order of overrepresentation) by genes involved in the TCA cycle and electron transport chain (TCA/ETC), and then those that are involved in nucleotide biosynthesis (**Figure 3.5**). Subdivision of the information pathway group, on the other hand, revealed that most of these genes were involved in translation (**Figure 3.6**).

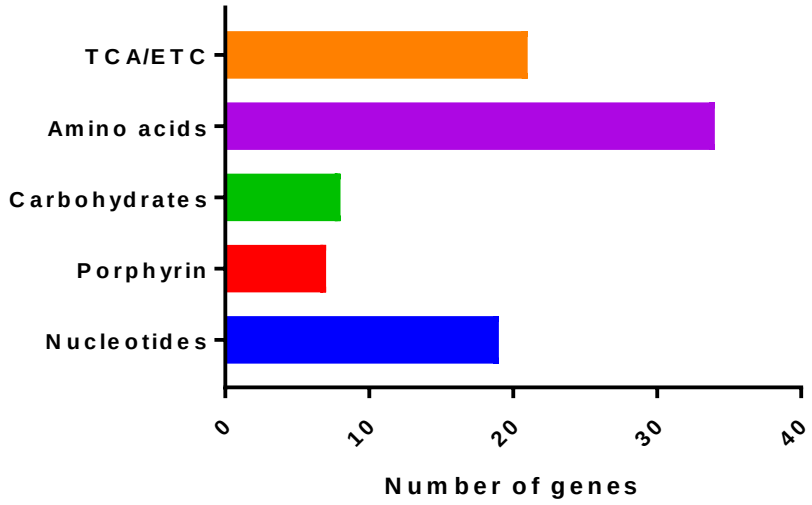


Figure 3. 5 Classification of essential genes that belong to the *Intermediary Metabolism and Respiration* group. MTB orthologs of all genes were first identified using Mycobrowser, with functions inferred from data contained in the Universal Protein resource (UniProt, © 2002 – 2019 UniProt Consortium).TCA/ETC: Tricarboxylic acid cycle/electron transport chain.

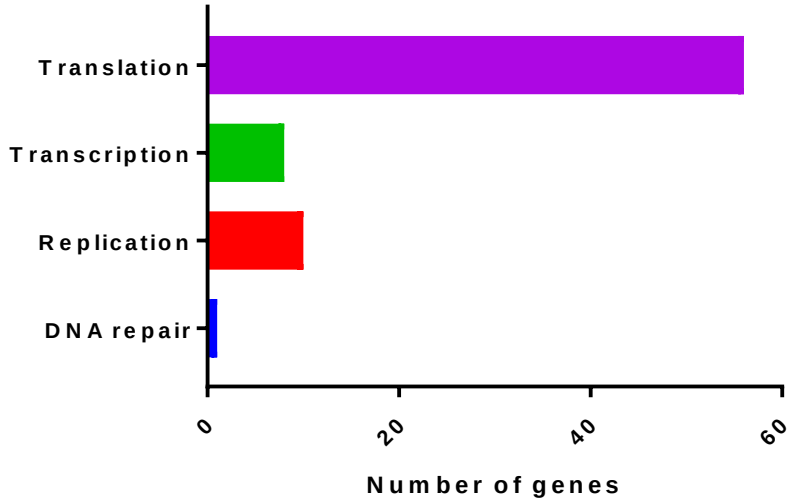


Figure 3. 6 Classification of essential genes belonging to the *Information Pathways* group. MTB orthologs of all genes were first identified using Mycobrowser, with functions inferred from data contained in the Universal Protein resource (UniProt,© 2002 – 2019 UniProt Consortium).

3.3.7 Essentiality of genes in the duplicated region of the MSM chromosome

Genome sequencing has shown that a chromosomal region is duplicated in MSM mc^2155 (Galamba et al., 2001). The duplicated regions are approximately 56 kb in size (spanning MSMEG0991 to MSMEG1044 and MSMEG2285 to MSMEG2337, respectively;(http://www.tigr.org) and make up a total of approximately 108 genes (Galamba et al., 2001, Warner et al., 2006). Interpretation of TnSeq data relies on the accurate mapping of transposon-junction reads to the genome of the mutagenized strain. Therefore, data analysis can be confounded by sequence variants (Carey et al., 2018). Duplicated genes are one of the challenges faced in TnSeq data analysis. Duplications may result in false non-essentiality calls because duplicated regions will contain more Tn-insertions compared to other non-duplicated regions (Carey et al., 2018). To assess the accuracy of our TnSeq prediction, we analyzed the essentiality calls of genes in the duplicated region with an expectation that these will either be NE or GA. **Table 3.6** shows some of the common genes (only those with gene names are shown) in this region, most of which were classified as GA, as expected.

Table 3. 6 Essentiality calls for selected genes (only those with gene names) within the MSM mc^2155 duplicated region.

MSM **ORF**	**Gene name**	**Gene description**	**Essentiality call**
MSMEG1014/2294	*dinB2*	DNA polymerase IV	GA
MSMEG1033	*nrdF2*	ribonucleotide-diphosphate reductase subunit beta	GA
MSMEG0994	*resD*	DNA-binding response regulator ResD	NE
MSMEG1041/2321	*serB*	phosphoserine phosphatase	NE
MSMEG2337	*fni*	isopentenyl pyrophosphate isomerase	GA

GA; Growth advantage, NE; Non-Essential

An MSM mutant, dubbed ΔDRKIN, was previously generated which lacks the chromosomal duplication (Warner et al., 2006). To study the essentialities of genes in the duplicated region, we constructed Tn libraries in ΔDRKIN. Three biological replicates were constructed and sequenced. **Table 3.7** summarizes the statistics that were generated for the libraries after TPP analysis. Overall, the libraries showed excellent genome coverage (>50 %) and minimal phage contamination.

Table 3. 7 A summary of TPP data analysis for ΔDRKIN TnSeq libraries

Tn library	Total reads	Mapped reads	No. TAs hit	% genome coverage	NZ mean	Phage contam. (%).
DRKIN1	6 306 607	4 292 559	46 232	59.5	52.2	1.5
DRKIN2	3 963 705	2 629 232	46 375	59.6	25.6	3.5
DRKIN3	4 871 008	3 191 628	40 027	51.5	61.0	3.1
DRKIN_3BIOREP	15 141 320	9 031 943	55 729	71.7	107	2.1
Notes: DRKIN_3BIOREP is a pooled library of all 3 biological replicates						

To determine which genes are conditionally essential for growth in the ΔDRKIN derivative, we used the resampling (permutation) method embedded in TRANSIT to compare WT and ΔDRKIN libraries. Comparative analysis by this method is carried out by summing the read counts at each gene site within a condition. The difference between the sum of read counts at each condition is calculated, and the significance in this difference is evaluated by comparing to a resampling distribution (DeJesus et al., 2015). For this analysis, the WT library (WT_3BIOREP) was used as a control and the ΔDRKIN library (DRKIN_3BIOREP) was set as an experiment. As expected, our results showed that there was an insignificant difference in genes required for optimal growth between the two strains, except for genes located in the duplicated region (**Table 3.8**). Specifically, only ribonucleotide-diphosphate reductase subunit alpha and beta were "more" essential in the ΔDRKIN strain as compared to the WT (**Table 3.8**) indicating that these genes are in fact essential for *in vitro* growth of MSM.

Table 3. 8 Comparison between WT and ΔDRKIN untreated libraries by resampling method. A negative and a positive delta mean shows that the gene is more essential in the experimental condition (DRKIN) or the control (WT) respectively. An adjusted p-value of ≤0.05 was used as a cut-off for a significant difference**.**

#Orff	Name	Description	Delta Mean	Adj. p-value
MSMEG_1019	-	ribonucleotide-diphosphate reductase subunit alpha	-487.4	0
MSMEG_1033	*nrdF2*	ribonucleotide-diphosphate reductase subunit beta	-487.8	0
MSMEG_2299	-	ribonucleotide-diphosphate reductase subunit alpha	-478.4	0
MSMEG_2313	*nrdF2*	ribonucleotide-diphosphate reductase subunit beta	-500.3	0

3.3.8 Comparison of MSM essential gene list with published essential genes of MTB

Next, we compared our 'core' essential genes in MSM with those identified in MTB (DeJesus et al., 2017) to establish which genes were specifically essential in MSM but not MTB. (**Figure 3.7**). All 277 'core' essential genes in MSM possessed MTB orthologs and, of these, >90 % were also identified as essential in MTB (**Figure 3.7**). However, this proportion is only about 40% of the MTB essential gene list, which has a larger number of essential genes compared to MSM (**Figure 3.7**).

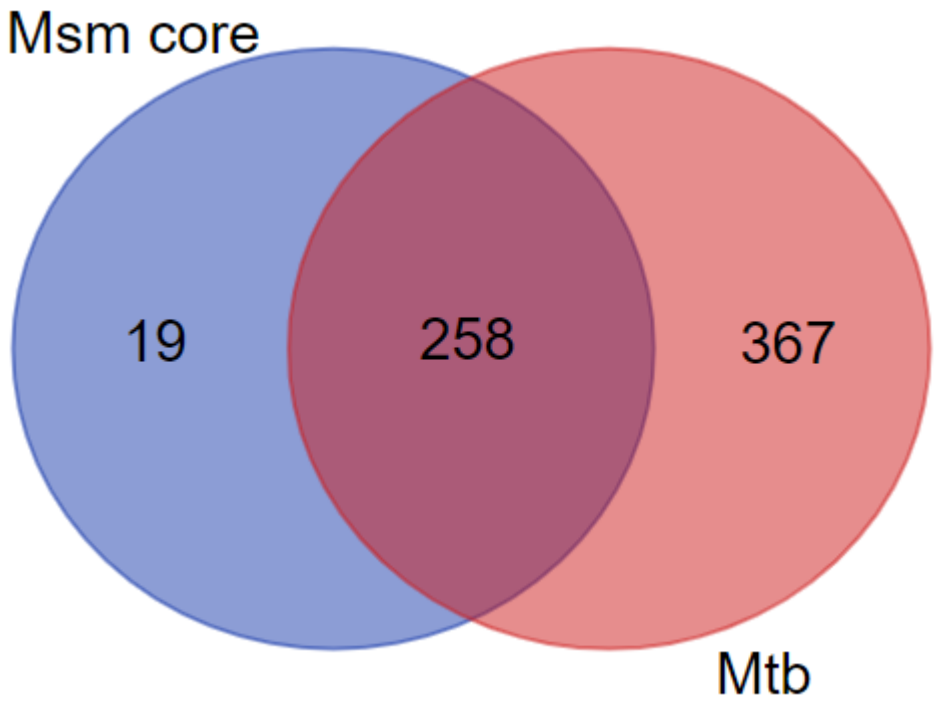


Figure 3. 7 A Venn diagram of MSM mc^{2}155 and MTB H37rv ES. 625 MTB essential genes (including ES, GD, ESD) were compared with 277 'core' essential genes from MSM. Both MSM and MTB gene sets were determined by TnSeq method.

Table 3.9 shows genes that are essential in MSM but not MTB. Complex II of the respiratory chain is formed by succinate dehydrogenase. This complex couples oxidative phosphorylation to central carbon metabolism by being an integral part of the citric acid cycle (Cecchini, 2013). MTB and MSM contains two operons which are predicted to encode succinate dehydrogenase enzymes (*sdh-1* and *sdh-2*). While both of these operons are dispensable for growth of MTB *in vitro* (DeJesus et al., 2017a), our TnSeq data show that MSMEG_1669C- MSMEG_1672C are essential for growth (**Table 3.9**), while MSMEG_0416- MSMEG_0420 is nonessential. These correspond to operons *sdh2* and *sdh1*, respectively (Pecsi et al., 2014). Our data (**Table 3.9**) also show that *metH* which encodes a coenzyme B_{12}-dependent methionine synthase (Cole et al., 1998) is essential only in MSM. Other metabolism genes which became differentially essential in MSM include malate: quinone-oxidoreductase (*mqo*), glutamine synthase type I (glnA) and 2-oxoglutarate dehydrogenase, E2 component, dihydrolipoamide succinyltransferase *(sucB*). From the information pathway functional group, we identified *rpsK* (ribosomal protein S11), *rplK* (ribosomal protein L11) and *mutT4* (MutT/nudix family protein) as differentially essential in MSM.

Table 3. 9 Genes which are unique to MSM during comparison with MTB essential gene set. MTB essential gene set used for the comparison was defined by DeJesus et al., 2017.

ORF ID	Gene name	Description	MTB ortholog.
MSMEG_4527		ferredoxin sulfite reductase	Rv2391
MSMEG_1670	*sdhA*	succinate dehydrogenase, flavoprotein subunit	Rv3318
MSMEG_6927	*mutT4*	MutT/nudix family protein	Rv3908
MSMEG_6182		conserved hypothetical protein	Rv3669
MSMEG_5049	*sucA*	2-oxoglutarate dehydrogenase, E1 component	Rv1248
MSMEG_4185	*metH*	methionine synthase	Rv2124
MSMEG_0949		HAD-superfamily protein subfamily protein IB hydrolase, TIGR01490	Rv0505
MSMEG_1346	*rplK*	ribosomal protein L11	Rv0640
MSMEG_1671	*sdhD*	succinate dehydrogenase hydrophobic membrane anchor protein SdhD	Rv3317
MSMEG_5435		acyl-CoA synthase	Rv1013
MSMEG_1672	*sdhC*	succinate dehydrogenase, cytochrome b556 subunit	Rv3316
MSMEG_6184		hydrolase, alpha/beta fold family protein	Rv3670
MSMEG_6900		penicillin-binding protein 1	Rv0050
MSMEG_1522	*rpsK*	ribosomal protein S11	Rv3459
MSMEG_4283	*sucB*	2-oxoglutarate dehydrogenase, E2 component, dihydrolipoamide succinyltransferase	Rv2215
MSMEG_2613	*mqo*	malate: quinone-oxidoreductase	Rv2852
MSMEG_4294	*glnA*	glutamine synthetase, type I	Rv2222
MSMEG_3213		conserved hypothetical protein	Rv3263

3.3.9 Validation of MSM essential genes using CRISPR- Cas9 technology

The HMM model provides a robust mechanism to analyze TnSeq data for essential regions (DeJesus and Ioerger, 2013). However, the accuracy and predictive power of TnSeq can be affected by several factors including library saturation, sequencing depth and normalization of reads counts (Chao et al., 2016). Hence, inferred gene functions/essentialities must be followed up with other, often lower-throughput approaches – for example, targeted gene inactivation via allelic exchange mutagenesis. To validate the TnSeq data, we employed the recently published anhydrotetracycline (ATc)-inducible CRISPR interference (CRISPRi) technology for mycobacteria (Rock et al., 2017). The genes selected for validation included *sdhA*, *sdhC* and *rpsK*, all of which were identified by TnSeq as MSM essential genes. On solid medium, growth of the CRISPRi knockdown mutants was eliminated in the presence of ATc-induction (**Figure 3.8**), confirming the inferred essentialities for *in vitro* growth.

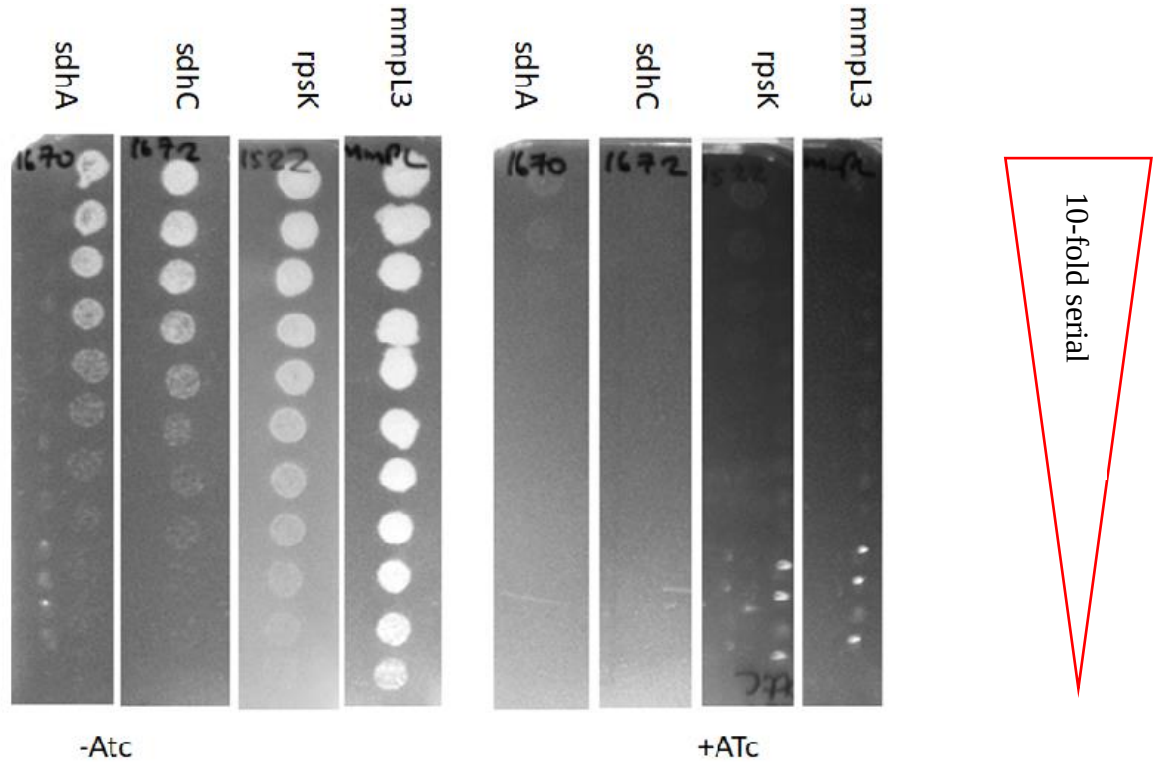


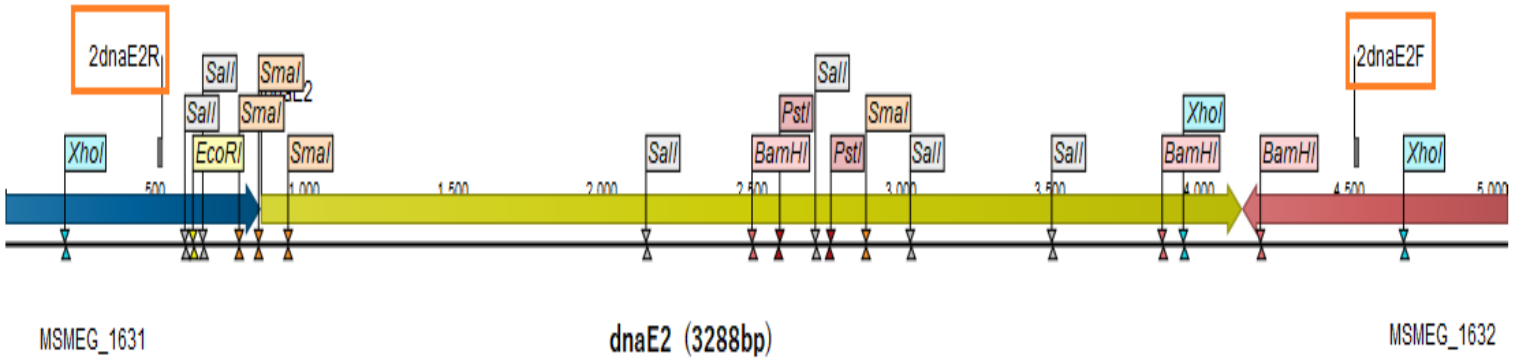
Figure 3. 8 CRISPRi validation of TnSeq essential genes. sgRNAs targeting *sdhA*, *sdhC* and *rpsK* were co-expressed with dCas9$_{sth1}$. The knockdown phenotype was monitored by plating tenfold serial dilutions (from top to bottom of plate) of cells on 7H10 solid media with/without ATc. *mmpL3* was used as a positive control. Essentiality is defined by no growth on ATc containing plates.

3.3.10 Comparison of essential genes between WT MSM, *uvrB* and *dnaE2* mutant strains

Our primary objective was to elucidate genes/pathways that might compensate for the absence of UvrB and DnaE2 under different DNA damaging conditions. Therefore, we next aimed to establish if there are any baseline differences in essential genes between the respective mutants' libraries and the WT parental strain under standard (non-DNA-damaging) conditions. First, though, we validated the genotypes of all mutant strains using PCR and whole-genome sequencing.

3.3.11 Confirmation of the genotypes of *dnaE2* and *uvrB* deletion mutants

Both *dnaE2* and *uvrB* mutants were generated previously in our laboratory. Initial confirmation of the mutant genotypes was done using PCR by designing primers flanking the genes of interest. Primers 2dnaE2F and 2dnaE2R were used to verify the genotype of Δ*dnaE2*. These primers were designed such that they anneal on the genes flanking *dnaE2* (**Figure 3.9**). A 4060 bp product was expected for the WT and a 973 bp product was expected for Δ*dnaE2*, as shown in **Figure 3.9**. Primers uvrBF and uvrBR were used to verify the genotype of Δ*uvrB*. For Δ*uvrB*, a product size of 2160 bp was expected for the WT and a 600 bp fragment was expected for Δ*uvrB* as shown in **Figure 3.10**. Both PCR reactions produced expected results hence validating the genotypes of our mutant strains.



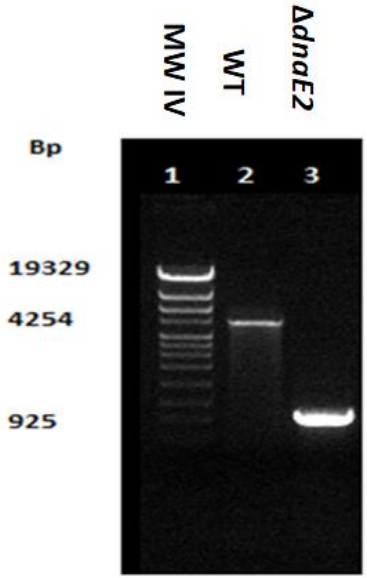
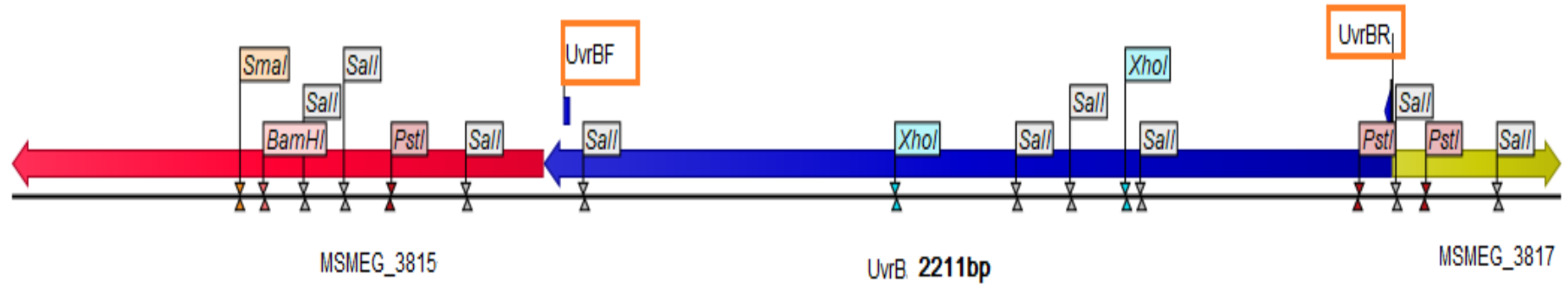


Figure 3. 9 Confirmation of Δ*dnaE2* deletion mutant by PCR . Top panel is a schematic representation of the *dnaE2* and flanking genes. Information on the genomic locus shown was taken from https://mycobrowser.epfl.ch/. The positions of PCR primers, 2dnaE2F and 2dnaE2R that were used for genotypic confirmation are shown in orange boxes. Image was generated using Qiagen CLC main workbench 20.0.4. The bottom panel is a 1% agarose gel showing separation of the PCR products; MW IV, molecular weight marker IV (from Roche).



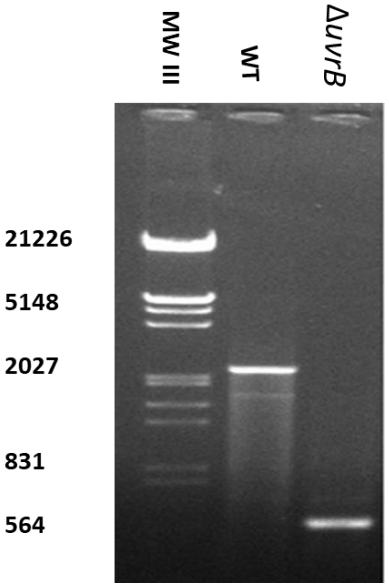


Figure 3. 10 Confirmation of the genotype of the Δ*uvrB* deletion mutant by PCR. Top panel is a schematic representation of the *uvrB* and flanking genes. Information on the genomic locus shown was taken from https://mycobrowser.epfl.ch/. Image was generated using Qiagen CLC main workbench 20.0.4. The positions of PCR primers, uvrBF and uvrBR are shown in orange boxes. The bottom panel shows PCR results; MIII, Roche molecular weight marker III.

3.3.11 Whole genome sequencing (WGS)

Additional confirmation of the deletion mutants was also carried out using WGS. To verify deletions in our genes of interest, a plot of read coverage along the genomic locus of *dnaE2* and *uvrB* was made. A drop in read coverage in these diagrams corresponds to the deleted region of the gene (**Figure 3.11 and 3.12**).

Additional single nucleotide polymorphism (SNP) analysis of the WGS data showed that the Δ*dnaE2* strain contained 6 SNPs in MSMEG_5106. Both the WT and Δ*uvrB* strains did not carry SNPs in this gene, which does not have an ortholog in MTB and is annotated as a putative HNM endonuclease in MSM (https://mycobrowser.epfl.ch/). Our TnSeq data showed that it was non-essential in both the WT, Δ*dnaE2* and Δ*uvrB* libraries under standard *in vitro* conditions. However, the reason for the accumulation of SNPs in MSMEG_5106 and the potential functional consequences thereof remain unknown.

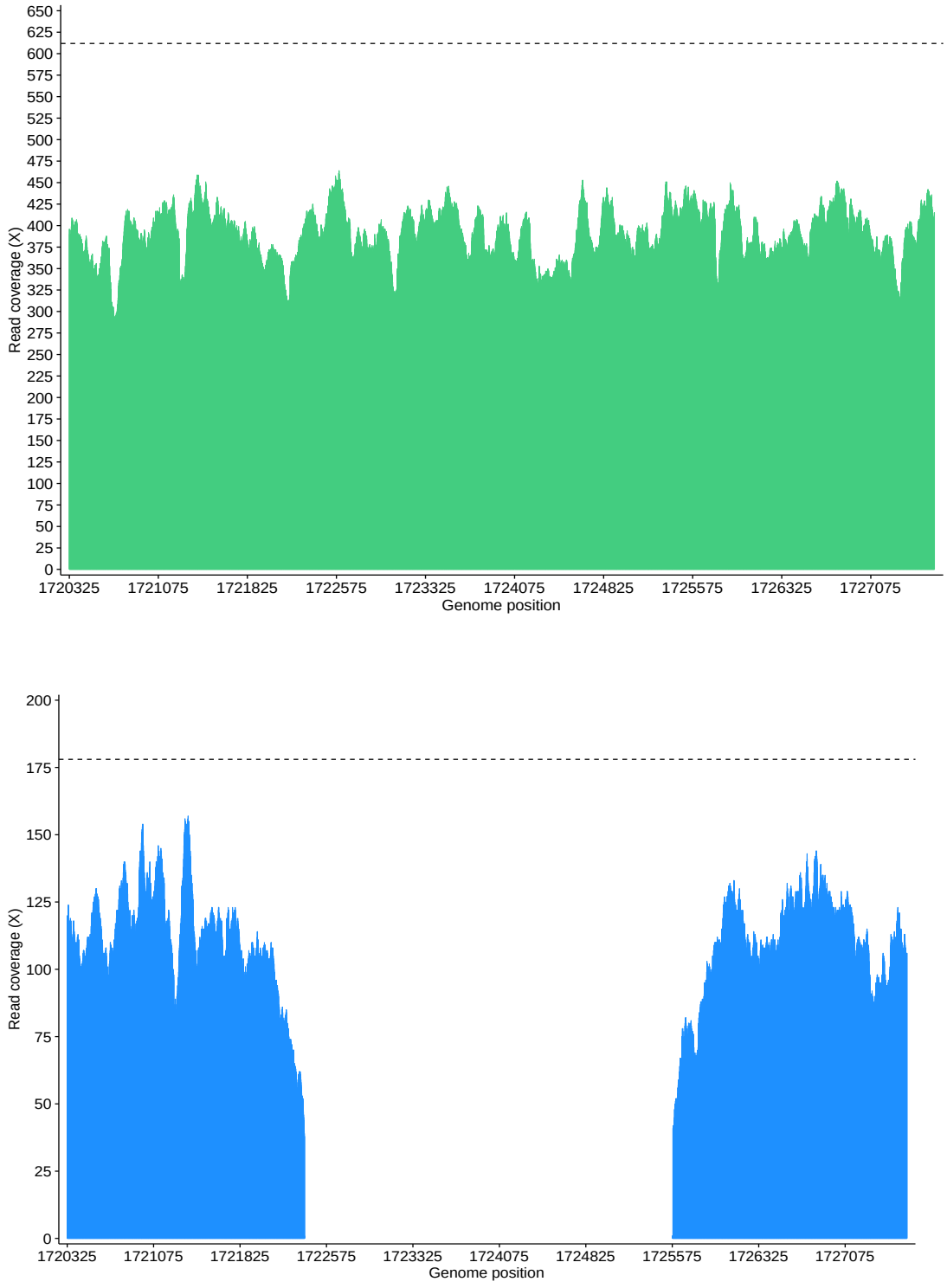


Figure 3. 11 Read coverage along the genomic coordinates of *dnaE2* in MSM. The coordinates of *dnaE2* in the genome of MSM mc^{2}155 span from 1722325-1725612. The top (green) image shows the WT strain and the bottom (blue) image shows the Δ*dnaE2* mutant. The average coverage for each strain is indicated by a dashed line. The drop in read depth in Δ*dnaE2* corresponded to the deleted region of the gene.

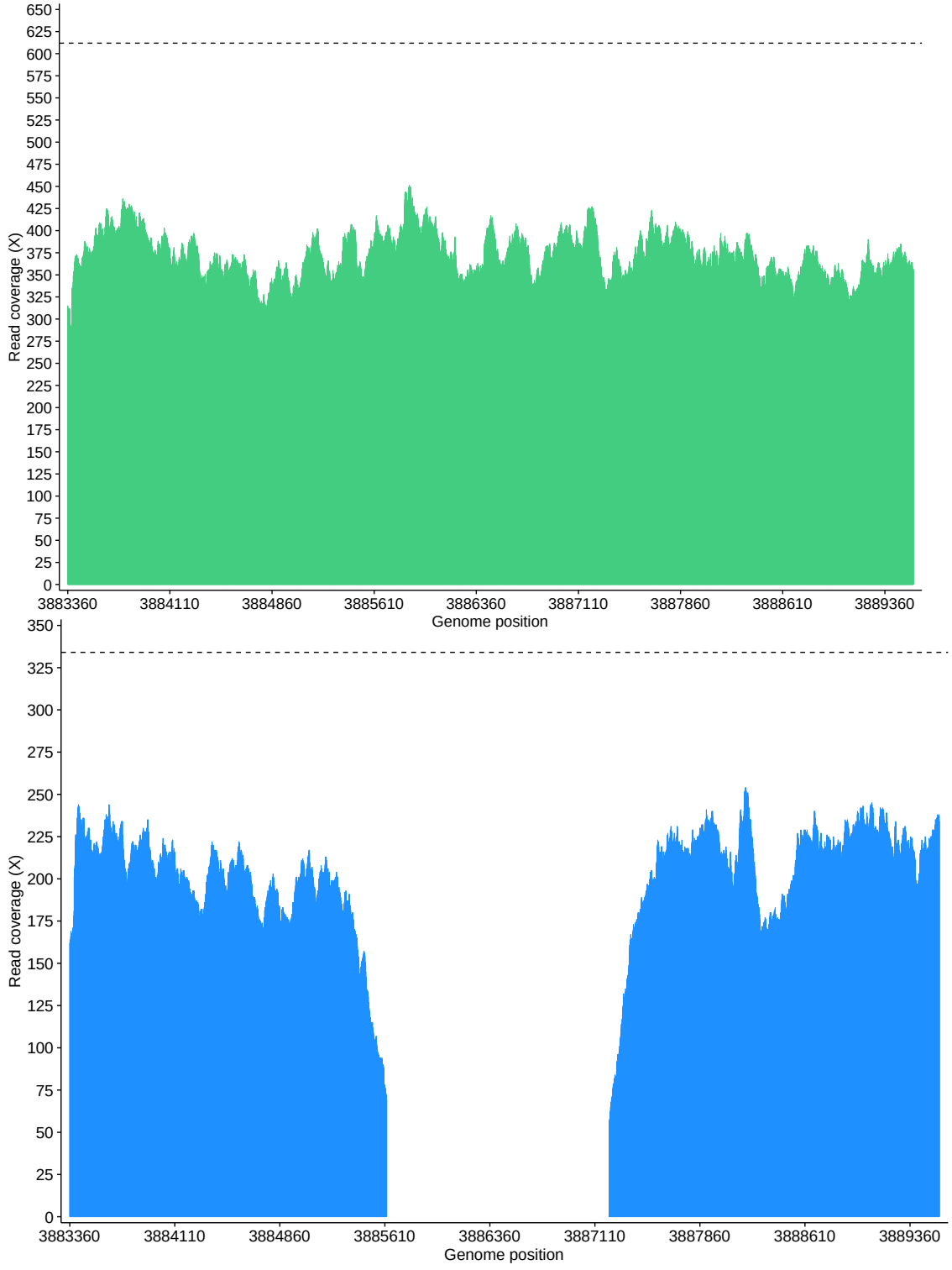


Figure 3. 12 A Read coverage along the genomic coordinates of *uvrB* in MSM. The coordinates of *uvrB* in the genome of MSM mc^2_155 span from 3885360- 3887570. The top green image shows the WT strain and the bottom blue image shows *uvrB* deleted strain. The average coverage for each strain is indicated by a dashed line. The drop-in read depth in Δ*uvrB* strain corresponded to the deleted region of the gene.

3.3.12 Tn library generation in Δ*dnaE2* and Δ*uvrB* strains of MSM

Tn insertion libraries were generated in *dnaE2* and *uvrB* backgrounds using the same approach described for the WT strain**. Table 3.11** summarizes the results of the TPP analysis. All libraries showed genome coverage of >50%. Libraries characterized by significant phage/vector contamination were generated using high phage concentration during the earlier phases of the research before the protocol was adapted to overnight transduction with low phage concentration.

Table 3. 10 A summary of TPP sequence analysis of Δ*dnaE2* and Δ*uvrB* Tn libraries

Tn library	Total reads	Mapped reads	No. TAs Hit	Insertion density	NZ mean	Phage contamination (%)
dnaE2_1	5159153	1040932	44788	57.7	22.7	32
dnaE2_2	2996518	616409	55192	71	11	0.1
dnaE2_2b	8065448	1718575	59350	76.3	28.2	21
uvrB_1	4252018	307399	43888	56.4	6.9	61
uvrB_2	2506152	588098	54164	69.7	10.6	0.3

The resampling method was used to compare WT with Δ*dnaE2* and WT with Δ*uvrB* libraries. The results showed that baseline essentiality calls in these strains were generally comparable with very few differentially essential genes detected. As expected, *dnaE2* in a Δ*dnaE2* library had a few Tn insertions compared to numerous Tn insertions in a WT. The same observation applied to the *uvrB* gene in the Δ*uvrB* library. These results confirmed deletion of the genes in the respective libraries, and therefore served as effective positive control for the Tn analysis. It was notable, however, that F_0F_1 ATP synthases and UDP-N-acetylglucosamine 1-carboxyvinyltransferase (*murA*) were more essential in the *dnaE2* background compared to WT, as evidenced by a significant reduction in mean Tn counts in Δ*dnaE2* compared to WT (**Figure 3.13**). MSM F_1F_0 ATP synthase is encoded in an operon which itself is part of a duplicated region (Tran and Cook, 2005). Dragset et al., 2019 recently established that the

duplication runs from MSMEG_4926-4946, which includes *murA* (Dragset et al., 2019), a gene that was also found to be differentially essential in Δ*dnaE2*. Our WT essentiality data support the presence of a duplicated region surrounding and including the ATP synthase operon because the genes in this operon gave a "growth advantage" essentiality call (Supplementary data 1). However, the "increased" essentiality of this operon in the *dnaE2* background prompted us to analyze our WGS data for this duplicated region in WT and in the *uvrB* and *dnaE2* mutant strains (**Figure 3.14**). A shift in read coverage which exceeds the average signifies the presence of more than one copy of the genome sequence; importantly, this is observed in WT and Δ*uvrB* but not in Δ*dnaE2* (**Figure 3.14**), suggesting the loss of the duplication in the *dnaE2* mutant.

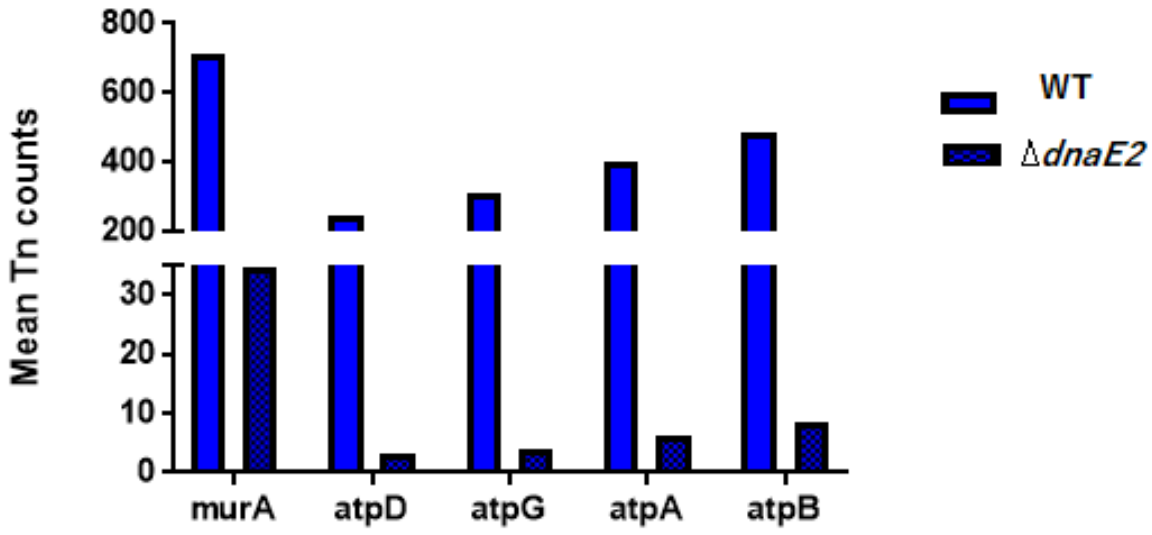
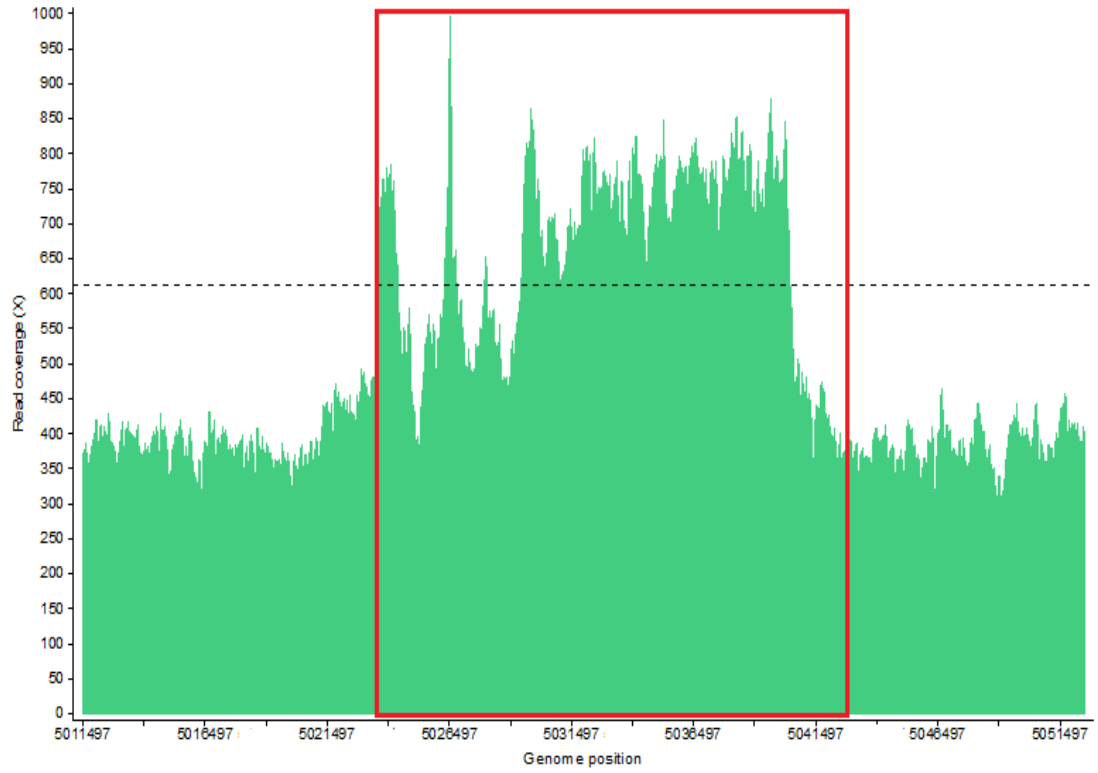
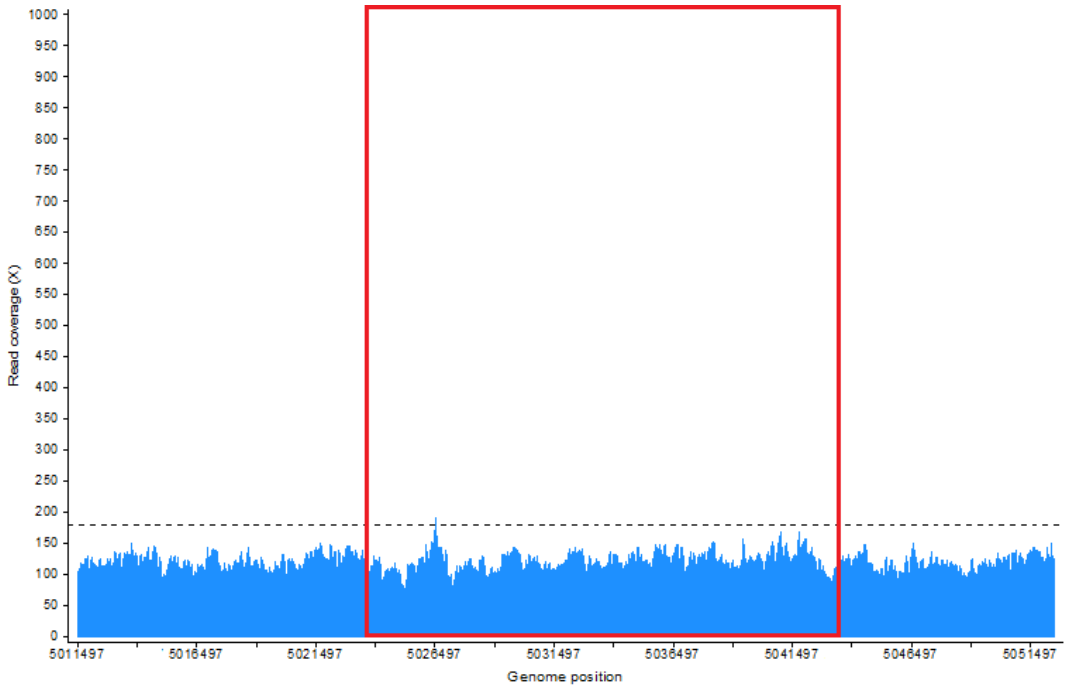


Figure 3. 13 Genes more essential in a Δ*dnaE2* background when compared to the WT.

WT


Read coverage (X)
Genome position
5011497
5016497
5021497
5026497
5031497
5036497
5041497
5046497
5051497

ΔdnaE2


Read coverage (X)
Genome position
5011497
5016497
5021497
5026497
5031497
5036497
5041497
5046497
5051497

Δ***uvrB***

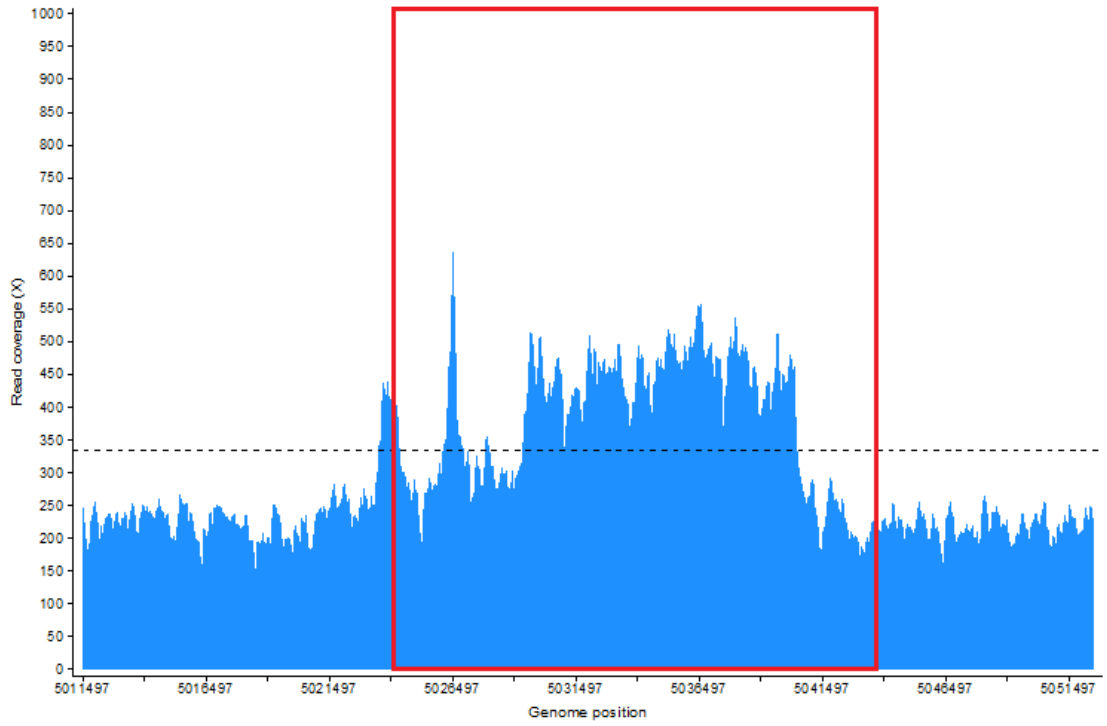


Figure 3. 14 Read coverage along the genomic coordinates of the duplicated region. It encompasses the F_1F_0 ATP synthase genes in WT, Δ*dnaE2* and Δ*uvrB* strains. The duplicated region spans from MSMEG_4926 to MSMEG_4946 with coordinates ranging from 5021832-5042826 which are highlighted in red boxes. The average read coverage for each strain is indicated by a dashed line. A shift in read coverage which exceed the average signifies the presence of more than one copy of this genome sequence. A shift in read coverage is observed in WT and Δ*uvrB* but not in Δ*dnaE2*.

3.4 Discussion

According to our data, complex and vector-free MSM Tn libraries can be achieved by overnight transduction with low phage titers. We initially generated our Tn libraries using 50mL cultures at OD_{600} of 2.0, infected with 1×10^{11}pfu/mL for 5h. This yielded very low titer libraries with fewer than 100 000 CFU/mL. Increasing phage concentration and the infection time to 7hrs improved the library titers. Nevertheless, upon sequence data analysis, we observed very high phage contamination. Phage contamination affected downstream analysis as only a small proportion of the sequenced reads mapped to the genome and consequently very low library saturation was observed. The protocol was therefore modified by using 200 mL culture at OD_{600} of 0.6 – 0.8 and infection with low phage concentrations ($<1\times10^{11}$) overnight. These changes greatly improved the library titer and there was also a reduction in vector contamination. At an OD_{600} of 2.0, cells are either in early stationary or stationary phase (Peñuelas-Urquides et al., 2013, Jaishankar and Srivastava, 2017). In this phase, cells activate the stringent response mechanism, the cell size decreases, the cell envelop become rigid, membrane fluidity is reduced and the cell wall is highly cross-linked (Jaishankar and Srivastava, 2017). It then follows that bacteriophage infections will be less efficient at the stationary phase compared to the exponential phase where the cell membranes are less rigid. The success of this new protocol shows that effective transduction of MSM with ΦMycoMarT7 is perhaps optimal in logarithmic-phase cells and that the infection process might be made more efficient through overnight incubation. Furthermore, our data shows that MSM does not replicate in MP buffer (used during phage infection), hence overnight infection appears not to carry any risk of selective clonal proliferation of cells.

Even though this adapted protocol offered better library results, we were still puzzled by the low number of reads that mapped to the genome (only ~25%) in sample WT2 which had very low phage contamination (0.8%). There are several factors that can contribute to this. For example, where excessive primer dimers carried over from library preparation are also sequenced together with the library, they do not map to the genome (DeJesus et al., 2015). Another common problem is adapter contamination, a problem which usually arises when fragments of the sequence library are shorter than the read length, and which means that the chromosomal region of interest is completely sequenced in both reads (Sturm et al., 2016). Lastly, sequences not mapping to the reference genome could be an indication of poor sequencing accuracy. Therefore, good library generation and preparation is a prerequisite of

good library sequencing data. Since all protocols carry inherent limitations, user errors, stochastic differences and experimental errors in sequencing can be mitigated by increasing sequencing read depth, doing technical replicates and also by using biological replicates (Robasky et al., 2014). Our data greatly support these mitigation strategies as we noted overall good library data after we pooled our biological replicates: increased total reads directly affected the percentage of mapped reads and, consequently, the overall saturation of the library.

We further validated the quality of the essential gene list of WT MSM by specifically looking at the essentiality calls of genes in the duplicated region. A significant number of these genes were classified as growth advantaged (GA). These were expected results: higher than average read counts indicate genes/regions which are duplicated (DeJesus and Ioerger, 2013). A further comparison between WT and ΔDRKIN strain showed that the only essential genes in the duplicated region are the ribonucleotide-diphosphate reductase subunit beta (*nrdF2*) and alpha (*nrdE*). This is consistent with previous work showing that insertional inactivation of *nrdF2* in a ΔDRKIN strain was not possible while the gene could be readily deleted in WT MSM mc^2155 which has the duplication (Mowa et al., 2009). Furthermore, *nrdF2* is essential in MTB (DeJesus et al., 2017a).

The comparison of our essential gene list with that of Dragset et al., 2019 showed large overlap in predicted gene essentialities. Although expected given that the strains and growth medium were the same in both studies, this result was nonetheless reassuring. However, it is worth noting that almost 100 of our essential genes did not appear in the Dragset list. Most of these genes received a "GD" assignment in our analysis, which is given when the mean read count of the gene is 'intermediate' when compared to the global mean. This state occurs somewhere between the "ES" and "NE" states (DeJesus et al., 2015). We observed an unusually large number of GD in WT1 library which had genome coverage of 31%. This shows that low coverage has a strong negative impact on TnSeq analysis since it makes it difficult to make essentiality calls confidently. However, the presence of minor differences in essential genes between these libraries was expected.

Studies have shown that bacteria of the same strain in different laboratories can behave differently. In their opinion paper, where they discuss plasticity in bacterial genomes, Fux et al., 2005 highlight that bacterial genomes evolve during serial *in vitro* passage. Repeated sub-culturing of bacteria can result in generations of various genotypes overtime(Fux et al., 2005): This could be brought about by deletions, single nucleotide polymorphisms and duplications,

as have been the case with the BCG vaccine. These genetic alterations are thought to account for the phenotypic differences observed between current BCG vaccine strains (Fux et al., 2005). It is because of this high degree of genetic flexibility that the same strains of bacteria from different laboratories may express different phenotypes under identical growth conditions. In fact, selection for rapidly growing bacteria in planktonic cultures is thought to occur at the expense of genes that are unnecessary *in vitro*. The loss of these genes could provide substantial savings in energy and amino acid requirements, and hence, the designation of 'laboratory reference strain', cannot truly represent a species. Rather, the complete species genome is dispersed throughout the population (Fux et al., 2005). Evidence for *in vitro* evolution has been provided for *M. tuberculosis* (Ioerger et al., 2010); it would be interesting to conduct a similar analysis of global *M. smegmatis* genomes.

A comparison of the TnSeq and CRISPRi methods showed a 60 % overlap with many genes being essential only by CRISPRi. It should be noted however that the CRISPRi analysis was restricted to MSM genes that carry an ortholog in MTB i.e. the CRISPRi analysis was not genome wide. However, this analysis highlighted core differences between the two methods. The F_1F_0 ATP synthase and *nrdF2* genes which are contained in a duplicated region were among those genes that were essential with CRISPRi but not with TnSeq. This confirms that CRISPRi is not affected by duplications which give it an advantage over the TnSeq method. Operonic genes, *kdpA* and *kdpB*, were also essential only with the CRISPRi method but not with TnSeq. Both belong to a potassium uptake system in mycobacteria which becomes particularly important during low potassium levels (Ali et al., 2017). Generally, Kdp-ATPase complex consist of four subunits, encoded in the *kdpFABC* operon (Ali et al., 2017). The CRISPRi method is confounded by polar effects for genes which are in operons; *i.e.*, any operonic gene downstream the dCas9 binding site will be transcriptionally silenced together with the gene of interest (Rock et al., 2017). Tn insertions within an operon may also exert a polar effect on downstream genes which may lead to incorrect fitness calls (Carey et al., 2018, Van Opijnen et al., 2009) however, our results suggest that polar effects are more profound in CRISPRi compared to TnSeq. It is possible that these are some of the reasons why many genes designated as essential by CRISPRi were not identified as essential by TnSeq analysis.

The comparison of TnSeq and CRISPRi data sets yielded core essential genes in MSM. Analysis of these "core essentials" showed that most genes required for optimal growth in MSM function in intermediary metabolism and respiration, information pathways and cell wall and cellular processes. Biological processes that were abundant in this data included translation

and amino acids biosynthesis pathways. These functions are consistent with those that have been found in MTB essential gene list (Sassetti et al., 2003). Nonetheless, a recent study has identified 1166 genes that make up the core genome of MTB as established from 183 complete and draft genomes of MTB strains (Zakham et al., 2021). Compared to 625 MTB ES (DeJesus et al., 2017a) and 277 MSM core gene list (our data), this number is much higher. This implies that the concept of core genome as determined by *in silico* analysis does not necessarily translate into *in vitro* essentiality. Bacteria rapidly adapt to *in vitro* conditions and gene essentiality *in vitro* may depend on the exact experimental conditions and genetic alterations (Fux et al., 2005).

By comparing the MSM essential gene list to that of MTB (DeJesus et al., 2017a), we observed that most MSM essential genes have MTB orthologs that are also essential in MTB. This agrees with the recent publication by Dragset et al., where they noted that nearly 96% of genes required in MSM has mutual orthologs in MTB from which 90% are also essential in MTB. However, compared to MSM's 451 essential gene set (our data), MTB has 625 essential genes (DeJesus et al., 2017a) suggesting that the number of ES genes does not scale with genome size as MSM has a larger genome than MTB (Gcebe et al., 2016). The large number of essential genes in MTB as compared to MSM could be attributable to lifestyle differences; MTB is an obligate human pathogen whereas the non-pathogenic MSM is an environmental saprophyte. However, many of these essential genes in MTB have been found to lack orthologues in other bacteria, suggesting that the minimal gene set required for *in vitro* survival varies greatly between bacterial species (Sassetti et al., 2003).

It is important to understand similarities and differences in MSM and MTB essential processes as this may help in data interpretation and will influence our choice of study organism. Several genes involved in intermediary metabolism and respiration were differentially essential in MSM. These included Complex II of the respiratory chain which is formed by succinate dehydrogenase. This complex couples oxidative phosphorylation to central carbon metabolism by being an integral part of the citric acid cycle (Cecchini, 2013). MTB and MSM contain two operons which are predicted to encode succinate dehydrogenase enzymes (*sdh-1* and *sdh-2*). While both of these operons are non-essential for growth of MTB in-vitro (DeJesus et al., 2017a), our TnSeq data show that MSMEG_1669C-MSMEG_1672C (*sdh2*) is essential for growth while MSMEG_0416-MSMEG_0420 (*sdh1*) is non-essential. These results are consistent with those generated by allelic gene replacement which showed that the *sdh1* operon was non-essential in MSM while the *sdh2* operon could only be deleted in a merodiploid

background, demonstrating that *sdh2* is essential for MSM growth (Pecsi et al., 2014). Malate: quinone-oxidoreductase (MQO) is also differentially essential in MSM. In MTB, the conversion of malate to oxaloacetate is carried out by either MQO or malate dehydrogenase (Mdh). MSM, however, does not possess *mdH* (Berney and Cook, 2010). Consistent with the different gene complement, our analysis using two TnSeq data sets and CRISPRi showed that, while both of these genes are not essential in MTB (DeJesus et al., 2017a), *mqo* is essential in MSM.

Another enzyme, MetH, was also found to be essential in MSM but not in MTB. In both MSM and MTB, a B_{12}-dependent methionine synthase (MetH) and a B_{12}-independent methionine synthase (MetE) catalyze the synthesis of L-methionine (Young et al., 2015, Kipkorir et al., 2020). However, *metH* is non-essential in MTB but essential in MSM. Conversely, *metE* is essential in MTB but non-essential in MSM. The difference in essentiality of these genes in the two species indicate core differences in vitamin B_{12} metabolism, as reported recently (Kipkorir et al., 2020). Another interesting observation was the essentiality of *mutT4*, a DNA repair protein in MSM. There are 4 MutT/NUDIX family proteins in MTB: *mutT1*, *mutT2*, *mutT3* and *mutT4*, all of which have MSM homologs (Dos Vultos et al., 2006). None of these genes is essential for optimal growth of MTB (DeJesus et al., 2017a), however the essentiality of *mutT4* in *MSM* was reported previously (Dos Vultos et al., 2006). It is not clear why only *mutT4* is essential in MSM, but this might point to a potential role which is separate from DNA repair in MSM. Again, while these differences could reflect adaptations to different lifestyles and environmental niches, they emphasize that caution should be taken when extrapolating data between the two species.

Our analysis of a limited set of clinical MTB isolates revealed no to few SNPs in *uvrB* and *dnaE2*, respectively, suggesting stable maintenance of these sequences. It is tempting to interpret this result as indicating the importance of these genes in DNA repair during host infection. However, a much larger sample size is required to make any conclusive statements. The identification of 6 SNPs in MSMEG_5106 in the Δ*dnaE2* strain was surprising, but the biological implications (if any) remain unclear.

The Δ*dnaE2* mutant was also found to have "lost" the duplicated chromosomal region containing the ATP synthase genes. Duplications in the genome are thought to evolve for different reasons: They may act as backup copies in case of the loss of functionally related genes (Moleirinho et al., 2011), they may result from a selective pressure (e.g., a new niche,

antibiotic pressure), or, in case of laboratory strains, duplications could result from *in vitro* passage (Domenech et al., 2010). Considering these causes of duplications, loss of a duplicated region could happen by chance or as a result of loss of selective pressure. In this context, it is worth noting that a derivative of MSM mc^{2}155 dubbed ΔDRKIN, was previously generated which lacks the chromosomal duplication (Warner et al., 2006).This mutant was identified during the attempted construction of a *dinB* knockout mutant, thereby providing another example of loss of a genomic duplication during the course of propagation and genetic manipulation of MSM (Warner et al., 2006), underscoring the genomic plasticity of this organism and the importance of analyzing engineered strains by WGS.

In conclusion, this study established baseline essential gene lists for MSM WT and mutant strains under optimal *in vitro* conditions. These results were critical for the next phase of the research, which aimed to investigate conditional gene essentialities in different genetic backgrounds during exposure to DNA damage.

Chapter 4:

Investigating conditional gene essentialities in DNA repair-deficient *Mycobacterium smegmatis* under genotoxic stress

4.1. Introduction

It is important to understand how DNA repair mechanisms are coordinated and what happens when certain DNA repair mechanisms are disabled. This is particularly important because these mechanisms, together with their complementary pathways, ensure mycobacterial survival in the face of otherwise lethal genotoxic stress provided by both the host and anti-TB drug, a phenomenon which facilitates drug resistance. Inhibition of these mechanisms together with their complementary pathways might therefore be used to target drug resistance TB. In this research, libraries of mutants containing random, genome-wide Tn insertions in WT MSM mc^2155 and derivative *uvrB* and *dnaE2* deletion mutant backgrounds were created and treated with DNA crosslinking agent, MMC, and the gyrase inhibitor, MOX.

Mitomycin C DNA damage

MMC is an antibiotic that is produced by *Streptomyces caespitosus* (Hata et al., 1956). It is a DNA damage chemotherapeutic drug currently used for a variety of cancers (Bradner, 2001). Upon entry into cells, MMC's quinolone functional group is spontaneously reduced and it initiates G-G inter-strand crosslinks (ICLs) between DNA strands and also causes DNA alkylation (Tomasz and Palom, 1997). DNA damage caused by MMC results in mutagenesis, blocked replication/transcription and DSBs (Roy and Schärer, 2016). This damage induces a DNA damage response in which over 100 genes are transcribed in mycobacteria. Most of these transcribed genes function in DNA repair pathways and include the SOS response regulatory genes, *recA* and *lexA* (Müller et al., 2018). ICLs caused by MMC result in blocks in DNA replication and transcription. Processing of ICLs involves unhooking of the ICL from one of the strands and synthesis past the lesion (Roy and Schärer, 2016). This involves excision of the ICL which consequently results in DSBs (Zhang and Walter, 2014). Repair of ICLs has been shown to rely heavily on proteins involved in HR, TLS and NER (Noll et al., 2006).

Moxifloxacin induced DNA damage

MOX belongs to the fluoroquinolone class of antibiotics. Generally, fluoroquinolones act by inhibiting DNA gyrase, DNA topoisomerase IV and homologous type II topoisomerase (Pestova et al., 2000). In mycobacteria, fluoroquinolones bind primarily to GyrA (Kumar et al., 2014). Fluoroquinolones cause bacterial cell death by arresting gyrase-DNA complexes, leading to accumulation of DSBs and ultimately activation of the SOS response (Kumar et al., 2014). Apart from mutations in the *gyrA* gene, resistance to fluoroquinolones is also facilitated

by active efflux or reduced influx, both of which decrease intracellular drug concentrations (Hooper and Jacoby, 2016). Mycobacteria encode an ABC transporter responsible for fluoroquinolone efflux (Viveiros et al., 2012).

4.2. Aims and objectives

This study aimed to elucidate otherwise cryptic repair pathways that might become conditionally essential under genotoxic stress in defined mutants lacking either one of two dominant DNA repair pathways: *uvrB*-dependent NER or *dnaE2*-dependent TLS. The specific objectives were:

- To expose Δ*uvrB* and Δ*dnaE2* mutants of MSM to whole-genome transposon mutagenesis libraries to different genotoxic conditions including MMC and MOX.
- To sequence the libraries and analyse the data for conditionally essential genes.
- To identify aggravating / antagonistic genetic interactions.
- To validate any genes of interest using CRISPRi.

4.3. Results

4.3.1. Phenotypic characterization of *dnaE2* and *uvrB* deletion mutants

MICs of MMC and MOX were determined for all three strains (WT, Δ*uvrB*, Δ*dnaE2*) by the resazurin microtiter assay (Palomino et al., 2002). This confirmed that Δ*uvrB* was hypersusceptible to MMC (MIC_{90} = 3.9 ng/ml) while the MIC for Δ*dnaE2* (MIC_{90} = 62 ng/ml) was half that determined for the WT parental strain (**Figure 4.1**). In contrast, both Δ*dnaE2* and Δ*uvrB* mutants had MIC_{90} values for MOX which were comparable to WT (**Figure 4.1**).

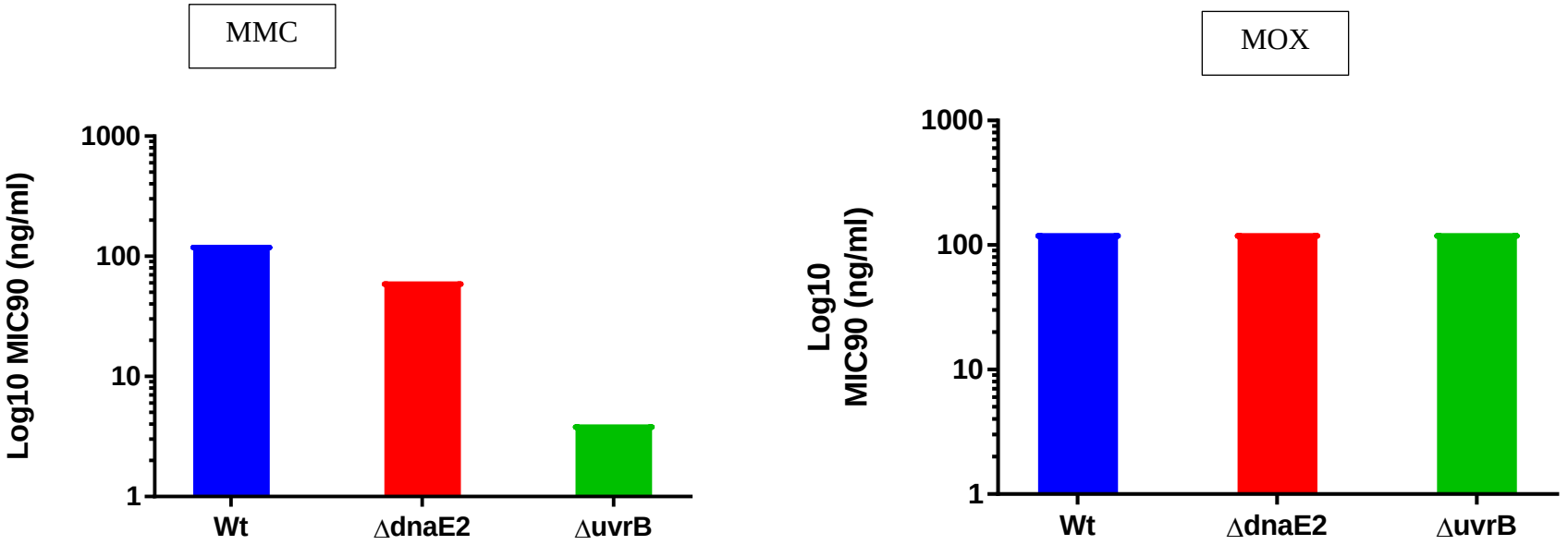


Figure 4. 1 Determination of MIC_{90} values for MMC and MOX. MICs were determined using resazurin microtiter assay (Palomino et al., 2002). The MIC_{90} was defined as the lowest drug concentration that prevented resazurin colour change from blue to pink. Data are representative of three biological replicates for each strain. The bacterial isolates were cultured in middlebrook 7H9 broth.

4.3.2. Phenotypic characterization: MMC DNA damage tolerance assay

Having established the relative susceptibilities of the different strains to the two agents, DNA damage tolerance assays were performed to determine the concentration of each drug to be used for Tn library selection. The aim was to find a concentration of MMC which was high enough to induce genotoxic stress but low enough to support growth at a very high inoculum; this was important to ensure enough biomass to extract high quality DNA. To this end, 10 µl of log_{10}-fold dilutions of WT, Δ*dnaE2* and Δ*uvrB* (each grown to a final OD_{600} ~0.6) were plated on standard 7H10 medium containing different concentrations of MMC (**Figure 4.2**).

For WT, the optimal concentration was determined to be 0.031 µg/ml MMC which was equivalent to ¼× MIC; for Δ*dnaE2*, it was 0.016 µg/ml (¼×MIC) and for Δ*uvrB* it was 0.001µg/ml (also ¼×MIC) (**Figure 4.2**). For MOX, the concentration was found to be 0.031 µg/ml which was equivalent to ¼×MIC for all three strains (**Figure 4.3).**

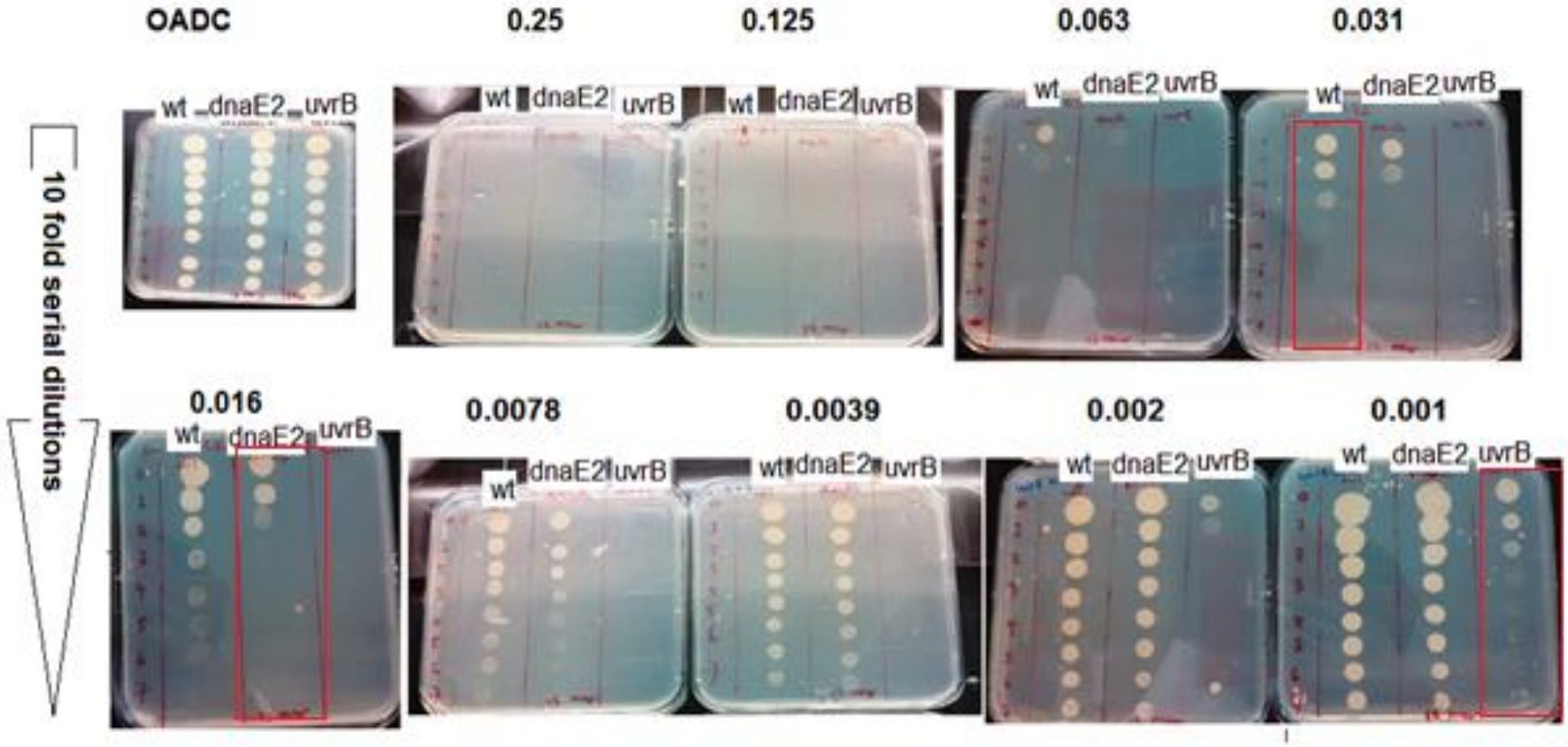


Figure 4. 2 MMC damage survival assay. The spots represent 10-fold serial dilutions which runs from undiluted (10^0) to 10^{-7} for WT, Δ*dnaE2* and Δ*urvB* strains. 10 µl of the serial dilutions were plated on 7H9/OADC medium containing different concentrations of MMC- shown as bold number(all concentrations are in µg/ml). The concentrations range from 2× to ¼ MIC. 10 µl of cells were also spotted on 7H10/OADC only as a positive control. Red rectangles indicate the drug concentration chosen for library selection for each strain.

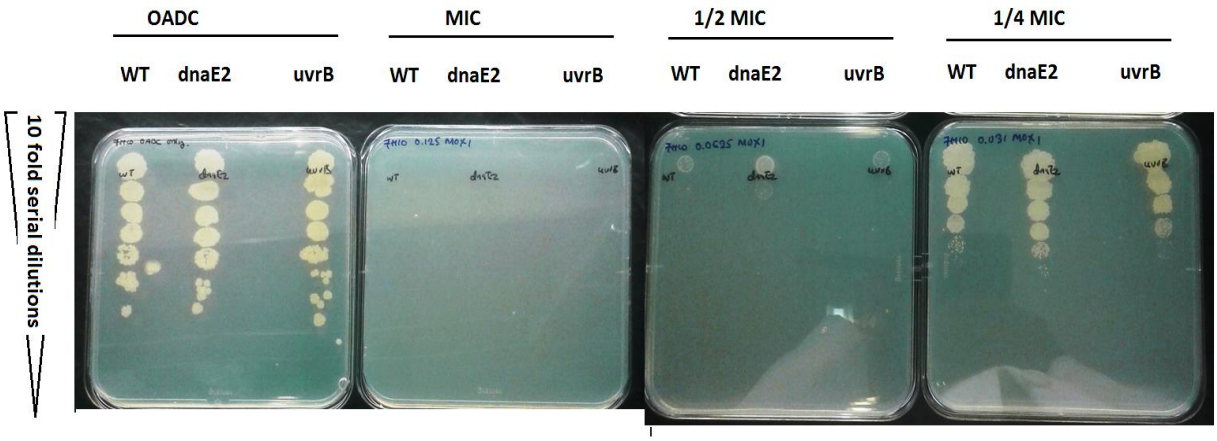


Figure 4. 3 MOX damage survival assay. The spots represent 10 µl of 10-fold serial dilutions of WT, Δ*dnaE2* and Δ*urvB* strains which runs from undiluted (10^0) to 10^{-7} plated on 7H9/OADC medium containing different concentrations of MOX. The concentrations range from 1× to ¼×MICs. 10 µl of cells were also spotted on 7H10/OADC only as a positive control.

4.3.3. Tn-Seq library selection

Tn libraries were generated in MSM WT, Δ*dnaE2* and Δ*uvrB* strains and selected on media containing either MMC or MOX. For library selection, two million cells per library were cultured on 7H10 plates containing ¼× MIC of MMC (specific to each strain) and MOX. Initial analyses of the sequence data were carried out using 'TRANSIT pre-processor' (TPP) which extracts read counts from raw sequences and maps them to the reference genome (DeJesus et al., 2015). TPP also gives information about the quality of the raw sequence data; *e.g*., percentage of TA sites in the genome which have Tn insertions. **Table 4.1** summarises the statistics that were generated for the libraries after TPP analysis. Most of the treated libraries met the minimum required transposon library genome coverage of 50% and the non-zero mean (NZ mean) of >10 (DeJesus et al., 2015); the exceptions were the Δ*dnaE2*_MMC and Δ*uvrB*_MOX libraries in which all the replicates did not achieve the genome insertion coverage of 50% (**Table 4.1**).

Table 4. 1 A summary of TPP results for MMC-treated Tn libraries. Two million cells in each library were cultured on 7H10 agar supplemented with kanamycin, Tween80 and MMC at ¼×MIC for each strain. For each library, cells were cultured in two disposable polystyrene Petri dishes (245mm × 245mm). Sequence reads were analyzed using TPP.

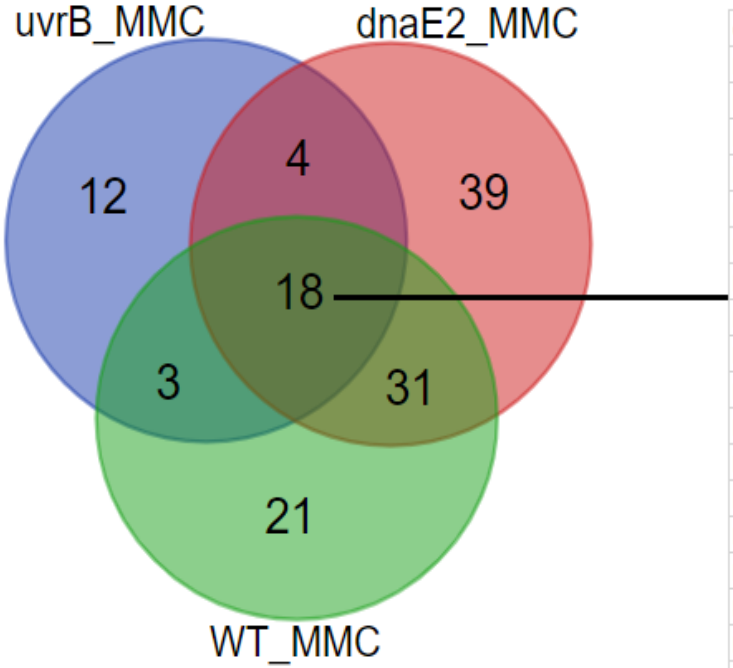
Tn library	Total reads	Mapped reads	Insertion density	NZ mean
		MMC		
WT_MMC1	4 623 343	1 096 428	29.7	42.7
WT_MMC2	6 915 913	3 777 078	57.4	33.7
dnaE2_MMC1	5 559 378	4 558 240	39.9	130.2
dnaE2_MMC2	1 130 108	676 745	32.2	21.4
uvrB_MMC1	4 068 760	3 014 815	56.3	63.1
		MOX		
WT_MOX1	1914402	895 354	62.3	11.4
WT_MOX2	4431117	2 399 962	73.2	32.5
dnaE2_MOX1	520 283	253 433	42.4	7.3
dnaE2_MOX2	3 673 742	2 135 841	56.8	36.3
uvrB_MOX1	726 548	361 287	23.7	12.5
uvrB_MOX2	2 942 251	1 240 378	45.9	38

NZ mean: Non-zero mean, a default normalization procedure.

4.3.4. Analysis of conditionally essential genes during MMC treatment

The analysis of conditionally essential genes was carried out using the re-sampling (permutation) method which compares the untreated Tn library with the treated Tn library. Comparative analysis by this method sums the read counts at each gene site within a given condition. The difference between the sum of read counts at each condition is calculated, and the significance of this difference is evaluated by comparing to a resampling distribution. Conditionally essential genes are those with a $P_{adjusted}$ value of ≤0.05 (DeJesus et al., 2015). For each strain, we compared pools grown under standard laboratory conditions (entered in Resampling interface as controls) with pools treated with MMC (entered in Resampling interface as experiment) to define genes or pathways which are conditionally essential during MMC exposure.

In total, our analysis identified 73, 92 and 37 conditionally essential genes during MMC treatment in WT, Δ*dnaE2* and Δ*uvrB* Tn libraries, respectively (**Figure 4.4 and Tables 4.2, 4.3 and 4.4**). The analysis identified 18 genes which were conditionally essential in all libraries, irrespective of strain background. These included the ATP-dependent DNA helicase, PcrA; members of the phosphate ABC transporter, PstC and PstS; mannosyltransferase, PimE; the antigen 85A mycolyltransferase; the cell division protein, FtsX; and asparagine synthase, AsnB (**Figure 4.4**).



ORF	Name	Description
MSMEG_4243		serine/threonine-protein kinase PK-1
MSMEG_5534	pcrA	ATP-dependent DNA helicase PcrA
MSMEG_5781	pstC	phosphate ABC transporter, permease protein PstC
MSMEG_6363		cysteine desulfurase family protein
MSMEG_5149	pimE	mannosyltransferase
MSMEG_5782	pstS	phosphate ABC transporter, phosphate-binding protein PstS
MSMEG_3641		hypothetical protein MSMEG_3641
MSMEG_4269	asnB	asparagine synthase (glutamine-hydrolyzing)
MSMEG_0786		serine/threonine protein kinase
MSMEG_4245		hypothetical protein MSMEG_4245
MSMEG_2090		putative cell division protein FtsX
MSMEG_0788		hypothetical protein MSMEG_0788
MSMEG_0386		NAD dependent epimerase/dehydratase family protein
MSMEG_0035		FHA domain-containing protein
MSMEG_5041		acyltransferase
MSMEG_6398		antigen 85-A
MSMEG_1654		isocitrate dehydrogenase, NADP-dependent
MSMEG_6387		arabinosyltransferase A

Figure 4. 4 A comparison of gene sets predicted to be conditionally essential during MMC treatment in WT, Δ*dnaE2* and Δ*uvrB* libraries. The libraries were selected on Middlebrook 7H10 agar plates.

Reassuringly, *dnaE2* and the NER genes (*uvrA*, *uvrB* and *uvrC*) were conditionally essential in the MMC-treated WT library (**Table 4.2**)as has been shown in other studies(Müller et al., 2018, Rand et al., 2003). However, fewer Tn insertions were identified in *dnaE2* than *uvrB* in the WT_MMC library (**Figure 4.5**); this suggested that even though they are both conditionally essential, *dnaE2* might be more important under this condition than *uvrB.*

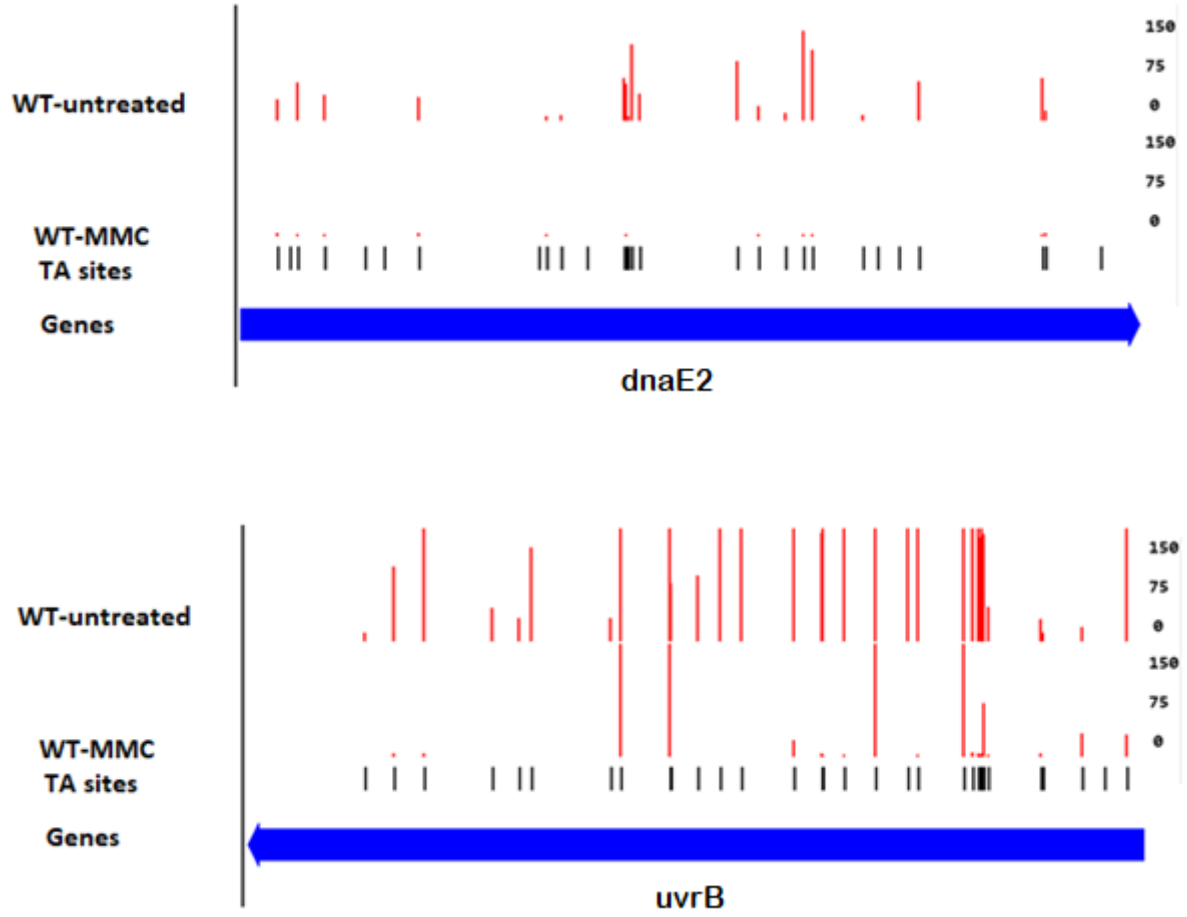


Figure 4. 5 Tn insertions counts in *dnaE2* and *uvrB* genes in the WT_MMC Tn libraries. Thick blue arrows show the gene of interest and the short black lines show the positions of -TA- dinucleotides in the gene. The red lines show Tn insertion counts at each potential -TA- site in the gene. Numbers on the right show the magnitude of Tn insertions per -TA- site. Few or no insertions under a specific condition indicates conditional gene essentiality. Images were generated using TRANSIT software. TOP: Tn insertions in *dnaE2* for untreated and WT_ MMC Tn libraries. BOTTOM: Tn insertions in *uvrB* for untreated WT and WT_MMC Tn library.

We also expected NER to compensate the loss of DnaE2 in the *dnaE2* mutant library, a prediction that was confirmed by the sizeable decrease in Tn insertion counts in *uvrB* in the Δ*dnaE2* library relative to WT (**Figure 4.6, TOP**). The converse was expected to be true for

dnaE2 in the Δ*uvrB* background. Surprisingly, however, similarities in the frequency and distribution of Tn insertions in the *dnaE2* gene before and after treatment of Δ*uvrB* library with MMC (**Figure 4.6, BOTTOM**) implied that DnaE2 does not compensate for the loss of UvrB.

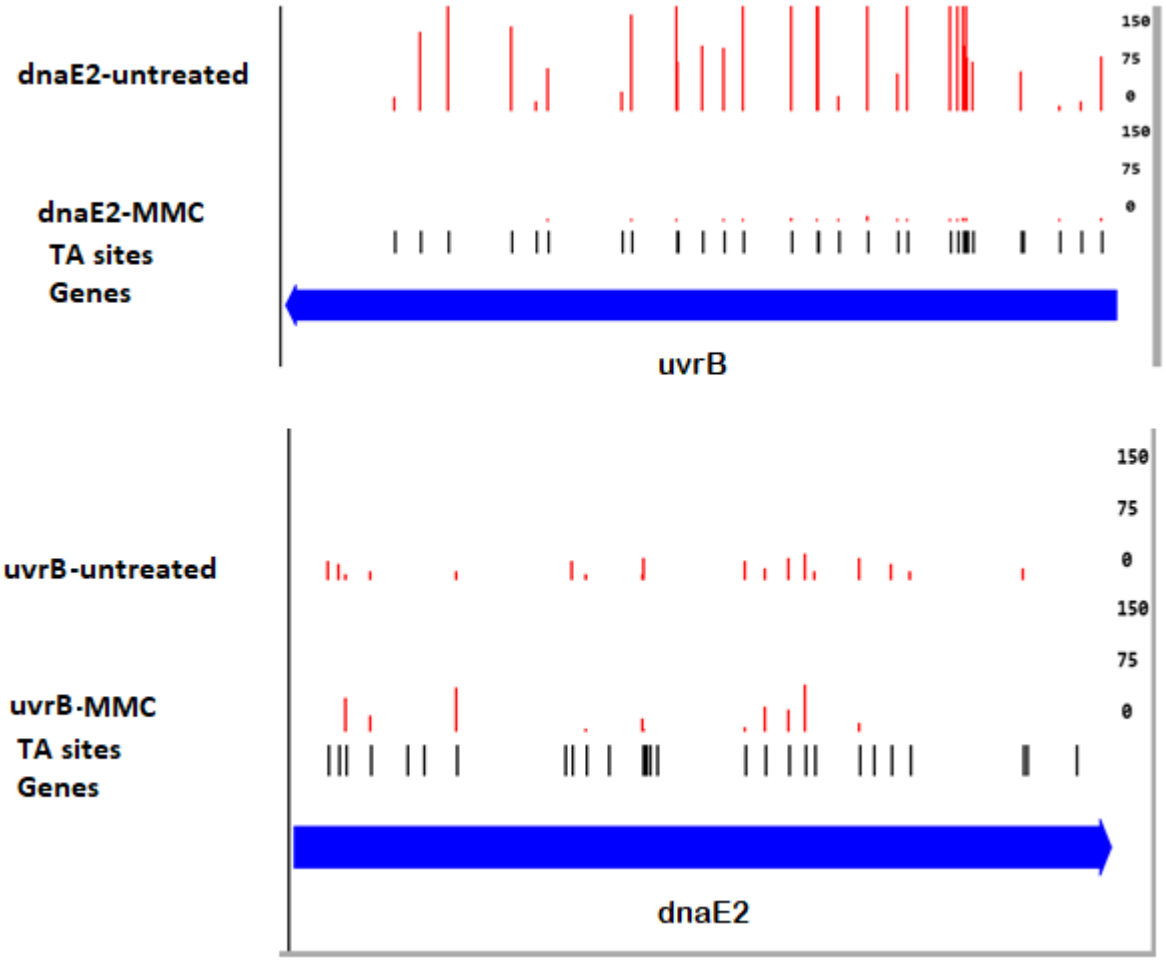


Figure 4. 6 Differential essentialities of *uvrB* and *dnaE2* in the different mutant libraries. Thick blue arrows show the genes of interest and the short black lines show the positions of -TA- dinucleotides in the gene. The red lines show Tn insertion counts at each potential -TA- site in the gene. Numbers on the right show the magnitude of Tn insertions per -TA- site. Few or no insertions under a specific condition indicates conditional gene essentiality. Images were generated using TRANSIT software. TOP: Tn insertions in *uvrB* of untreated and Δ*dnaE2*_MMC Tn libraries. BOTTOM: Tn insertion in *dnaE2* in untreated and Δ*uvrB*_MMC Tn libraries.

Table 4. 2 Conditionally essential genes in WT_MMC. A total of 73 genes were found to be conditionally essential. **All** conditionally essential genes **shown here had** adjusted P value ≤0.05 and delta mean of < 0 according to TRANSIT software analysis (statistical analysis results are shown in Supplementary data 2). Genes shared with all strains (WT and the 2 mutants) under MMC treatment are highlighted in blue, genes shared only with Δ*dnaE2*_MMC are highlighted in green, genes shared only with Δ*uvrB*_MMC are highlighted in red and genes unique to WT_MMC are highlighted in yellow.

Orf #	Name	Description
MSMEG_0786	-	serine/threonine protein kinase
MSMEG_0788	-	hypothetical protein MSMEG_0788
MSMEG_1654	-	isocitrate dehydrogenase, NADP-dependent
MSMEG_2090	-	putative cell division protein FtsX
MSMEG_3641	-	hypothetical protein MSMEG_3641
MSMEG_4269	asnB	asparagine synthase (glutamine-hydrolyzing)
MSMEG_5041	-	acyltransferase
MSMEG_5149	pimE	mannosyltransferase
MSMEG_5534	pcrA	ATP-dependent DNA helicase PcrA
MSMEG_6363	-	cysteine desulfurase family protein
MSMEG_6398	-	antigen 85-A
MSMEG_5782	pstS	phosphate ABC transporter, phosphate-binding protein PstS
MSMEG_6387	-	arabinosyltransferase A
MSMEG_4243	-	serine/threonine-protein kinase PK-1
MSMEG_5781	pstC	phosphate ABC transporter, permease protein PstC
MSMEG_0386	-	NAD dependent epimerase/dehydratase family protein
MSMEG_0035	-	FHA domain-containing protein
MSMEG_4245	-	hypothetical protein MSMEG_4245
MSMEG_3187	-	acyltransferase domain-containing protein
MSMEG_3839	-	DNA polymerase I
MSMEG_0866	-	DNA or RNA helicase of superfamily protein II
MSMEG_0031	-	penicillin binding protein transpeptidase domain-containing protein
MSMEG_0034	-	FHA domain-containing protein
MSMEG_0234	-	metallopeptidase
MSMEG_0382	-	putative transport protein
MSMEG_0400	-	peptide synthetase
MSMEG_0401	-	putative non-ribosomal peptide synthase
MSMEG_0402	-	linear gramicidin synthetase subunit D
MSMEG_1757	-	DEAD/DEAH box helicase
MSMEG_1809	-	putative thiosulfate sulfurtransferase
MSMEG_1929	-	hypothetical protein MSMEG_1929
MSMEG_1943	-	ATP-dependent DNA helicase
MSMEG_1959	-	hypothetical protein MSMEG_1959
MSMEG_3078	uvrC	excinuclease ABC subunit C
MSMEG_3808	uvrA	excinuclease ABC subunit A
MSMEG_3816	uvrB	excinuclease ABC subunit B

MSMEG_3888	-	hypothetical protein MSMEG_3888
MSMEG_5447	-	dolichyl-phosphate-mannose-protein mannosyltransferase
MSMEG_5693	-	transporter, major facilitator family protein
MSMEG_5225	-	hypothetical protein MSMEG_5225
MSMEG_5416	-	LpqU protein
MSMEG_2369	-	LppZ protein
MSMEG_6160	-	ATP-dependent rna helicase, dead/deah box family protein
MSMEG_6939	-	Soj family protein
MSMEG_0314	zwf	glucose-6-phosphate 1-dehydrogenase
MSMEG_3493	-	putative secreted protein
MSMEG_4323	aceE	pyruvate dehydrogenase subunit E1
MSMEG_5468	-	hypothetical protein MSMEG_5468
MSMEG_5780	pstA	phosphate ABC transporter, permease protein PstA
MSMEG_3146	-	invasin 1
MSMEG_4724	-	oligoribonuclease
MSMEG_5908	-	acyl-CoA synthetase
MSMEG_0381	-	Mmp14a protein
MSMEG_2752	-	RNA polymerase sigma factor SigB
MSMEG_3890	-	proteasome component
MSMEG_3902	-	ATPase, AAA family protein
MSMEG_1633	dnaE2	error-prone DNA polymerase
MSMEG_6138	-	metallopeptidase
MSMEG_6201	-	transglycosylase
MSMEG_6410	-	Rieske 2Fe-2S family protein
MSMEG_0846	-	putative monovalent cation/H+ antiporter subunit D
MSMEG_3654	-	preprotein translocase subunit SecA
MSMEG_6087	-	beta-lactamase
MSMEG_0032	-	cell cycle protein, FtsW/RodA/SpoVE family protein
MSMEG_0403	-	integral membrane protein
MSMEG_3884	-	hypothetical protein MSMEG_3884
MSMEG_6283	-	FAD binding domain-containing protein
MSMEG_4483	-	deoxyguanosinetriphosphate triphosphohydrolase-like protein
MSMEG_5423	mfd	transcription-repair coupling factor

Orf: Open reading frame; Delta mean is the variation in Tn insertions before and after treatment for each gene. A negative delta mean value signifies conditional essentiality of a gene in the treated condition.

Table 4. 3 Conditionally essential genes in Δ*dnaE2*_MMC. A total of 92 genes were found to be conditionally essential after exposure of the Δ*dnaE2* library to MMC. **All** conditionally essential genes **shown here had** adjusted P value ≤0.05 and delta mean of <0 according to TRANSIT software analysis (statistical analysis results are shown in Supplementary data 2). Genes shared with all strains are highlighted in blue, genes shared only with WT_MMC are highlighted in green, genes shared with Δ*uvrB*_MMC are highlighted in red and genes unique to Δ*dnaE2*_MMC are highlighted in yellow**.**

Orf #	Name	Description
MSMEG_0786	-	serine/threonine protein kinase
MSMEG_0788	-	hypothetical protein MSMEG_0788
MSMEG_4269	asnB	asparagine synthase (glutamine-hydrolyzing)
MSMEG_1654	-	isocitrate dehydrogenase, NADP-dependent
MSMEG_2090	-	putative cell division protein FtsX
MSMEG_4243	-	serine/threonine-protein kinase PK-1
MSMEG_3641	-	hypothetical protein MSMEG_3641
MSMEG_5041	-	acyltransferase
MSMEG_5149	pimE	mannosyltransferase
MSMEG_5534	pcrA	ATP-dependent DNA helicase PcrA
MSMEG_6363	-	cysteine desulfurase family protein
MSMEG_6387	-	arabinosyltransferase A
MSMEG_6398	-	antigen 85-A
MSMEG_4245	-	hypothetical protein MSMEG_4245
MSMEG_5781	pstC	phosphate ABC transporter, permease protein PstC
MSMEG_5782	pstS	phosphate ABC transporter, phosphate-binding protein PstS
MSMEG_0035	-	FHA domain-containing protein
MSMEG_0386	-	NAD dependent epimerase/dehydratase family protein
MSMEG_0031	-	penicillin binding protein transpeptidase domain-containing protein
MSMEG_0034	-	FHA domain-containing protein
MSMEG_0234	-	metallopeptidase
MSMEG_0314	zwf	glucose-6-phosphate 1-dehydrogenase
MSMEG_0382	-	putative transport protein
MSMEG_0400	-	peptide synthetase
MSMEG_1757	-	DEAD/DEAH box helicase
MSMEG_1809	-	putative thiosulfate sulfurtransferase
MSMEG_1943	-	ATP-dependent DNA helicase
MSMEG_1959	-	hypothetical protein MSMEG_1959
MSMEG_3078	uvrC	excinuclease ABC subunit C
MSMEG_3808	uvrA	excinuclease ABC subunit A
MSMEG_3816	uvrB	excinuclease ABC subunit B
MSMEG_3888	-	hypothetical protein MSMEG_3888
MSMEG_4323	aceE	pyruvate dehydrogenase subunit E1
MSMEG_5225	-	hypothetical protein MSMEG_5225
MSMEG_5447	-	dolichyl-phosphate-mannose-protein mannosyltransferase
MSMEG_5468	-	hypothetical protein MSMEG_5468

MSMEG_5693	-	transporter, major facilitator family protein
MSMEG_5780	pstA	phosphate ABC transporter, permease protein PstA
MSMEG_6939	-	Soj family protein
MSMEG_1929	-	hypothetical protein MSMEG_1929
MSMEG_5416	-	LpqU protein
MSMEG_3146	-	invasin 1
MSMEG_6160	-	ATP-dependent rna helicase, dead/deah box family protein
MSMEG_0401	-	putative non-ribosomal peptide synthase
MSMEG_3493	-	putative secreted protein
MSMEG_4724	-	oligoribonuclease
MSMEG_2369	-	LppZ protein
MSMEG_0402	-	linear gramicidin synthetase subunit D
MSMEG_2403	recG	ATP-dependent DNA helicase RecG
MSMEG_0690	-	iron-sulfur cluster-binding protein
MSMEG_5470	-	molybdopterin biosynthesis protein MoeA 1
MSMEG_4916	-	alpha-amylase family protein
MSMEG_0406	-	acyl-CoA-dehydrogenase
MSMEG_0860	-	CDP-diacylglycerol--serine O-phosphatidyltransferase
MSMEG_2418	rnc	ribonuclease III
MSMEG_2934	-	lipid A biosynthesis lauroyl acyltransferase
MSMEG_3021	-	recombination factor protein RarA
MSMEG_3147	moxR	ATPase, MoxR family protein
MSMEG_3222	-	prolipoprotein diacylglyceryl transferase
MSMEG_3873	-	cobalamin biosynthesis protein cobIJ
MSMEG_4259	-	hypothetical protein MSMEG_4259
MSMEG_4853	-	peptidase, M24 family protein
MSMEG_4979	cysD	sulfate adenylyltransferase subunit 2
MSMEG_6099	-	transmembrane protein rich in alanine
MSMEG_5004	-	DNA repair exonuclease
MSMEG_4235	mraW	S-adenosyl-methyltransferase MraW
MSMEG_4237	-	transmembrane protein
MSMEG_5049	kgd	alpha-ketoglutarate decarboxylase
MSMEG_5601	-	hypothetical protein MSMEG_5601
MSMEG_6143	-	hypothetical protein MSMEG_6143
MSMEG_6596	-	hypothetical protein MSMEG_6596
MSMEG_2868	-	transcriptional regulator, PadR family protein
MSMEG_5072	-	RNA polymerase sigma factor SigE
MSMEG_5121	-	N-succinyldiaminopimelate aminotransferase
MSMEG_6079	radA	DNA repair protein RadA
MSMEG_3069	-	aminoglycosides/tetracycline-transport integral membrane protein
MSMEG_4491	recO	DNA repair protein RecO
MSMEG_5545	-	formamidopyrimidine-DNA glycosylase
MSMEG_5779	pstB	phosphate ABC transporter, ATP-binding protein
MSMEG_0674	-	ErfK/YbiS/YcfS/YnhG family protein
MSMEG_1527	truA	tRNA pseudouridine synthase A
MSMEG_1876	-	lipoprotein LpqB

MSMEG_2503	-	hypothetical protein MSMEG_2503
MSMEG_3100	opcA	OpcA protein
MSMEG_6284	-	cyclopropane-fatty-acyl-phospholipid synthase
MSMEG_0233	-	lipoprotein Lpps
MSMEG_4627	ndk	nucleoside diphosphate kinase
MSMEG_4818	-	putative cytochrome P450 123
MSMEG_5002	-	hypothetical protein MSMEG_5002
MSMEG_6940	gidB	methyltransferase GidB

Orf: Open reading frame; Delta mean is the variation in Tn insertions before and after treatment for each gene. A negative delta mean value signifies conditional essentiality of a gene in the treated condition.

Table 4. 4 Conditionally essential genes in Δ*uvrB*_MMC. A total of 37 genes were found to be conditionally essential after treatment of Δ*uvrB* library with MMC. **All** conditionally essential genes **shown here had** adjusted P value of ≤0.05 and delta mean of < 0 according to TRANSIT software analysis (statistical analysis results are shown in Supplementary data 2). Genes shared with all strains are highlighted in blue, genes shared only with WT_MMC are highlighted in green, genes shared with Δ*dnaE2*_MMC are highlighted red and genes unique to Δ*uvrB*_MMC highlighted in yellow.

Orf #	Name	Description
MSMEG_0786	-	serine/threonine protein kinase
MSMEG_1654	-	isocitrate dehydrogenase, NADP-dependent
MSMEG_2090	-	putative cell division protein FtsX
MSMEG_4243	-	serine/threonine-protein kinase PK-1
MSMEG_4269	asnB	asparagine synthase (glutamine-hydrolyzing)
MSMEG_3641	-	hypothetical protein MSMEG_3641
MSMEG_5041	-	acyltransferase
MSMEG_5149	pimE	mannosyltransferase
MSMEG_5534	pcrA	ATP-dependent DNA helicase PcrA
MSMEG_5781	pstC	phosphate ABC transporter, permease protein PstC
MSMEG_6363	-	cysteine desulfurase family protein
MSMEG_6387	-	arabinosyltransferase A
MSMEG_6398	-	antigen 85-A
MSMEG_4245	-	hypothetical protein MSMEG_4245
MSMEG_0788	-	hypothetical protein MSMEG_0788
MSMEG_0035	-	FHA domain-containing protein
MSMEG_5782	pstS	phosphate ABC transporter, phosphate-binding protein PstS
MSMEG_0386	-	NAD dependent epimerase/dehydratase family protein
MSMEG_2403	recG	ATP-dependent DNA helicase RecG
MSMEG_4916	-	alpha-amylase family protein
MSMEG_5470	-	molybdopterin biosynthesis protein MoeA 1
MSMEG_0690	-	iron-sulfur cluster-binding protein

MSMEG_0866	-	DNA or RNA helicase of superfamily protein II
MSMEG_3187	-	acyltransferase domain-containing protein
MSMEG_3839	-	DNA polymerase I
MSMEG_3494	-	putative secreted protein
MSMEG_2772	-	amino acid permease
MSMEG_3886	tatC	twin arginine-targeting protein translocase TatC
MSMEG_0777	-	F420-dependent glucose-6-phosphate dehydrogenase
MSMEG_4189	cysS	cysteinyl-tRNA synthetase
MSMEG_4918	glgB	glycogen branching enzyme
MSMEG_1930	-	DEAD/DEAH box helicase
MSMEG_4900	-	Pks14 protein
MSMEG_5672	gltA	type II citrate synthase
MSMEG_0840	-	hypothetical protein MSMEG_0840
MSMEG_3027	-	hypothetical protein MSMEG_3027

Orf: Open reading frame; Delta mean is the variation in Tn insertions before and after treatment for each gene. A negative delta mean value signifies conditional essentiality of a gene in the treated condition.

4.3.4.1. Taking a closer look at MMC-essential genes

Conditionally essential genes were further classified into functional groups. In all libraries, most genes belonged to cell wall and cell processes and intermediary metabolism and respiration functional groups (**Figure 4.7)**. In the WT library, genes in cell wall and cellular processes were abundant. Intermediary metabolism and respiration genes were more abundant in the Δ*dnaE2* and Δ*uvrB* Tn libraries (**Figure 4.7**). Mannosyltransferase, *pimE*, putative cell division protein, *ftsX* and phosphate ABC transporter genes were some of the cell wall and cell processes genes that appeared in all the libraries. Asparagine synthase, *asnB*, is an intermediary metabolism and respiration gene that also appeared in all the libraries. Since the treatment that we used directly damages DNA, we were also interested in the information pathways functional group (**Table 4.5**) which contains DNA repair genes. Genes in the information pathways were the third most highly represented in WT_MMC and Δ*dnaE2*_MMC libraries (**Figure 4.7**). Apart from NER and TLS, some of the information pathways genes represented included DNA polymerase I which was identified in WT and Δ*uvrB* libraries and the ATP-dependent DNA helicase, RecG, which was identified in the Δ*uvrB* and Δ*dnaE2* libraries (**Table 4.5**). Overall, we observed an overrepresentation of helicases across the libraries in the information pathway genes. Furthermore, DNA repair pathways most represented in our data were NER, TLS and HR (**Table 4.5**).

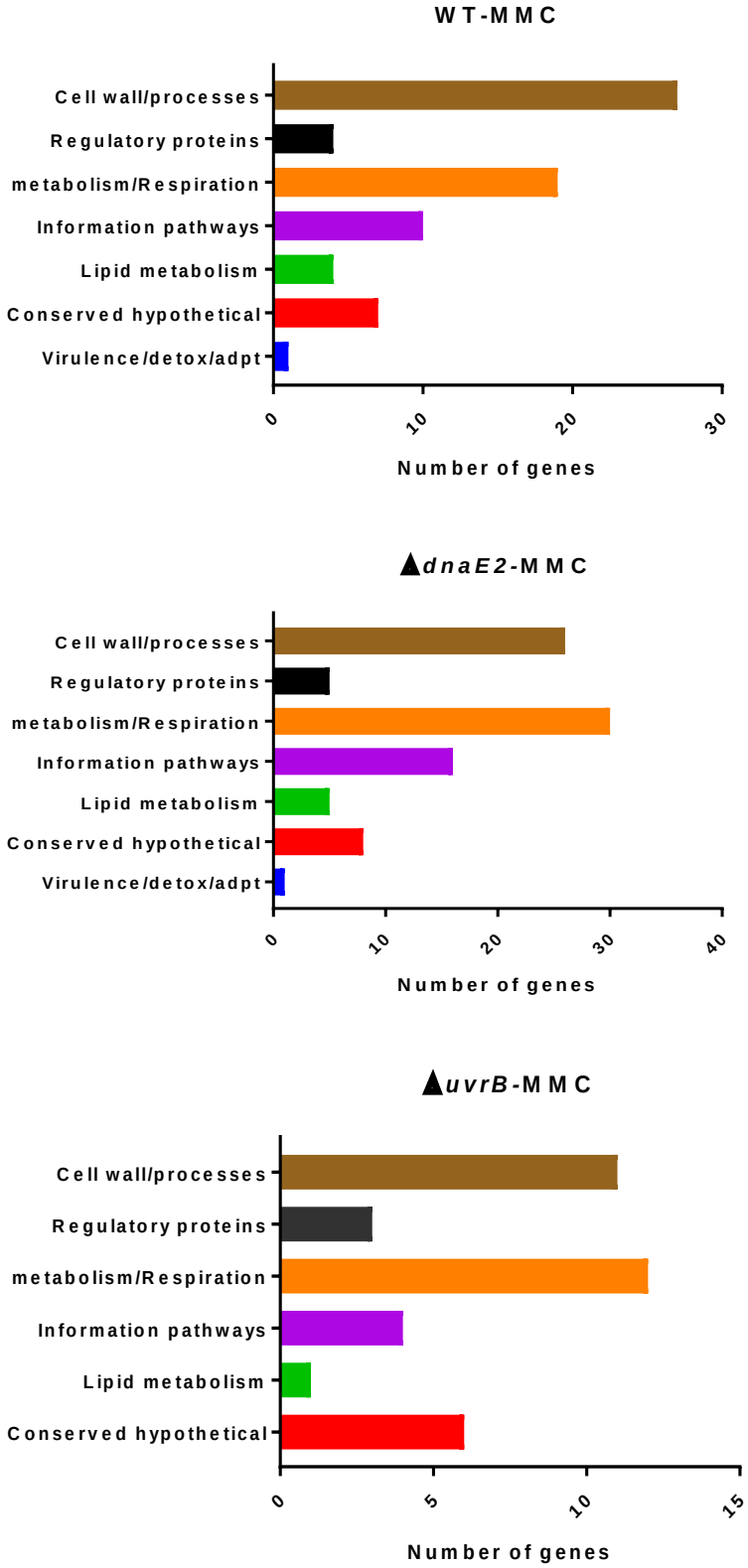


Figure 4. 7 Functional classification of "MMC essential" genes in WT, Δ*uvrB* and Δ*dnaE2* Tn libraries. MTB orthologs of all the conditionally essential genes were first identified using Mycobrowser (https://mycobrowser.epfl.ch/). The classifications of genes with MTB orthologs were based on functional classes as they appear in the Mycobrowser

database. For those genes that did not have MTB orthologs, functional classification was inferred from UniProt database: https://www.uniprot.org/.

Table 4. 5 Genes that belong to the information pathway functional group in MMC-treated libraries. Genes shared with all strains (WT and the 2 mutants) under MMC treatment are highlighted in blue, genes shared between Δ*dnaE2*_MMC and WT_MMC are highlighted in green, genes shared between Δ*uvrB*_MMC and Δ*dnaE2*_MMC are highlighted in red and genes shared between WT_MMC and Δ*uvrB*_MMC are highlighted in purple and strain unique genes are highlighted in yellow.

WT_MMC			
ORF	*Name*	*Description*	*Biological process*
MSMEG_1757	-	DEAD/DEAH box helicase	Helicase activity
MSMEG_3078	*uvrC*	excinuclease ABC subunit C	NER
MSMEG_3808	*uvrA*	excinuclease ABC subunit A	NER
MSMEG_3816	*uvrB*	excinuclease ABC subunit B	NER
MSMEG_3839	-	DNA polymerase I	DNA replication/repair
MSMEG_5534	*pcrA*	ATP-dependent DNA helicase PcrA	DNA replication
MSMEG_2752	-	RNA polymerase sigma factor SigB	Transcription
MSMEG_6160	-	ATP-dependent rna helicase, dead/deah box family protein	Helicase activity
MSMEG_5423	*mfd*	transcription-repair coupling factor	NER
MSMEG_1633	*dnaE2*	error-prone DNA polymerase	Trans-lesion synthesis
Δ*dnaE2*_MMC			
MSMEG_1943	-	ATP-dependent DNA helicase	Helicase activity
MSMEG_2418	*rnc*	ribonuclease III	RNA processing
MSMEG_4491c	*recO*	DNA repair protein RecO	Recombination
MSMEG_1527	*truA*	tRNA pseudouridine synthase A	tRNA pseudouridine synthesis
MSMEG_3021	-	recombination factor protein RarA	DNA replication/repair
MSMEG_1757c	-	DEAD/DEAH box helicase	Helicase activity
MSMEG_3078	*uvrC*	excinuclease ABC subunit C	NER
MSMEG_3808c	*uvrA*	excinuclease ABC subunit A	NER
MSMEG_3816c	*uvrB*	excinuclease ABC subunit B	NER
MSMEG_2403	*recG*	ATP-dependent DNA helicase RecG	Recombination

MSMEG_5534c	*pcrA*	ATP-dependent DNA helicase PcrA			DNA replication
Δ*uvrB*_MMC					
MSMEG_2403	*recG*	ATP-dependent DNA helicase RecG			Recombination
MSMEG_3839c	-	DNA polymerase I			DNA replication/repair
MSMEG_5534c	*pcrA*	ATP-dependent DNA helicase PcrA			DNA replication
MSMEG_1930c	-	DEAD/DEAH box helicase			Helicase activity

NER: Nucleotide excision repair

4.3.4.2. Gene essentiality and gene expression under MMC treatment

A recent transcriptomic analysis of MSM gene expression under MMC treatment showed that genes which are involved in replication, repair and recombination were significantly upregulated during MMC treatment (Müller et al., 2018). Most of the genes identified in that study did not emerge as "essential" in our MMC-treated WT Tn library: a comparison of our Tn-Seq data with expression profiling performed in MMC treated MSM showed little correlation between gene expression under MMC treatment and requirement for survival. Of the 270 genes which were upregulated under MMC in MSM *(*Müller *et al.,* 2018*)*, only eight were found to be conditionally essential in our WT_MMC data, all of which are involved in DNA repair and replication (**Table 4.6**).

Table 4. 6 TnSeq WT _MMC conditionally essential genes that were also identified as upregulated in RNA seq MMC data by Muller et al., 2018.

ORF	Name	Description
MSMEG_1633	*dnaE2*	error-prone DNA polymerase
MSMEG_1757	-	DEAD/DEAH box helicase
MSMEG_1943	-	ATP-dependent DNA helicase
MSMEG_3078	*uvrC*	excinuclease ABC subunit C
MSMEG_3808	*uvrA*	excinuclease ABC subunit A
MSMEG_3816	*uvrB*	excinuclease ABC subunit B
MSMEG_5534	*pcrA*	ATP-dependent DNA helicase PcrA
MSMEG_3839	-	DNA polymerase I

4.3.5. CRISPRi validation of MMC conditionally essential genes

TnSeq is a screening method whose accuracy and predictive power can be affected by several factors including library saturation, sequencing depth and normalisation of reads counts (Chao et al., 2016). Therefore, gene function leads must be followed up with other, often lower-throughput approaches; *e.g.*, creating targeted gene deletions using allelic exchange mutagenesis. To validate the TnSeq data generated in this study the recently published ATc-inducible CRISPRi technology (Rock et al., 2017) was employed to target *pstA*, one of the genes identified as conditionally essential under MMC treatment in WT and Δ*dnaE2* libraries. The *pstA* gene is a member of the phosphate specific transporter system. While MTB has several copies of the phosphate specific transporter (Pst) system, MSM only has a single copy organised as a *pstSCAB* operon (Gebhard et al., 2006). Most genes within this operon were conditionally essential in MMC-treated libraries (**Tables 4.2, 4.3 and 4.4**) however, the TnSeq screen did not identify *pstA* as conditionally essential in a *uvrB*_MMMC library as it did not achieve a $P_{adjusted}$ value cut off of ≤0.05.

As a control, we first investigated MSMEG_4259, a probable DnaQ-UvrC fusion protein, which has been shown to be essential in WT MSM under MMC treatment (Griffault, Ditse **et al**., personal communication).The TnSeq data indicated that MSMEG_4259 was essential under MMC treatment in the *dnaE2* KO strain with a $P_{adjusted}$ value of 0 and a dramatic decrease in the number of Tn insertions (**Figure 4.8**). In the WT library, MSMEG_4259 had an adjusted P value of 0.085 which was outside our cut off ≤0.05; however, closer inspection of the Tn trace showed a clear decrease in Tn insertions (**Figure 4.8**). Interestingly, in the Δ*uvrB* mutant, MSMEG_4259 was non-essential with a $P_{adjusted}$ value of 1 and with comparable Tn insertions between treated and untreated libraries (**Figure 4.8**). CRISPRi validated the essentiality of MSMEG_4259 in both WT and Δ*dnaE2* backgrounds: no growth was observed in the MSMEG_4259 CRISPRi mutants on MMC supplemented 7H10 solid media after ATc induced silencing of MSMEG_4259. In contrast, MSMEG_4259 was confirmed as non-essential in the Δ*uvrB* mutant, as shown by bacillary growth after ATc induction (**Figure 4.8**).

Having demonstrated the utility of CRISPRi with MSMEG_4259, there was confidence in its utility to validate the TnSeq data. For *pstA*, CRISPRi knockdown mutants could not grow on MMC supplemented 7H10 solid medium (**Figure 4.9**) in all three genetic backgrounds (WT,

Δ*dnaE2*, Δ*uvrB*). This was not surprising because even though *pstA* was not identified as conditionally essential in a *uvrB*_MMC library statistically, a closer inspection of the Tn traces showed an apparent decrease in Tn insertions in this gene in a Δ*uvrB*_MMC when compared to a uvrB-untreated Tn library (**Figure 4.9**). Therefore, the CRISPRi results recapitulated the TnSeq observations.

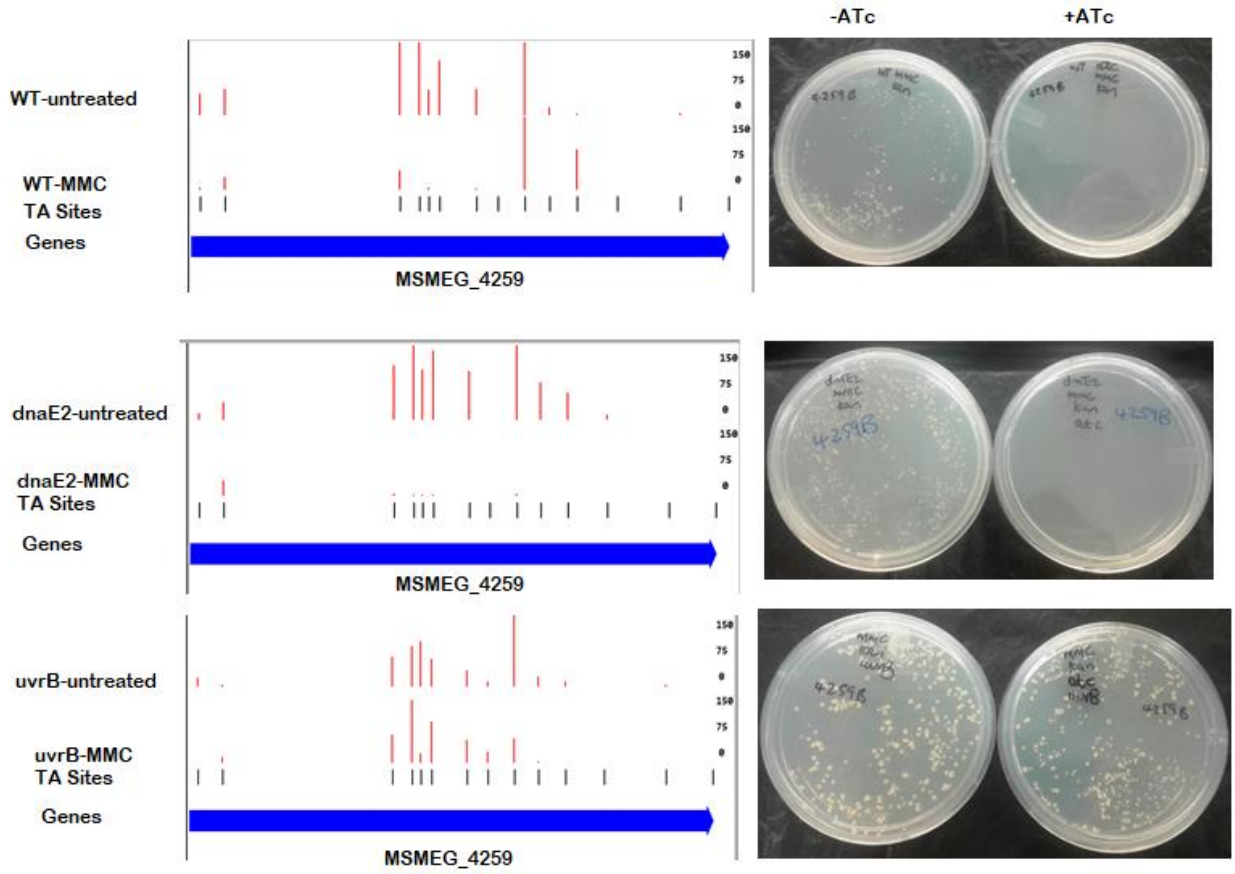


Figure 4. 8 Validation of conditional essentiality of MSMEG_4259 (*dnaQ-uvrC*) under MMC treatment by CRISPRi. Distribution of Tn insertions in MSMEG_4259 for each library are shown with the corresponding CRISPRi validation plates. For Tn insertion images (LEFT), the thick blue arrow shows the gene of interest and the red lines shows Tn insertion in the gene. The short black lines show the positions of -TA- dinucleotides in the gene. Numbers on the right show the frequencies of Tn insertions per TA site. Few or no insertions in a condition indicates essentiality. Images were generated using TRANSIT software. For CRISPRi validation (RIGHT), sgRNAs targeting MSMEG_4259 were co-expressed with dCas9$_{sth1}$ in WT, Δ*dnaE2* and Δ*uvrB* parental strains(Rock et al., 2017). The knockdown phenotype was monitored by plating 100 µl of cells at OD_{600} of 0.6 on 7H10 solid medium supplemented with MMC +/-ATc; absence of growth on +ATc plate indicates essentiality. The UPPER panel is WT, MIDDLE panel is Δ*dnaE2* and LOWER panel is Δ*uvrB*.

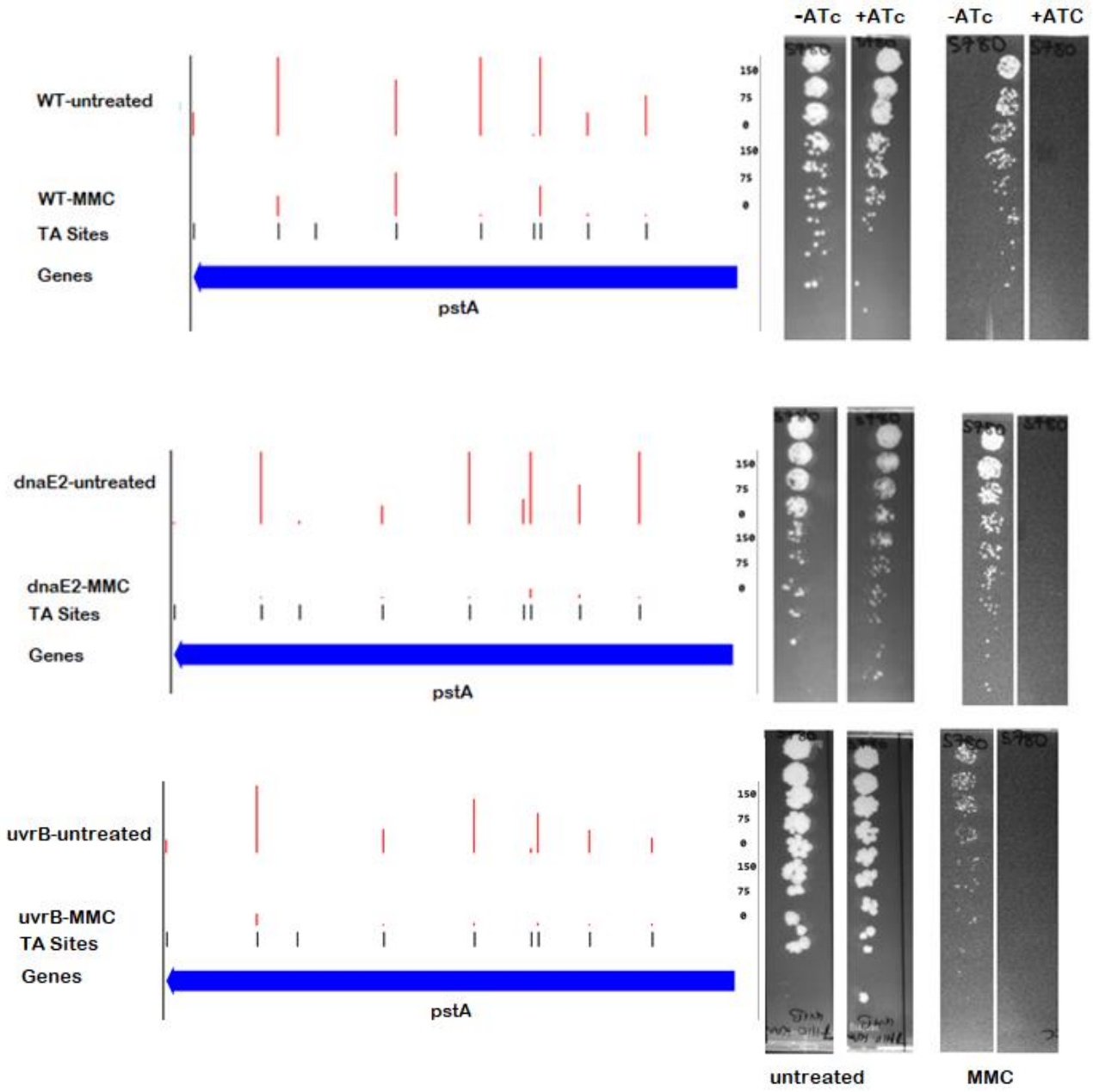


Figure 4. 9 Validation of the essentiality of *pstA* during MMC by CRISPRi. Distribution of Tn insertions in *pstA* for each library are shown with the corresponding CRISPRi validation plates. For Tn insertion images (LEFT), the thick blue arrow shows the gene of interest and the red lines show Tn insertions in the gene. The short black lines show the positions of -TA- dinucleotides in the gene. Numbers on the right show the frequency of Tn insertions per TA site. Few or no insertions in a condition indicate essentiality. Images were generated using TRANSIT software. For CRISPRi validation (RIGHT), sgRNAs targeting *pstA* were co-expressed with $dCas9_{sth1}$ in WT, Δ*dnaE2* and Δ*uvrB* parental strains. The knockdown phenotype was monitored by plating 5 µl of 10-fold serial dilutions of WT, Δ*dnaE2* and Δ*uvrB* strains which runs from undiluted 10^0 to 10^{-7} on 7H10 solid medium untreated or treated with MMC +/-ATc. The UPPER panel is WT, MIDDLE panel is Δ*dnaE2* and LOWER panel is Δ*uvrB*. Untreated panels are 7H10 plates without MMC which were included as controls and show that the gene is non-essential under standard *in vitro* conditions.

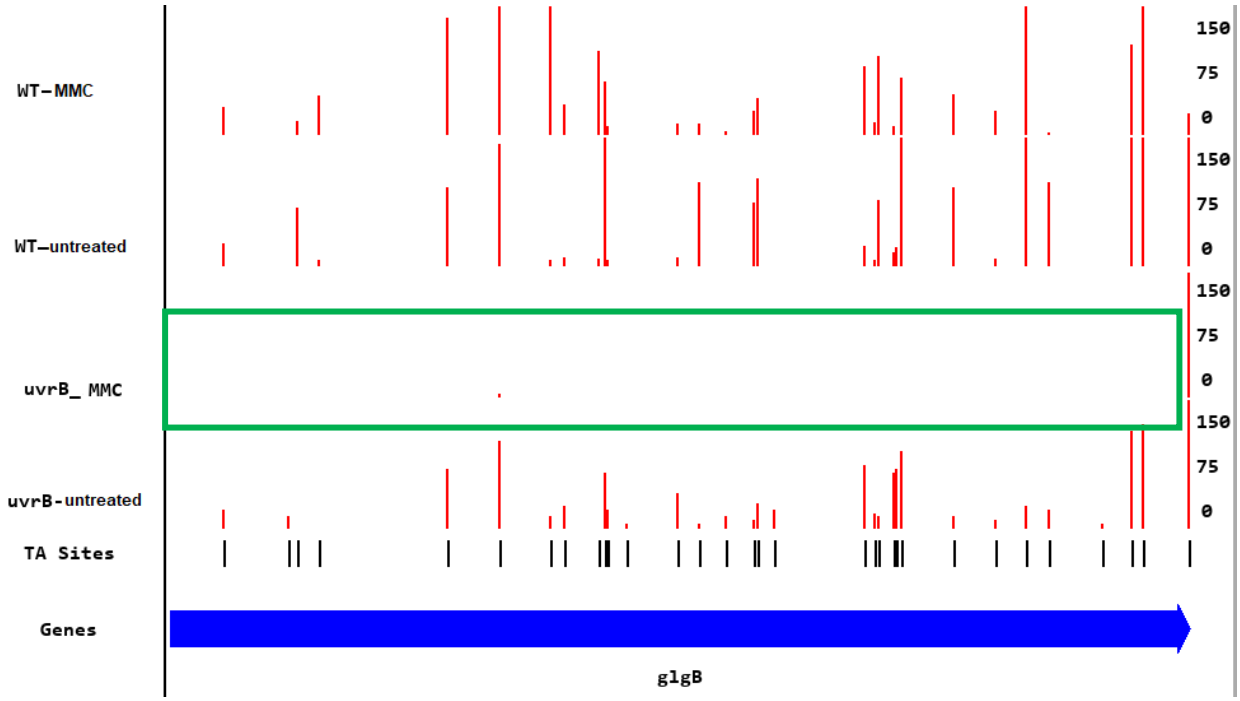
4.3.6. Genetic interaction analysis (GI): MMC

Comparing the MMC resampling results of all libraries, it was noticeable that a number of genes seemed specific for each of the three genetic backgrounds: that is, 21 genes appeared in the WT library only, 39 genes appeared only in the Δ*dnaE2* background, and 12 genes appeared only in the Δ*uvrB* background (**Figure 4.4**). While this might be expected given these are genotypically distinct strains, the possibility existed that these differences could reflect differences in library coverage. Also, the libraries had not been normalised together, instead comparisons were restricted to with/without MMC exposure (*e.g.*, WT-untreated was normalised to WT-MMC, etc.). Furthermore, while the resampling method identifies conditionally essential genes in two conditions, it does not specifically identify genetic interactions; that is, genes that interact with the genes of interest. To identify genes which interact specifically with *dnaE2* and *uvrB* under MMC, we carried out GI analysis. In this method, a library in a knockout strain is compared to WT in order to identify mutants whose abundance changes compared to WT strain. GI analysis requires a 4-way experimental design whereby WT-untreated and treated libraries are compared with mutant untreated and treated libraries (DeJesus et al., 2017b). In this analysis, a significant change in enrichment can be ascribed to differences in the genetic background (DeJesus et al., 2017b). For this research, the specific focus was on genetic interactions that might be aggravating that is, additional gene disruptions that make MSM hypersusceptible to MMC in either Δ*dnaE2* or Δ*uvrB* backgrounds.

Aggravating genetic interactions were identified as those whose delta-log fold change (delta-logFC) was significantly different from 0 (adjusted probability of falling within a region of practical equivalence (ROPE) was ≤0.05, (DeJesus et al., 2017b)). In this analysis, 9 genes were identified that had aggravating genetic interactions with *uvrB* (**Table 4.7**). Most of these were involved in metabolic processes including glycogen biosynthesis (*glgB*), mycothiol biosynthesis (MSMEG_0933 and MSMEG_5261), glycerol metabolism (*dhaK*) and the TCA cycle (*gltA*). The only information pathway protein was CysS which is involved in translation (**Table 4.7**). Tn insertions in some of these genes (those with the fewest observed Tn insertions) are shown in **Figure 4.10** which highlights the very clear reduction in the number of Tn insertions in *glgB*, MSMEG_5261 and CysS in the Δ*uvrB*-MMC library. However, *cysS* in the WT-untreated library also had fewer Tn insertions compared to its WT_MMC counterpart (**Figure 4.10**).

Table 4. 7 Genes exhibiting aggravating genetic interactions with *uvrB* during MMC treatment.

#ORF	Name	Description	Biological process
MSMEG_4918	*glgB*	glycogen branching enzyme	Glycogen biosynthesis
MSMEG_0933	-	hypothetical protein MSMEG_0933	Mycothiol biosynthesis
MSMEG_4189c	*cysS*	cysteinyl-tRNA synthetase	Translation
MSMEG_5261c	-	mycothiol conjugate amidase Mca	Mycothiol biosynthesis
MSMEG_2123c	*dhaK*	dihydroxyacetone kinase, DhaK subunit	Glycerol metabolism
MSMEG_4024c	-	transcriptional regulator, TetR family protein	Transcription regulator
MSMEG_2788	-	ATP/GTP-binding integral membrane protein	Conserved hypothetical
MSMEG_4900	-	Pks14 protein	Membrane protein
MSMEG_5672c	*gltA*	type II citrate synthase	TCA cycle



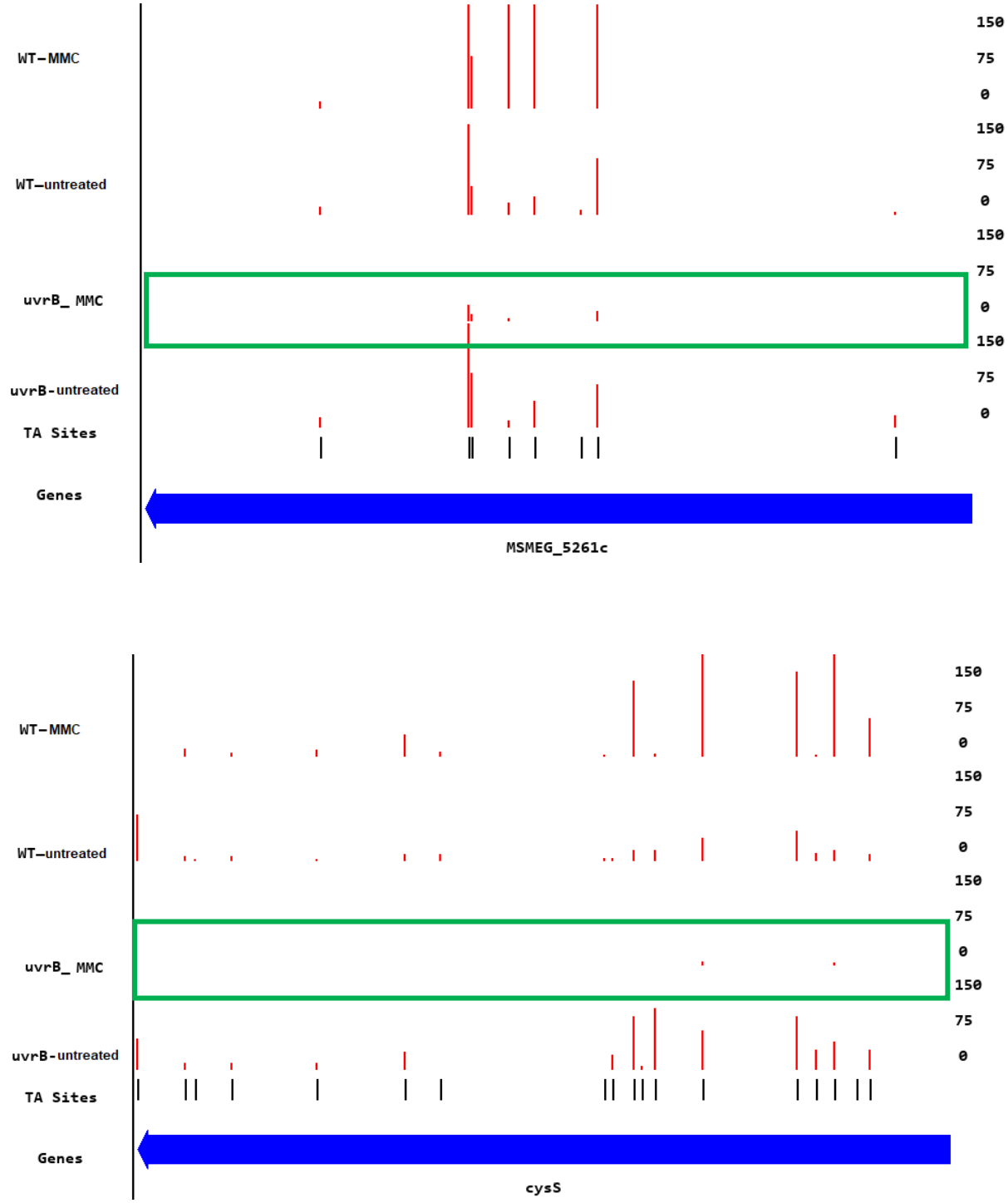
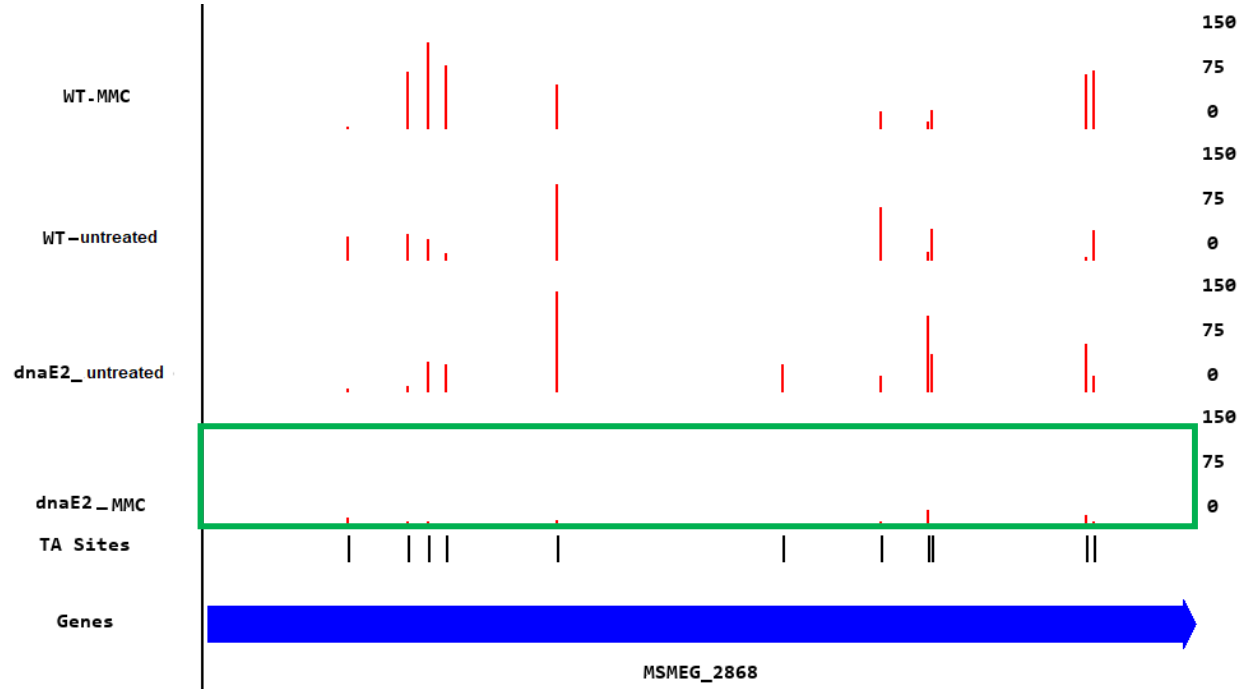


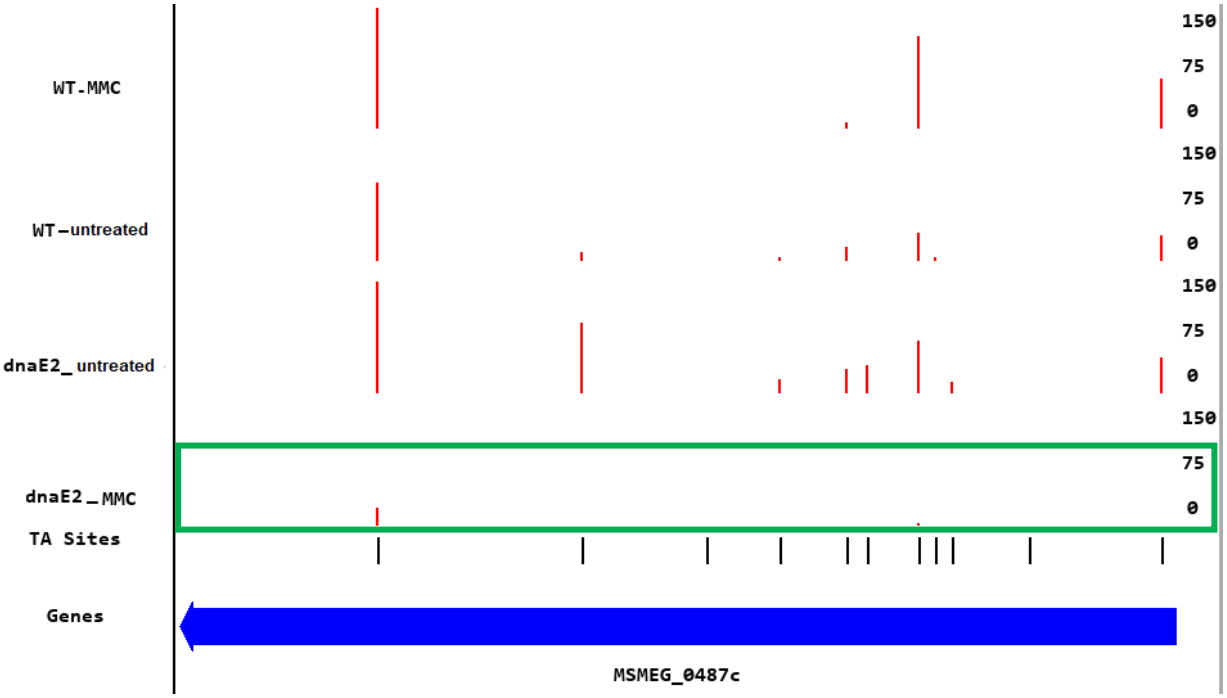
Figure 4. 10 Tn insertions in some of the genes which have aggravating genetic interactions with *uvrB* during MMC treatment. Insertions in the four libraries (WT_untreated, WT_MMC, Δ*uvrB*-untreated and Δ*uvrB*_MMC) that were analysed together are shown for each gene. Thick blue arrows indicate the gene and the red lines show Tn insertions in the -TA- sites (short black lines) of that gene. Green outlined rectangles show Tn insertions in the gene of interest in a *uvrB*-MMC library where few insertions or no insertions indicates essentiality. Numbers on the right represent the magnitude of the Tn insertions. Images were generated using TRANSIT software.

In a Δ*dnaE2*_MMC library, 6 significant aggravating genetic interactions were found (**Table 4.8**). Two of these were transcriptional regulators, PadR family and LysR family. A sizeable reduction in the number of Tn insertions was observed in MSMEG_2868 (PadR transcriptional regulator), MSMEG_0487 and MSMEG_0425 (**Figure 4.11**) MSMEG_2868 was interesting because it had the fewest Tn insertions. The data were therefore probed further to see if this pattern of Tn insertions was specific to a Δ*dnaE2*_MMC library. **Figure 4.12** shows Tn insertions in MSMEG_2868 in all MMC-treated strains. While there were almost no insertions in this gene in a Δ*dnaE2*_MMC_library, reduced Tn insertions were also observed in the Δ*uvrB* library compared to the WT (Figure 4.12). The essentiality of MSMEG_2868 in the Δ*dnaE2*_MMC library was validated by CRISPRi. **Figure 4.13** shows growth of the MSMEG_2868 CRISPRi KD mutant in untreated and MMC treated conditions. In untreated conditions, growth of bacterial cells was observed even after ATc induction of the CRISPRi system, confirming the non-essentiality of this gene in all strains under optimal conditions (Figure 4.13 Left). However, in agreement with the TnSeq data, there was no growth of the MSMEG_2868 CRISPRi mutant upon ATc induction in a Δ*dnaE2* strain while the mutant was able to grow in both Δ*uvrB* and WT strains under MMC treatment (**Figure 4.13 Right**).

Table 4. 8 Genes exhibiting aggravating genetic interactions with *dnaE2* during MMC treatment.

#ORF	Name	Description	Biological process
MSMEG_2868	-	transcriptional regulator, PadR family protein	Transcriptional regulator
MSMEG_3323	-	hypothetical protein MSMEG_3323	Conserved hypothetical
MSMEG_4510c	*mbtF*	peptide synthetase mbtf	Peptide synthase
MSMEG_0487c	-	ABC transporter permease	Transmembrane transport
MSMEG_4025c	-	transcriptional regulator, LysR family protein	Transcriptional regulator
MSMEG_4084c	-	putative acyl-CoA dehydrogenase	Oxidoreductase


WT-MMC
WT-untreated
dnaE2_untreated
dnaE2_MMC
TA Sites
Genes
MSMEG_2868
150
75
0


WT-MMC
WT-untreated
dnaE2_untreated
dnaE2_MMC
TA Sites
Genes
MSMEG_0487c
150
75
0

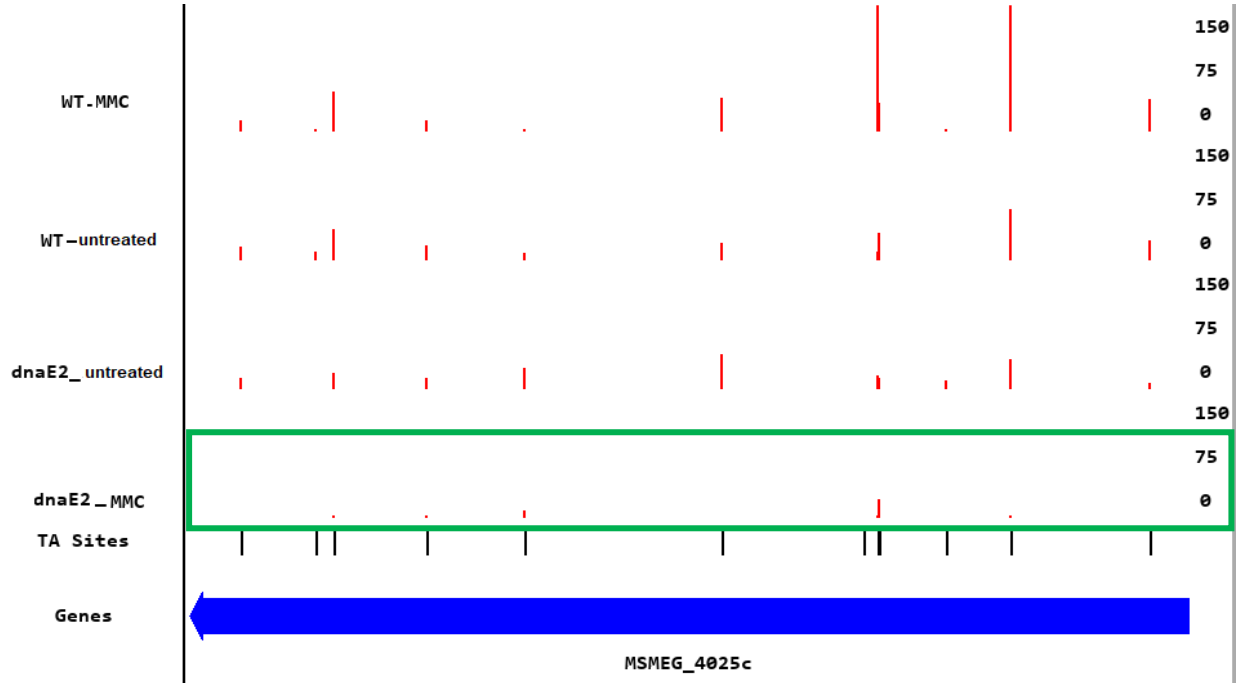


Figure 4. 11 Tn insertions in selected genes which have aggravating genetic interactions with *dnaE2* during MMC treatment. Insertions in the four libraries (WT-untreated, WT-MMC, Δ*dnaE2*-untreated and *dnaE2*-MMC) that were analysed together are shown for every gene shown. Thick blue arrows show the gene and the red lines show Tn insertion frequencies in the different -TA- sites (short black lines) of that gene; the. numbers on the right indicate the Tn insertion frequencies. Green outlined rectangles show Tn insertions in the gene of interest in a *dnaE2*-MMC library where few or no insertions indicate essentiality. Images were generated using TRANSIT software.

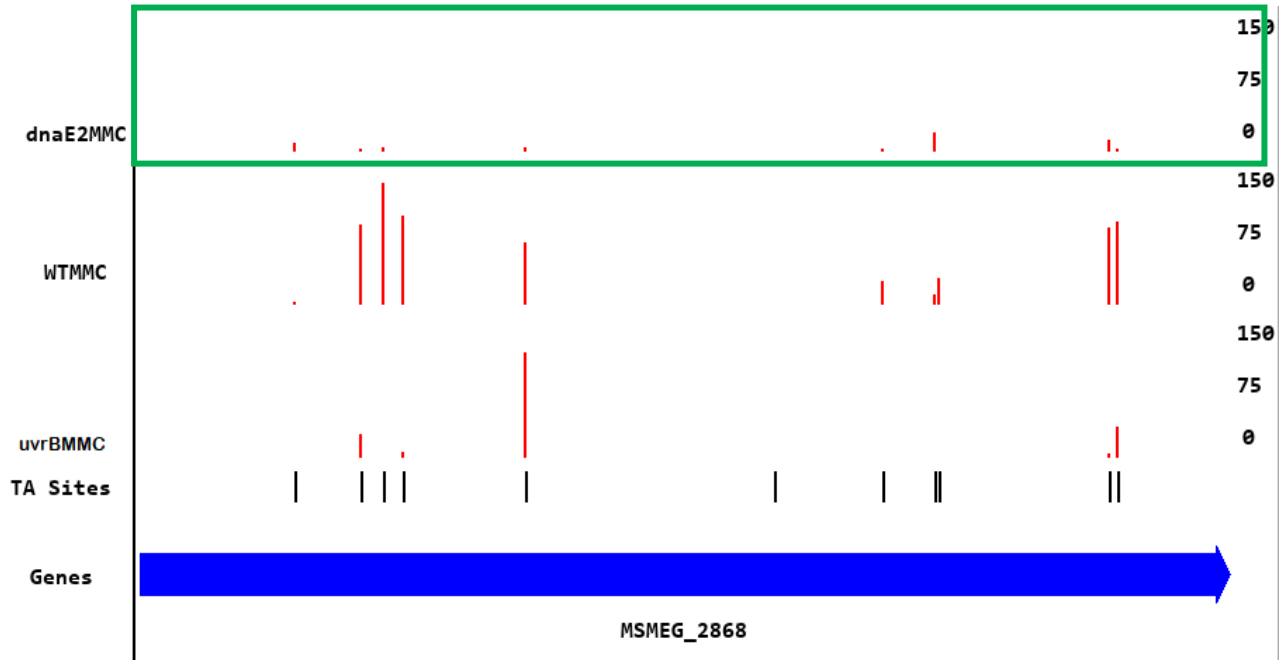


Figure 4. 12 Insertions in MSMEG_2868 in WT-MMC, Δ*uvrB*-MMC and Δ*dnaE2*-MMC libraries. Thick blue arrows show the genes and the red lines indicate Tn insertion frequencies in the different -TA- sites (short black lines). Green outlined rectangles show Tn insertions in the gene of interest in a dnaE2-MMC library. Numbers on the right represent the magnitude of the Tn insertions. Images were generated using TRANSIT software.

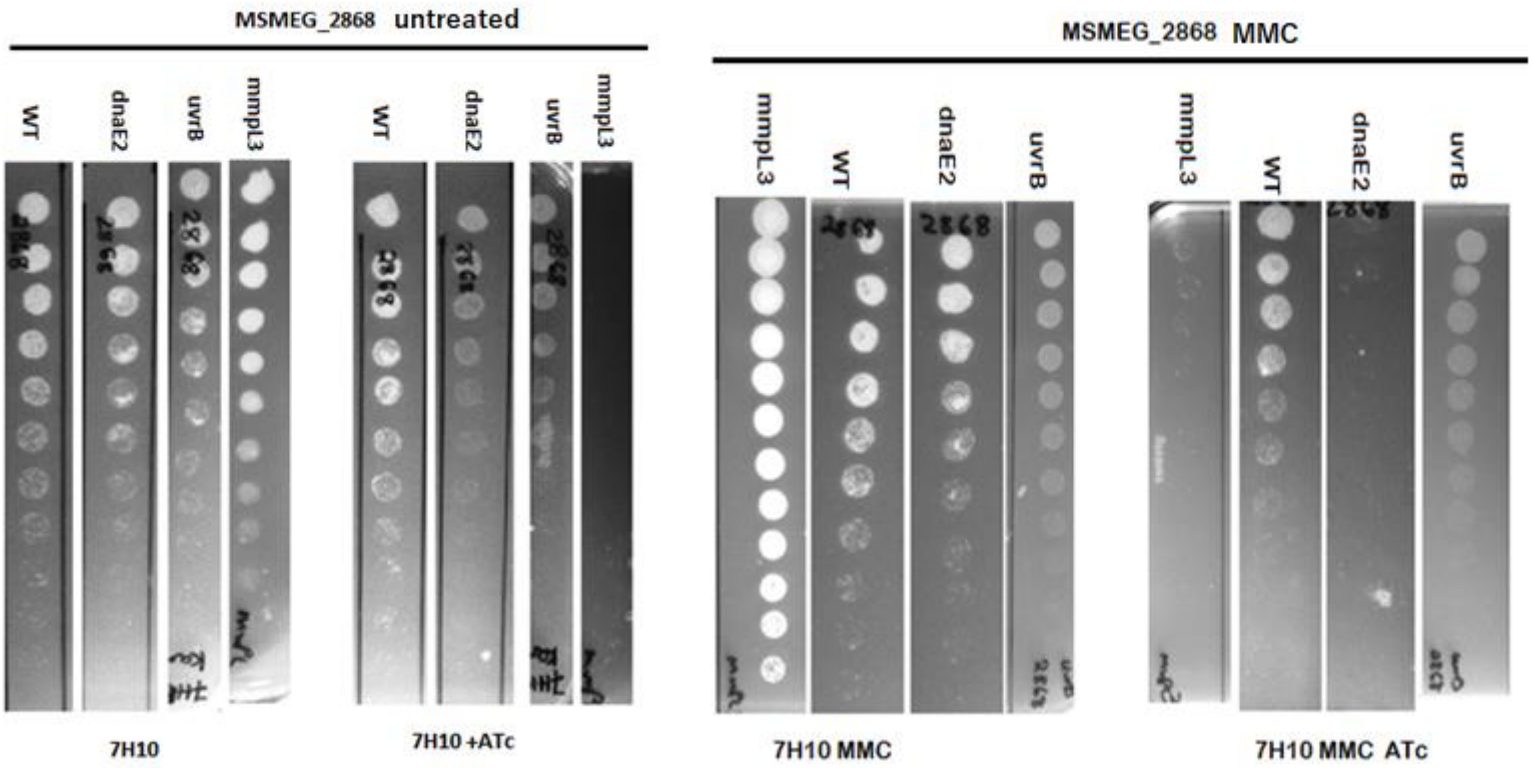


Figure 4. 13 Validation of MSMEG_2868 conditional essentiality using CRISPRi: sgRNA targeting MSMEG_2868 were co-expressed with dCas9$_{sth1}$ in WT, Δ*dnaE2* and Δ*uvrB* strains. The knockdown phenotype was monitored by plating cells on 7H10 solid media with or without MMC. Induction of the CRISPRi KD was done by addition of ATc to the medium. Bacillary growth was monitored by plating 5 µl of 10-fold serial dilutions of WT, Δ*dnaE2* and Δ*uvrB* strains which runs from undiluted (10^0) to 10^{-11} on 7H10 solid medium treated or untreated with MMC +/-ATc. No growth in +ATc plates indicates essentiality. mmpL3 CRISPRi KD in WT was used as a positive control. Left panel: Validation of MSMEG_2868 in untreated conditions (without MMC). Right panel: Validation of MSMEG_2868 essentiality under MMC treatment.

CRISPRi was also used to test the conditional essentiality of MSMEG_2868 during exposure to compounds with mechanisms of action (MOAs) that are different from MMC. Moxifloxacin and novobiocin (NOV), both of which are gyrase inhibitors (Fàbrega et al., 2009) but with distinct MOAs and griselimycin (GRY), which targets the *dnaN*-encoded β sliding clamp (Kling, 2015), were used. The results show that in contrast to the MMC phenotype, MSMEG_2868 was not essential under MOX, NOV and GRY treatments (**Figure 4.14).**

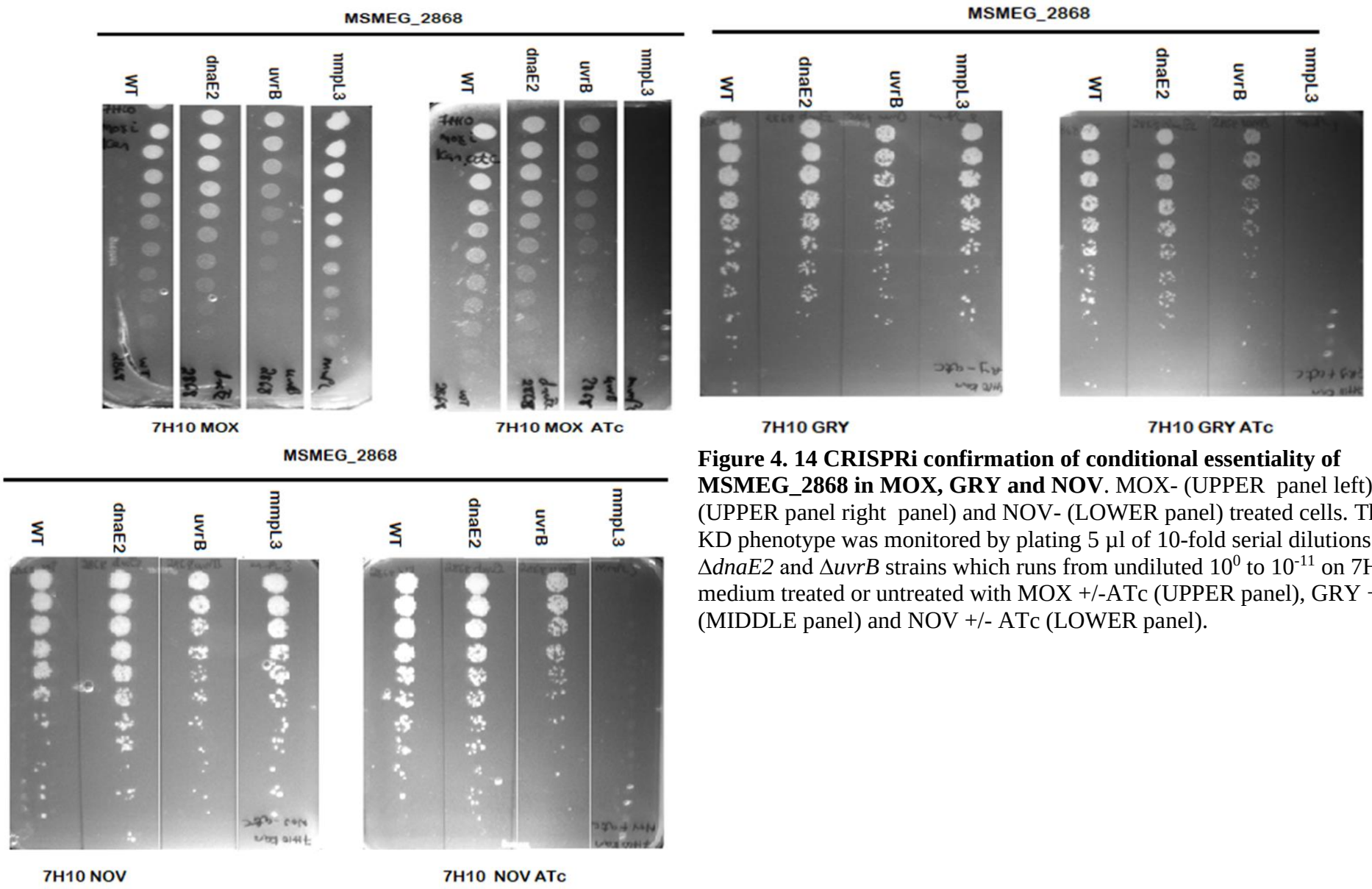


Figure 4. 14 CRISPRi confirmation of conditional essentiality of MSMEG_2868 in MOX, GRY and NOV. MOX- (UPPER panel left), GRY- (UPPER panel right panel) and NOV- (LOWER panel) treated cells. The CRISPRi KD phenotype was monitored by plating 5 µl of 10-fold serial dilutions of WT, Δ*dnaE2* and Δ*uvrB* strains which runs from undiluted 10^{0} to 10^{-11} on 7H10 solid medium treated or untreated with MOX +/-ATc (UPPER panel), GRY +/- Atc (MIDDLE panel) and NOV +/- ATc (LOWER panel).

4.3.7. Analysis of genes that are conditionally essential during MOX treatment

An investigation of potential conditionally essential genes under MOX treatment was carried out as described for MMC treatment. For MOX, 29, 64 and 889 genes were found to be conditionally essential in WT, Δ*dnaE2* and Δ*uvrB* libraries, respectively. **Tables 4.9**, **4.10**, **4.11** show the conditionally essential genes in these three libraries, respectively .

Table 4. 9 Conditionally essential genes in WT_MOX. A total of 29 genes were found to be conditionally essential after treatment of WT library with MOX. **All** conditionally essential genes **shown here had** adjusted P value ≤0.05 and delta mean of < 0 according to TRANSIT software analysis (statistical analysis results are shown in Supplementary data 3). Genes shared with all strains are highlighted in blue, genes shared only with Δ*uvrB*_MMC are highlighted in green, genes shared with Δ*dnaE2*_MMC are highlighted red and genes unique to WT_MMC are highlighted in yellow.

Orf#	Name	Description
MSMEG_0381	-	Mmp14a protein
MSMEG_0382	-	putative transport protein
MSMEG_0400	-	peptide synthetase
MSMEG_0401	-	putative non-ribosomal peptide synthase
MSMEG_0402	-	linear gramicidin synthetase subunit D
MSMEG_0866c	-	DNA or RNA helicase of superfamily protein II
MSMEG_1654	-	isocitrate dehydrogenase, NADP-dependent
MSMEG_1959	-	hypothetical protein MSMEG_1959
MSMEG_3187c	-	acyltransferase domain-containing protein
MSMEG_3888c	-	hypothetical protein MSMEG_3888
MSMEG_4269	asnB	asparagine synthase (glutamine-hydrolyzing)
MSMEG_4916	-	alpha-amylase family protein
MSMEG_5041c	-	acyltransferase
MSMEG_5149c	pimE	mannosyltransferase
MSMEG_5447	-	dolichyl-phosphate-mannose-protein mannosyltransferase
MSMEG_6939c	-	Soj family protein
MSMEG_1943	-	ATP-dependent DNA helicase
MSMEG_6398c	-	antigen 85-A
MSMEG_3839c	-	DNA polymerase I
MSMEG_5534c	pcrA	ATP-dependent DNA helicase PcrA
MSMEG_0034c	-	FHA domain-containing protein
MSMEG_4133	-	transposase
MSMEG_3816c	uvrB	excinuclease ABC subunit B
MSMEG_5672c	gltA	type II citrate synthase
MSMEG_6363c	-	cysteine desulfurase family protein
MSMEG_0031c	-	penicillin binding protein transpeptidase domain-containing protein
MSMEG_4869c	-	LAO/AO transport system ATPase
MSMEG_0403	-	integral membrane protein

MSMEG_6138	-	metallopeptidase

Orf: open reading frame; Delta mean is the variation in Tn insertions before and after treatment for each gene. A negative delta mean value signifies conditional essentiality of a gene in the treated condition.

Table 4. 10 Conditionally essential genes in *dnaE2*_MOX. A total of 64 genes were found to be conditionally essential after treatment of the Δ*dnaE2* Tn library with MOX. **All** conditionally essential genes **shown here had** adjusted P value ≤0.05 and delta mean of < 0 according to TRANSIT software analysis (statistical analysis results are shown in Supplementary table 4). Genes shared with all strains are highlighted in blue, genes shared only with WT_MOX are highlighted in green, genes shared with Δ*uvrB*_MOX are highlighted red and genes unique to Δ*dnaE2*_MOX are highlighted in yellow.

#Orf	Name	Description
MSMEG_0381	-	Mmp14a protein
MSMEG_0400	-	peptide synthetase
MSMEG_0401	-	putative non-ribosomal peptide synthase
MSMEG_0402	-	linear gramicidin synthetase subunit D
MSMEG_0382	-	putative transport protein
MSMEG_0866c	-	DNA or RNA helicase of superfamily protein II
MSMEG_1654	-	isocitrate dehydrogenase, NADP-dependent
MSMEG_1943	-	ATP-dependent DNA helicase
MSMEG_1959	-	hypothetical protein MSMEG_1959
MSMEG_3187c	-	acyltransferase domain-containing protein
MSMEG_3839c	-	DNA polymerase I
MSMEG_4269	asnB	asparagine synthase (glutamine-hydrolyzing)
MSMEG_5041c	-	acyltransferase
MSMEG_4916	-	alpha-amylase family protein
MSMEG_5149c	pimE	mannosyltransferase
MSMEG_5447	-	dolichyl-phosphate-mannose-protein mannosyltransferase
MSMEG_6398c	-	antigen 85-A
MSMEG_6939c	-	Soj family protein
MSMEG_5534c	pcrA	ATP-dependent DNA helicase PcrA
MSMEG_4702c	-	ABC-type transporter, permease components
MSMEG_0241c	-	MmpL11 protein
MSMEG_0860c	-	CDP-diacylglycerol--serine O-phosphatidyltransferase
MSMEG_0973	-	hypothetical protein MSMEG_0973
MSMEG_2775	nhaA	Na+/H+ antiporter NhaA
MSMEG_6143	-	hypothetical protein MSMEG_6143
MSMEG_3147	moxR	ATPase, MoxR family protein
MSMEG_1964	-	mitomycin radical oxidase
MSMEG_3222	-	prolipoprotein diacylglyceryl transferase
MSMEG_6284	-	cyclopropane-fatty-acyl-phospholipid synthase
MSMEG_3888c	-	hypothetical protein MSMEG_3888

MSMEG_4417c	msrA	methionine sulfoxide reductase A
MSMEG_2418	rnc	ribonuclease III
MSMEG_4505c	hrcA	heat-inducible transcription repressor
MSMEG_4852c	-	enoyl-CoA hydratase, putative
MSMEG_5600	-	hypothetical protein MSMEG_5600
MSMEG_0380	-	MmpS4 protein
MSMEG_4491c	recO	DNA repair protein RecO
MSMEG_3886c	tatC	twin arginine-targeting protein translocase TatC
MSMEG_0840c	-	hypothetical protein MSMEG_0840
MSMEG_0386	-	NAD dependent epimerase/dehydratase family protein
MSMEG_4189c	cysS	cysteinyl-tRNA synthetase
MSMEG_5002c	-	hypothetical protein MSMEG_5002
MSMEG_5049	kgd	alpha-ketoglutarate decarboxylase
MSMEG_5392	kdpA	potassium-transporting ATPase subunit A
MSMEG_5396	-	KDP operon transcriptional regulatory protein KdpE
MSMEG_5470c	-	molybdopterin biosynthesis protein MoeA 1
MSMEG_6012c	-	putative acyl-CoA dehydrogenase
MSMEG_6283	-	FAD binding domain-containing protein
MSMEG_6387	-	arabinosyltransferase A
MSMEG_6555c	-	transcriptional regulator, TetR family protein, putative
MSMEG_1929c	-	hypothetical protein MSMEG_1929
MSMEG_3227	pyk	pyruvate kinase
MSMEG_5004c	-	DNA repair exonuclease
MSMEG_5263	greA	transcription elongation factor GreA
MSMEG_5693	-	transporter, major facilitator family protein
MSMEG_2934	-	lipid A biosynthesis lauroyl acyltransferase
MSMEG_4265c	-	MmpS3 protein
MSMEG_1941	-	helicase, UvrD/Rep family protein
MSMEG_3247	-	branched-chain amino acid ABC transporter substrate-binding protein
MSMEG_4241	-	integral membrane protein
MSMEG_5672c	gltA	type II citrate synthase

Orf: open reading frame; Delta mean is the variation in Tn insertions before and after treatment for each gene. A negative delta mean value signifies conditional essentiality of a gene in the treated condition.

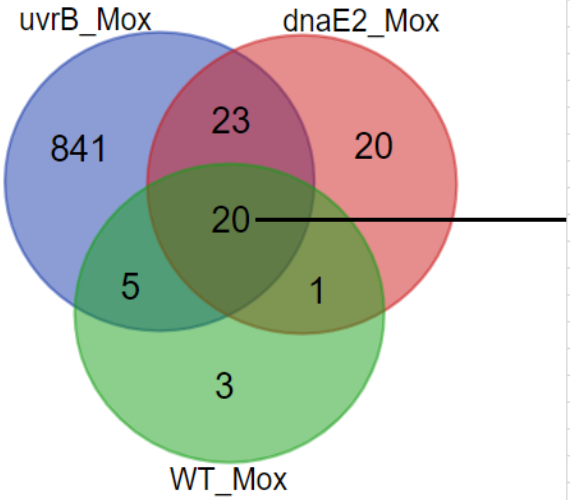
Table 4. 11 Conditionally essential genes in *uvrB*-MOX. A total of 889 genes were found to be conditionally essential after treatment of the Δ*uvrB* library with MOX. Since this is a huge list, only those with gene names as they appear on https://mycobrowser.epfl.ch/ are shown. **All** conditionally essential genes **shown here had** adjusted P value ≤0.05 and delta mean of < 0 according to TRANSIT software analysis (a full list of conditional essentials and statistical analysis results are shown in Supplementary data 3). Genes shared with all strains are highlighted in blue, genes shared only with Δ*dnaE2*_MOXare highlighted in green, genes unique to Δ*uvrB*_MOX are highlighted yellow.

#Orf	Name	Description
MSMEG_4269	*asnB*	asparagine synthase (glutamine-hydrolyzing)
MSMEG_5149c	*pimE*	mannosyltransferase
MSMEG_5534c	*pcrA*	ATP-dependent DNA helicase PcrA
MSMEG_3227	*pyk*	pyruvate kinase
MSMEG_4189c	*cysS*	cysteinyl-tRNA synthetase
MSMEG_5392	*kdpA*	potassium-transporting ATPase subunit A
MSMEG_5263	*greA*	transcription elongation factor GreA
MSMEG_4979c	*cysD*	sulfate adenylyltransferase subunit 2
MSMEG_2594c	*asnB*	asparagine synthase (glutamine-hydrolyzing)
MSMEG_6612c	*moxR*	ATPase, MoxR family protein
MSMEG_0418c	*sdhA*	succinate dehydrogenase flavoprotein subunit
MSMEG_0647c	*phnC*	phosphonate ABC transporter, ATP-binding protein
MSMEG_0974	*ccsB*	cytochrome c-type biogenesis protein CcsB
MSMEG_1713	*araB*	ribulokinase
MSMEG_2051c	*nuoM*	NADH dehydrogenase subunit M
MSMEG_3403	*amiF*	formamidase
MSMEG_3724c	*pqqB*	pyrroloquinoline quinone biosynthesis protein PqqB
MSMEG_4910c	*clpS*	ATP-dependent Clp protease adaptor protein ClpS
MSMEG_5056	*corA*	magnesium and cobalt transport protein CorA
MSMEG_5912	*gabD2*	succinic semialdehyde dehydrogenase
MSMEG_2294c	*dinB*	DNA polymerase IV
MSMEG_4114c	*menB*	naphthoate synthase
MSMEG_4235c	*mraW*	*S*-adenosyl-methyltransferase MraW
MSMEG_6301c	*ligD*	DNA polymerase LigD polymerase subunit
MSMEG_6940c	*gidB*	methyltransferase GidB
MSMEG_1852	*selD*	selenide, water dikinase
MSMEG_2123c	*dhaK*	dihydroxyacetone kinase, DhaK subunit
MSMEG_2403	*recG*	ATP-dependent DNA helicase RecG
MSMEG_4590c	*ssuD*	nitrilotriacetate monooxygenase component A
MSMEG_4415c	*tal*	transaldolase
MSMEG_3743c	*soj*	SpoOJ regulator protein
MSMEG_4552c	*ssuD*	nitrilotriacetate monooxygenase component A
MSMEG_1014c	*dinB*	DNA polymerase IV
MSMEG_1849c	*selB*	selenocysteine-specific translation elongation factor
MSMEG_2057c	*nuoG*	NADH dehydrogenase subunit G
MSMEG_5548c	*cobF*	precorrin 6A synthase

MSMEG_3771c	*argR*	arginine repressor
MSMEG_5782c	*pstS*	phosphate ABC transporter, phosphate-binding protein PstS
MSMEG_3648c	*gcvH*	glycine cleavage system protein H
MSMEG_1088	*gatA*	glutamyl-tRNA (Gln)/aspartyl-tRNA (Asn) amidotransferase, A subunit
MSMEG_2784c	*msrB*	methionine-R-sulfoxide reductase
MSMEG_6634c	*mdcB*	triphosphoribosyl-dephospho-CoA synthase MdcB
MSMEG_0994c	*resD*	DNA-binding response regulator ResD
MSMEG_1527	*truA*	tRNA pseudouridine synthase A
MSMEG_6809	*cbbQ*	CbbQ protein
MSMEG_5986c	*glmS*	glucosamine--fructose-6-phosphate aminotransferase, isomerizing
MSMEG_1946	*nudC*	NADH pyrophosphatase
MSMEG_2427	*glnD*	PII uridylyl-transferase
MSMEG_2412	*pyc*	pyruvate carboxylase
MSMEG_0733c	*dapB*	dihydrodipicolinate reductase N-terminus domain-containing protein
MSMEG_6053	*bluB*	cob (II)yrinic acid a,c-diamide reductase
MSMEG_2273	*hypF*	[NiFe] hydrogenase maturation protein HypF [Mycobacterium smegmatis str. MC2 155]
MSMEG_3548	*ehuC*	ectoine/hydroxyectoine ABC transporter, permease protein EhuC
MSMEG_4669	*fdhD*	formate dehydrogenase accessory protein
MSMEG_2724	*recX*	recombination regulator RecX
MSMEG_3878c	*cobL*	precorrin-6Y C5,15-methyltransferase (decarboxylating)
MSMEG_4091c	*ssuD*	nitrilotriacetate monooxygenase component A
MSMEG_4418c	*msrB*	methionine-R-sulfoxide reductase
MSMEG_6458c	*gltD*	glutamate synthase subunit beta
MSMEG_2649	*truB*	tRNA pseudouridine synthase B
MSMEG_5779c	*pstB*	phosphate ABC transporter, ATP-binding protein
MSMEG_1476	*sppA*	signal peptide peptidase SppA, 67K type
MSMEG_1328c	*recC*	exodeoxyribonuclease V, gamma subunit
MSMEG_5408	*lgt*	prolipoprotein diacylglyceryl transferase
MSMEG_0709	*dnaK*	molecular chaperone DnaK
MSMEG_1182	*hutI*	imidazolonepropionase
MSMEG_3184c	*treZ*	malto-oligosyltrehalose trehalohydrolase
MSMEG_5049	*kgd*	alpha-ketoglutarate decarboxylase
MSMEG_4374c	*speB*	agmatinase

Orf: Open reading frame; Delta mean is the variation in Tn insertions before and after treatment for each gene. A negative delta mean value signifies conditional essentiality of a gene in the treated condition.

The Δ*uvrB*_MOX library was found to contain a very large number (889) of conditionally essential genes during MOX treatment compared to other libraries (**Figure 4.15**). This observation was surprising because all the parental strains (WT, Δ*dnaE2* and Δ*uvrB*) had the same MIC for MOX (**Figure 4.3**). A Venn diagram comparing conditionally essential genes in these libraries show that 20 were found to be common to all three libraries. These include among others, *pcrA* and *pimE* (**Figure 4.15**) which were also common to all the strains under MMC treatment (**Figure 4.4**).



ORF	Name	Description
MSMEG_0866c		DNA or RNA helicase of superfamily protein II
MSMEG_3888c		hypothetical protein MSMEG_3888
MSMEG_3187c		acyltransferase domain-containing protein
MSMEG_0402		linear gramicidin synthetase subunit D
MSMEG_4269		asparagine synthase (glutamine-hydrolyzing)
MSMEG_0381		Mmp14a protein
MSMEG_0400		peptide synthetase
MSMEG_5149c	PimE	mannosyltransferase
MSMEG_0382		putative transport protein
MSMEG_1959		hypothetical protein MSMEG_1959
MSMEG_5447		dolichyl-phosphate-mannose-protein mannosyltransferase
MSMEG_3839c		DNA polymerase I
MSMEG_1943		ATP-dependent DNA helicase
MSMEG_6398c		antigen 85-A
MSMEG_5041c		acyltransferase
MSMEG_5534c	pcrA	ATP-dependent DNA helicase PcrA
MSMEG_6939c		Soj family protein
MSMEG_0401		putative non-ribosomal peptide synthase
MSMEG_4916		alpha-amylase family protein
MSMEG_1654		isocitrate dehydrogenase, NADP-dependent

Figure 4. 15 Venn diagram comparing gene sets predicted to be conditionally essential in WT, Δ*dnaE2* and Δ*uvrB* libraries under MOX treatment.

The results show that TLS by DnaE2 is not required in WT during MOX treatment. Also, while all the NER enzymes were required for MMC treatment, only UvrB is required in WT-MOX. In order to understand this result, we looked at the abundance of Tn insertions in *uvrB* and *dnaE2* in the two WT libraries. As observed in **Figure 4.16** there was no difference in Tn insertion of *dnaE2* between WT_untreated and WT_MOX (**Figure 4.16 BOTTOM**). However differences in Tn insertions were observed in *uvrB* which indicate that disruption of *uvrB* in WT causes a fitness defect when grown on MOX (**Figure 4.16 UPPER**). This was supported by a significant $P_{adjusted}$ value of 0 and a delta mean of -258 (Supplementary data 3). *uvrB* did not come up as conditionally essential in a Δ*dnaE2*_MOX library, and neither did *dnaE2* come up as conditionally essential in Δ*uvrB*_MOX library (**Figure 4.17** and **4.18** respectively).

However there seems to be domain essentiality of *dnaE2* in a Δ*uvrB*_MOX library because only the -TA- that are in the middle of this gene seems to be able to sustain insertions while the -TA-at the beginning and end of this gene were devoid of Tn insertions (**Figure 4.18**)

.

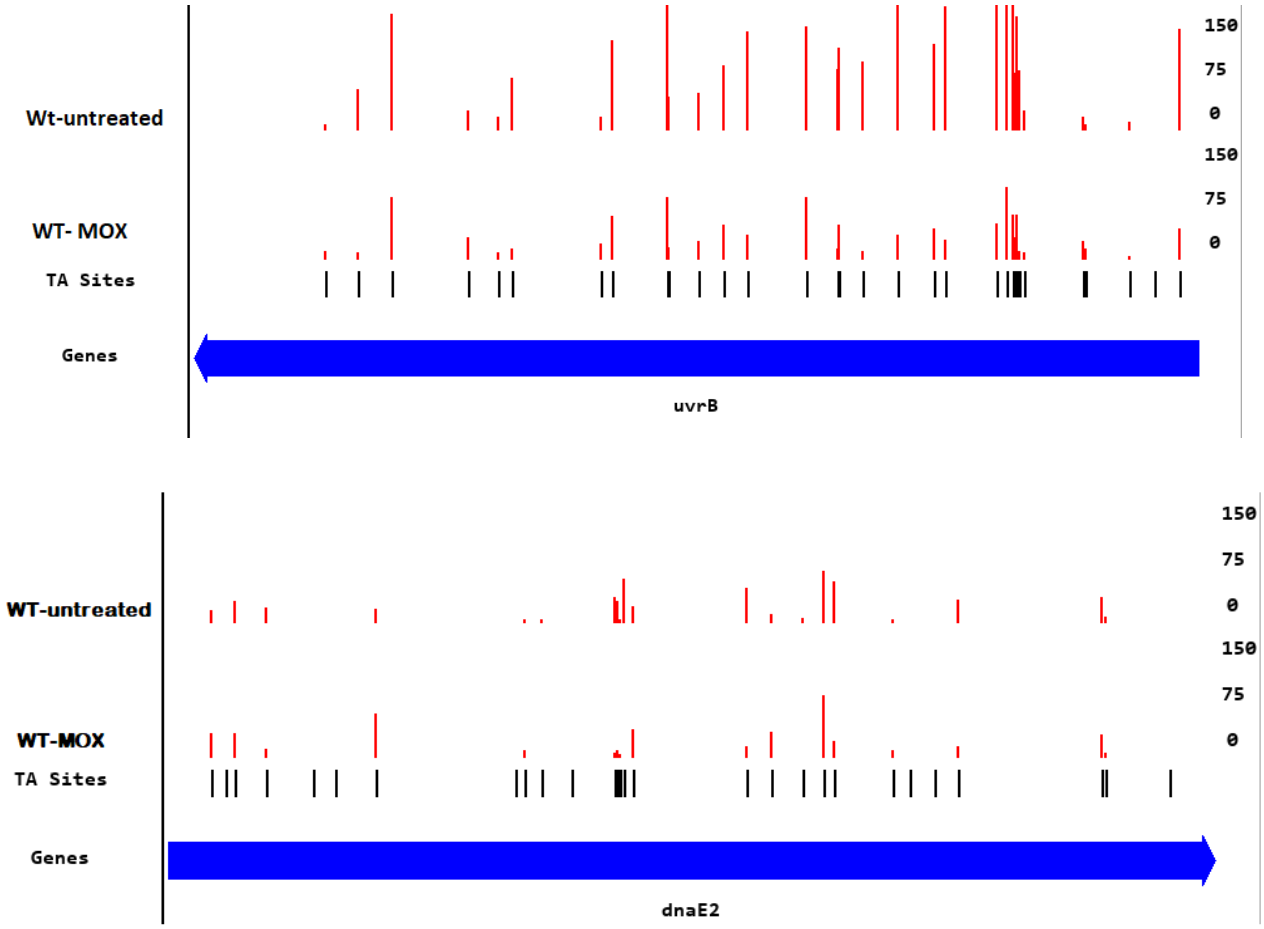


Figure 4. 16 Distribution of Tns insertions in *uvrB* (upper panel) and *dnaE2* (bottom panel) before after treatment of WT Tn library with MOX. Thick blue arrows show the genes and the red lines show Tn insertion frequencies for the different -TA- sites (short black lines) of each gene. Comparable magnitudes and patterns of insertions between the two conditions indicate non-essentiality. Images were generated using TRANSIT software.

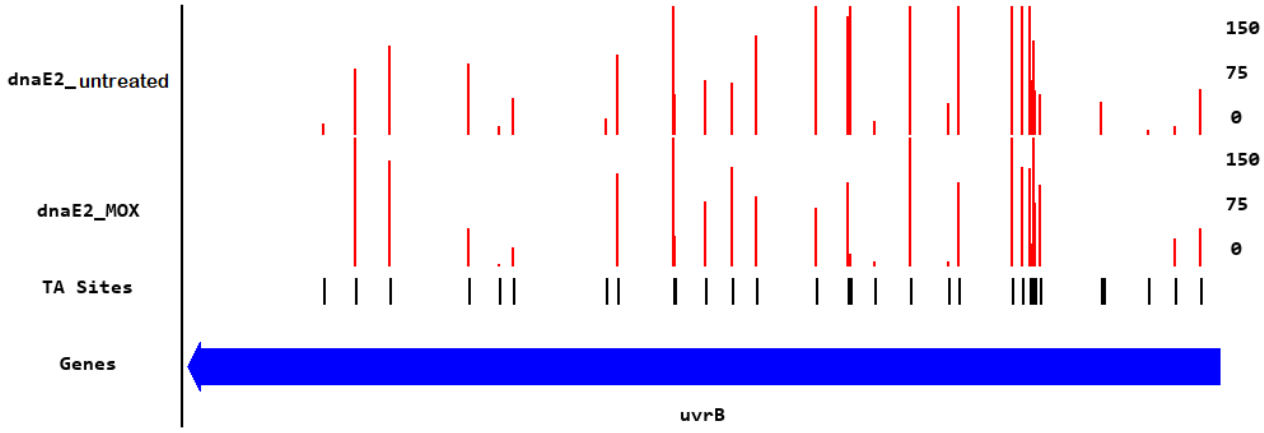


Figure 4. 17 Distribution of transposon insertions in *uvrB* before and after treatment of the Δ*dnaE2* Tn library with MOX. Thick blue arrows show the gene and the red lines indicate Tn insertion frequencies in each of the different -TA- sites Comparable magnitude and pattern of insertions between the two conditions indicates non-essentiality. Images were generated using TRANSIT software.

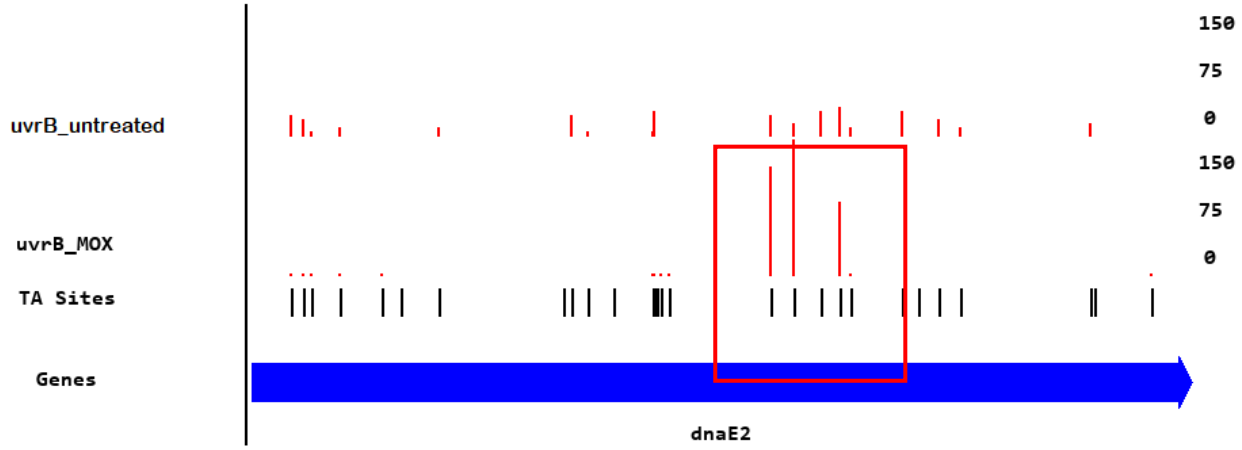


Figure 4. 18 Distribution of transposon insertions in *dnaE2* before and after treatment of Δ*uvrB* Tn library with MOX. Thick blue arrows show the genes and the red lines indicate Tn insertion frequencies in the different -TA- sites (short black lines). The section highlighted with a red rectangle shows the only region of *dnaE2* that can sustain Tn insertions in a Δ*uvrB*_MOX library. Images were generated using TRANSIT.

Functional classification of the conditionally essential genes shows that, similarly to MMC treatment genes in the functional classes of Intermediary metabolism and respiration and Cell wall and cell processes are abundant under MOX treatment (**Figure 4.19**).

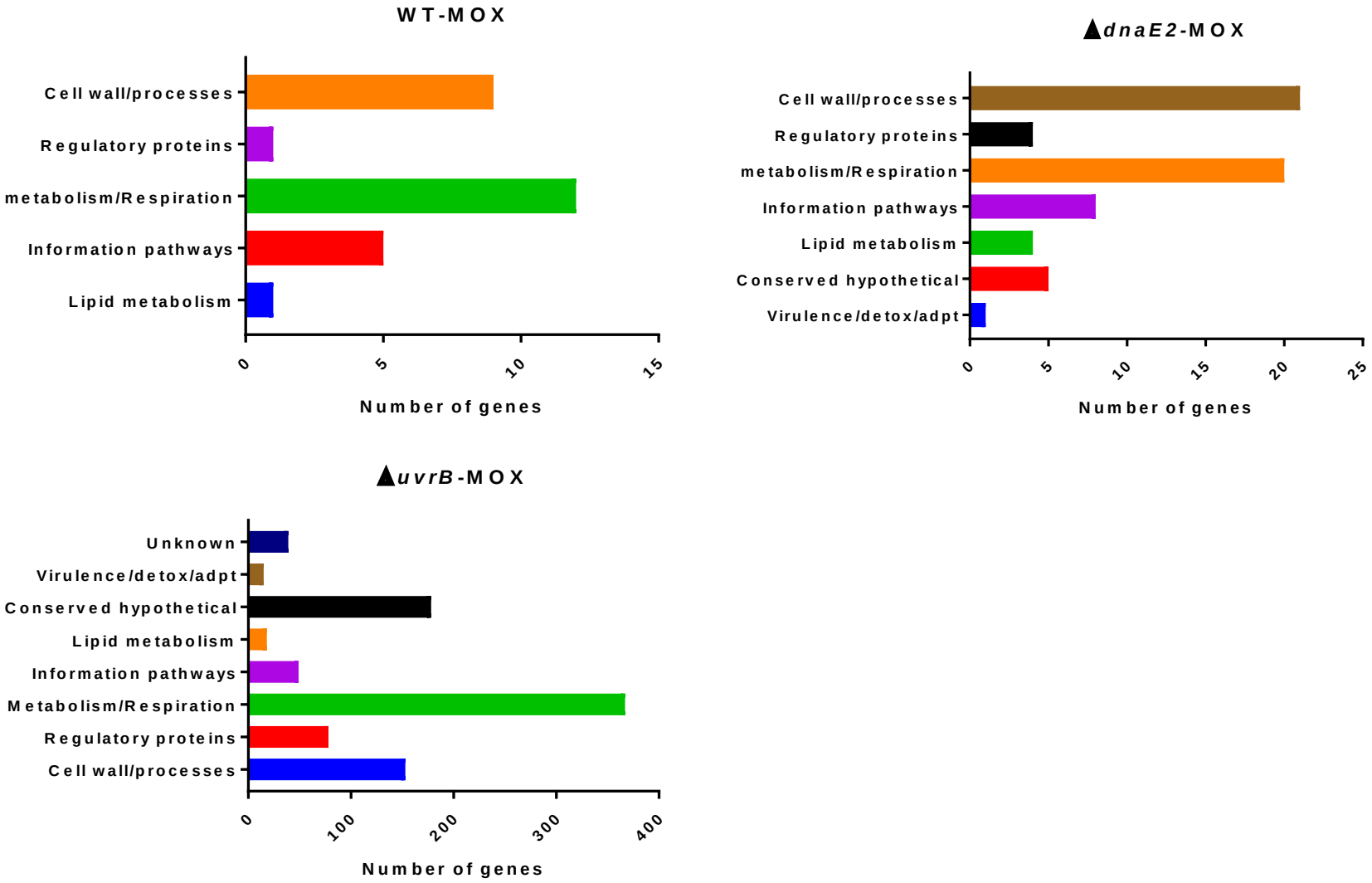


Figure 4. 19 Functional classification of conditionally essential genes identified in WT, Δ*uvrB* and Δ*dnaE2* libraries under MOX treatment. MTB homologs of all conditionally genes were first identified using Mycobrowser (https://mycobrowser.epfl.ch/). For those that did not have MTB homologs/orthologs, functional classification was inferred from UniProt database: https://www.uniprot.org/.

Next, CRISPRi was used to validate the MOX-dependent essentiality of *pimE* in the libraries. The TnSeq screen identified this gene as essential in all MOX-treated libraries (**Figure 4.20**). CRISPRi confirmed the essentiality of *pimE* in Δ*dnaE2* and Δ*uvrB* treated with MOX. However, in the WT strain, the phenotype was subtle: growth of the CRISPRi mutant was observed upon ATc induction, but the growth was not as dense at that observed in the non-induced (-ATc)control (**Figure 4.20**).

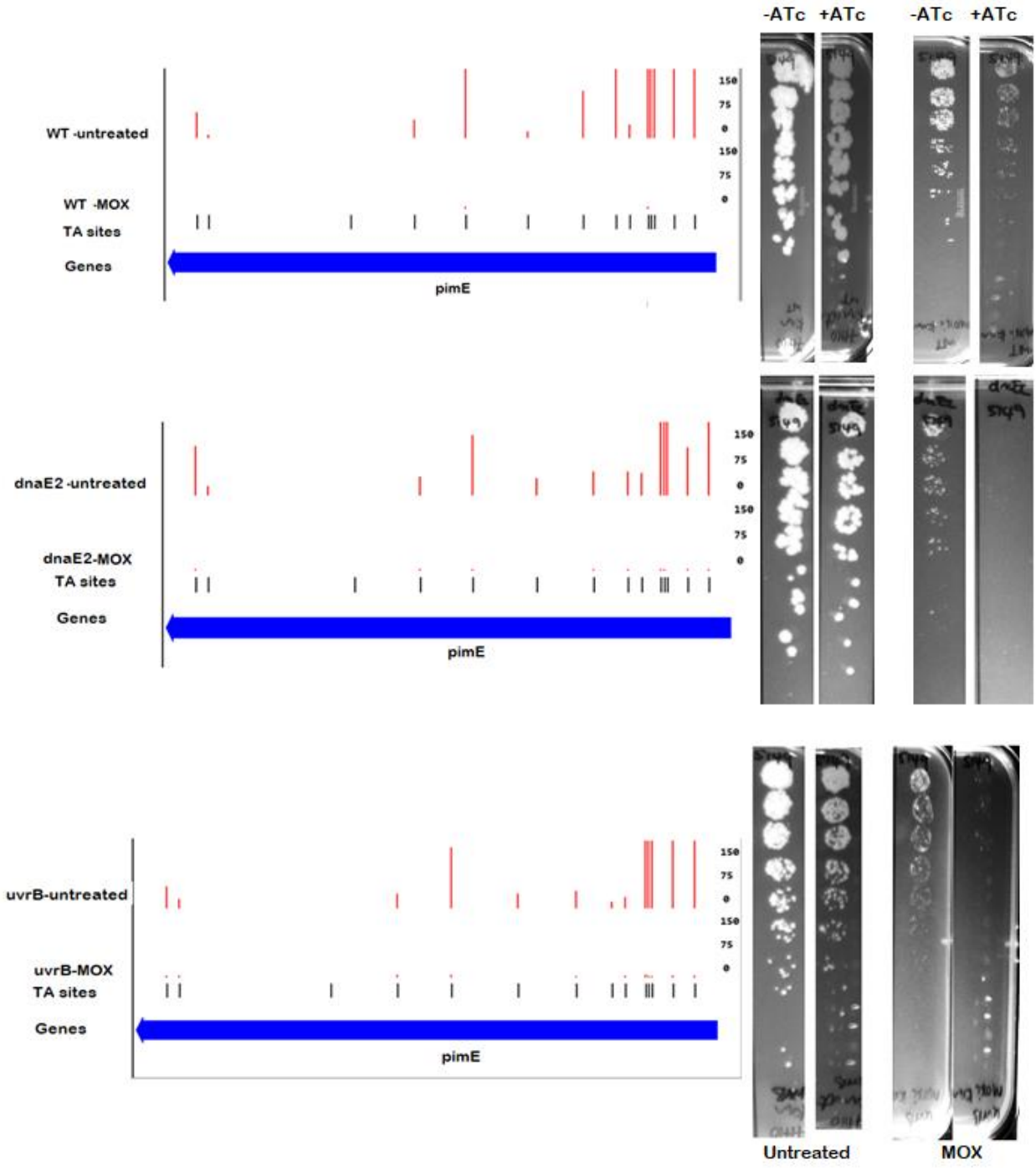


Figure 4. 20 Validation of the **essentiality of *pimE* during MOX treatment by CRISPRi.** The distributions of Tn insertions in *pimE* for each library are shown with the corresponding CRISPRi validation plates. For Tn insertion images (LEFT), the thick blue arrows show the gene of interest and the red lines indicate Tn insertion frequencies in the different -TA- dinucleotides. For CRISPRi validation (RIGHT), sgRNA targeting *pimE* were co-expressed with $dCas9_{sth1}$ in WT, Δ*dnaE2* and Δ*uvrB* parental strains. The KD phenotype was monitored by plating 5 µl of 10-fold serial dilutions of WT, Δ*dnaE2* and Δ*uvrB* strains which runs from undiluted 10^0 to 10^{-7} on 7H10 solid medium treated or untreated with MOX +/-ATc, where no growth on +ATc plate indicates essentiality. The UPPER panel is WT, MIDDLE panel is Δ*dnaE2* and LOWER panel is Δ*uvrB*. Untreated panels are 7H10 plates without MOX which were included as controls and show that the gene is non-essential under optimal conditions.

4.3.8. Genetic interactions under MOX treatment

Aggravating genetic interactions during MOX treatment were analysed as outlined above for MMC treatment. They were identified as those interactions whose adjusted probability of the delta-log fold change (delta-logFC) falling within region of practical equivalence (ROPE) was ≤ 0.05. This analysis yielded 16 genes that that had aggravating genetic interactions with Δ*dnaE2* (**Table 4.12**). Most of these were involved in metabolic processes including lipid degradation (MSMEG_5494, MSMEG_5297 and MSMEG_6012). A TetR family protein transcriptional regulator MSMEG_6807 was also identified, as were multiple proteins of unknown function including MSMEG_6338 and MSMEG_3820 (**Table 4.12**). While very few Tn insertions were observed in MSMEG_6235 and MSMEG_6012 in a Δ*dnaE2*_MOX library compared to other libraries (**Figure 4.21**), most genes that were identified as giving aggravating genetic interactions with Δ*dnaE2* in MOX treatment had observable comparable mean Tn insertions with those in an untreated WT library, *e.g*., *rhaB*, encoding rhamnulokinase, had 59.43 and 42.93 mean Tn insertions in WT-untreated and *dnaE2*-MOX respectively (Supplementary data 4).

Table 4. 12 Genes that have aggravating genetic interactions with *dnaE2* during MOX treatment.

#ORF	Name	Description	Biological process
MSMEG_6012c	-	putative acyl-CoA dehydrogenase	Lipid degradation
MSMEG_6235	-	thiopurine *S*-methyltransferase (tpmt) superfamily protein	Conserved hypothetical
MSMEG_3131c	-	AMP-binding protein	Lipid degradation
MSMEG_6338c	-	phosphoglycerate mutase family protein, putative	unknown
MSMEG_5494	-	acyl-CoA dehydrogenase fadE12	Lipid degradation; oxidoreductase
MSMEG_0591	*rhaB*	rhamnulokinase	Rhamnose catabolic process
MSMEG_6807c	-	transcriptional regulator, TetR family protein	Regulation of transcription
MSMEG_3820	-	hypothetical protein MSMEG_3820	unknown
MSMEG_6577	-	methylmalonyl-CoA carboxyltransferase 12S subunit	unknown
MSMEG_0186c	-	MmpS2 protein	Membrane protein
MSMEG_0224	-	O-methyltransferase MdmC	Methylation
MSMEG_1208	-	glycosyltransferase, group I	transferase
MSMEG_5297	-	putative oxalyl-CoA decarboxylase	Fatty acid alpha oxidation
MSMEG_6017	-	aspartate aminotransferase	Biosynthetic process
MSMEG_3505	-	6-aminohexanoate-cyclic-dimer hydrolase	hydrolase
MSMEG_5609	-	carotenoid oxygenase	Carotene catabolism

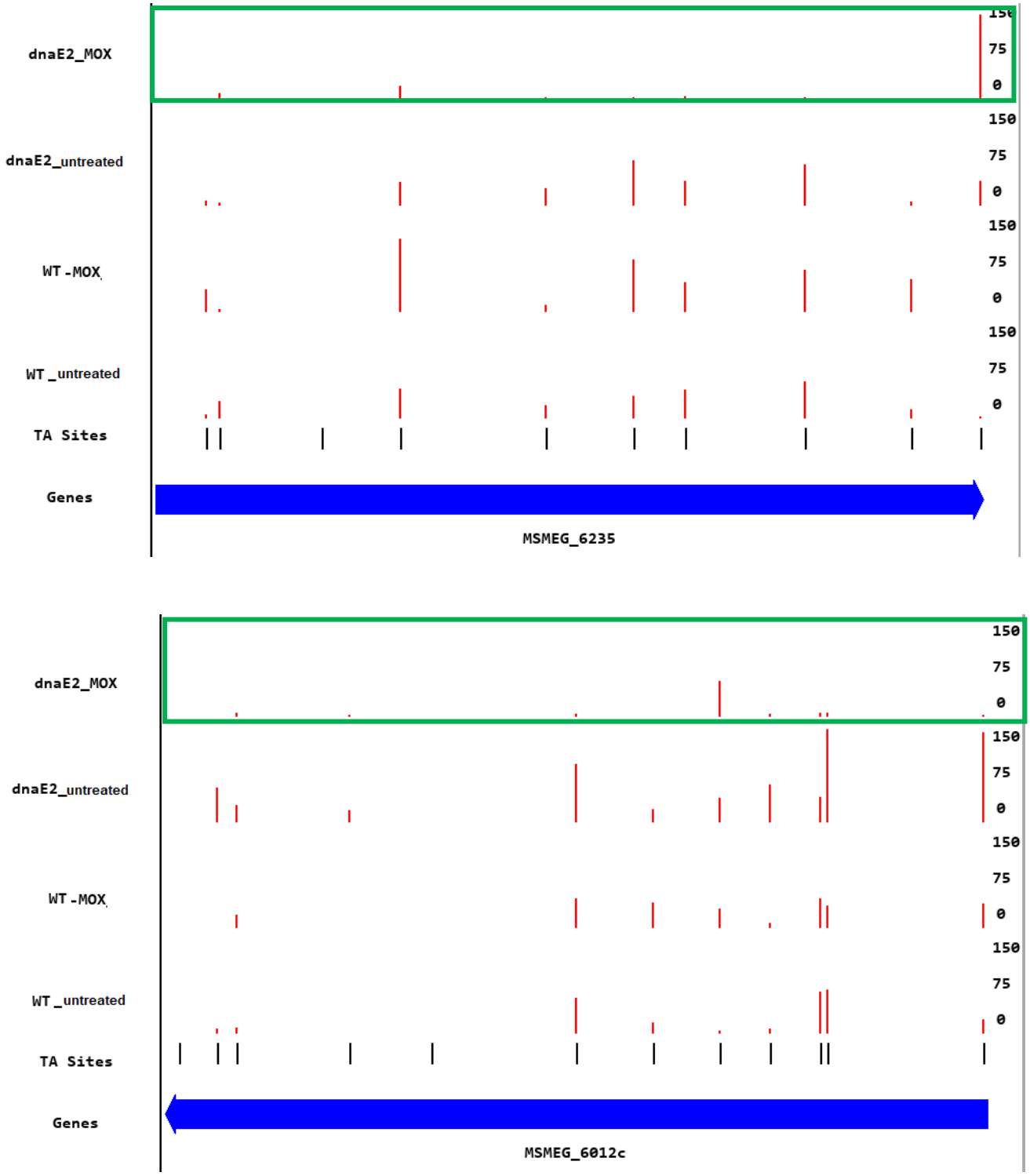


Figure 4. 21 Distribution of transposon insertions in selected genes with aggravating interactions with *dnaE2* during MOX treatment. Insertions in the four libraries (WT-untreated, WT-MOX, *dnaE2*-untreated and *dnaE2*-MOX) that were analysed together are shown for every gene shown. Thick blue arrows show the genes and the red lines indicate Tn insertion frequencies in the different -TA- sites (short black lines). Green outlined rectangles show Tn insertions in the gene of interest in a *dnaE2*-MMC library. Numbers on the right represent the magnitude of the Tn insertions. Images were generated using TRANSIT software.

We identified 163 genes that had aggravating genetic interactions with Δ*uvrB* during MOX treatment. Some of these genes (those with gene names as they appear on https://mycobrowser.epfl.ch/) are shown in **Table 4.12**. Most of the encoded proteins are predicted to function in the intermediary metabolism and respiration, and cell wall and cell processes functional classes. However, many hypothetical/conserved hypothetical genes were also identified (**Figure 4.22**). Most of the genes that belong to the intermediary metabolism and respiration class function in oxidation-reduction processes (Supplementary data 4). For genes that belong to the regulatory functional group, almost all were transcriptional regulators (Supplementary data 4). **Figure 4.23** shows Tn insertions across 4 libraries in *pqqB*, a pyrroloquinoline quinone biosynthesis protein; *amiF*, a formamidase; and *recC*, an exodeoxyribonuclease V gamma subunit. Aggravating interactions of these genes with Δ*uvrB* in MOX treatment are signified by the almost complete absence of Tn insertions in these genes in a Δ*uvrB*-MOX library compared to other libraries which had an abundance of Tn insertions in these genes (Figure 4.23).

Table 4. 13 Genes that have aggravating genetic interactions with *uvrB* during MOX treatment. **163 genes were identified. Only those with gene names are shown.**

ORF	Name	Description	Biological process
MSMEG_6301c	*ligD*	DNA polymerase LigD polymerase subunit	DNA replication
MSMEG_3403	*amiF*	formamidase	Nitrogen metabolic process
MSMEG_3724c	*pqqB*	pyrroloquinoline quinone biosynthesis protein PqqB	PQQ biosynthesis
MSMEG_5912	*gabD2*	succinic semialdehyde dehydrogenase	TCA, oxidoreductase
MSMEG_2123c	*dhaK*	dihydroxyacetone kinase, DhaK subunit	Glycerol metabolism
MSMEG_0418c	*sdhA*	succinate dehydrogenase flavoprotein subunit	Anaerobic respiration; oxidoreductase
MSMEG_5056	*corA*	magnesium and cobalt transport protein CorA	Ion transport
MSMEG_1713	*araB*	ribulokinase	Arabinose catabolism
MSMEG_4910c	*clpS*	ATP-dependent Clp protease adaptor protein ClpS	proteolysis

MSMEG_1328c	*recC*	exodeoxyribonuclease V, gamma subunit	DNA repair

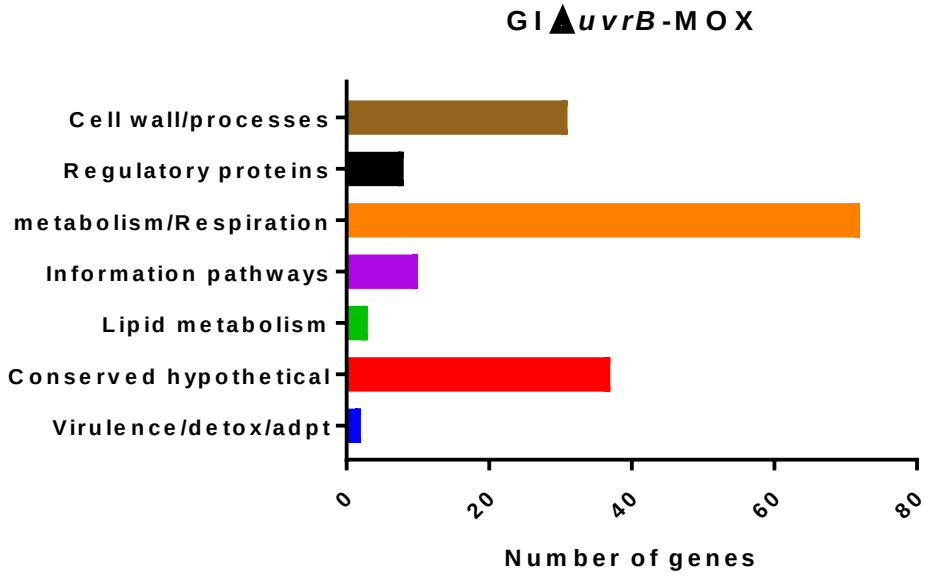
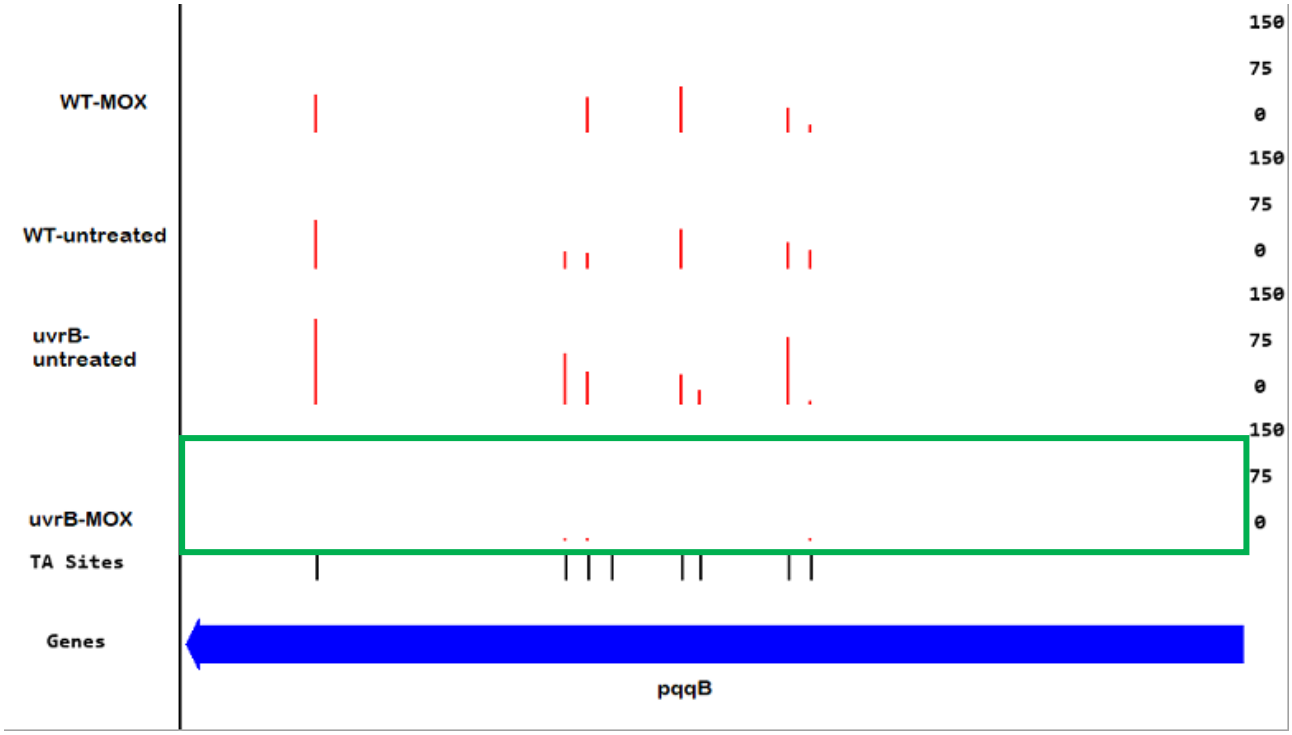
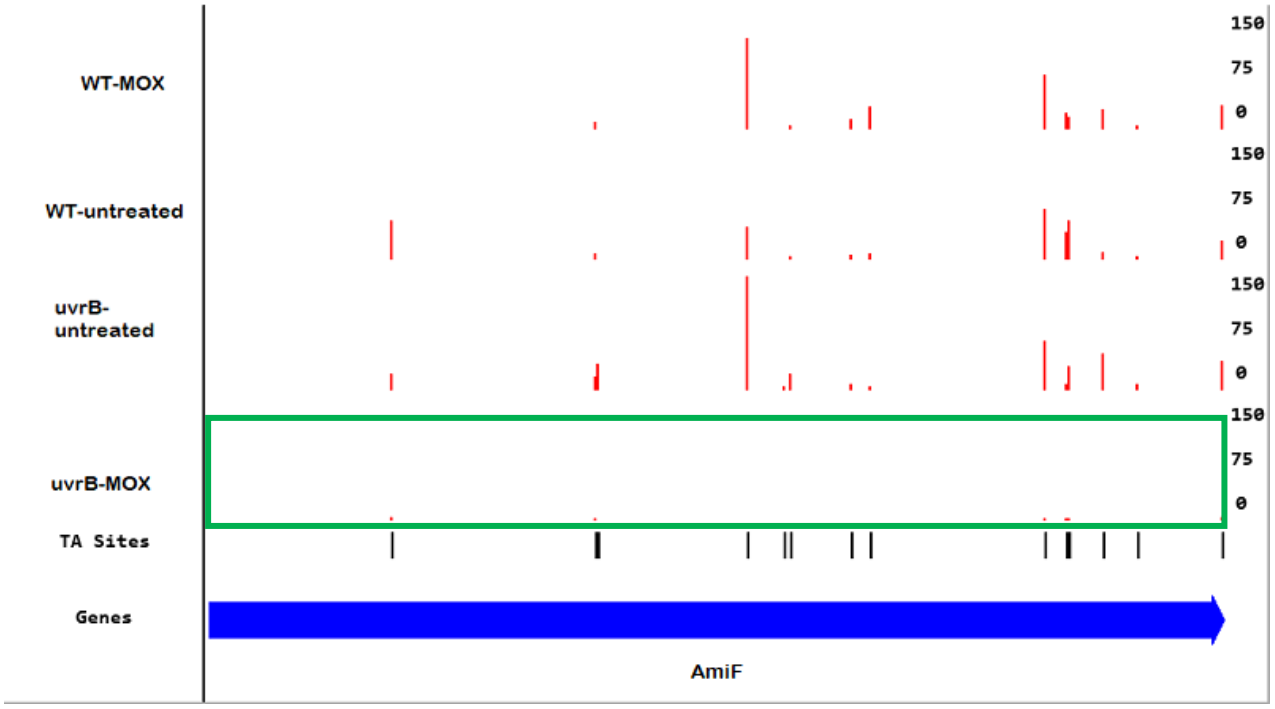


Figure 4. 22 Functional classification 163 genes that were identified as having aggravating genetic interactions with *uvrB* during MOX treatment. MTB homologs of all the conditionally genes were first identified using Mycobrowser (https://mycobrowser.epfl.ch/). For those that do not have MTB homologs/orthologs functional classification was inferred from UniProt database: https://www.uniprot.org/.


WT-MOX
WT-untreated
uvrB-
untreated
uvrB-MOX
TA Sites
Genes
pqqB
150
75
0
150
75
0
150
75
0
150
75
0


WT-MOX
WT-untreated
uvrB-
untreated
uvrB-MOX
TA Sites
Genes
AmiF
150
75
0
150
75
0
150
75
0
150
75
0

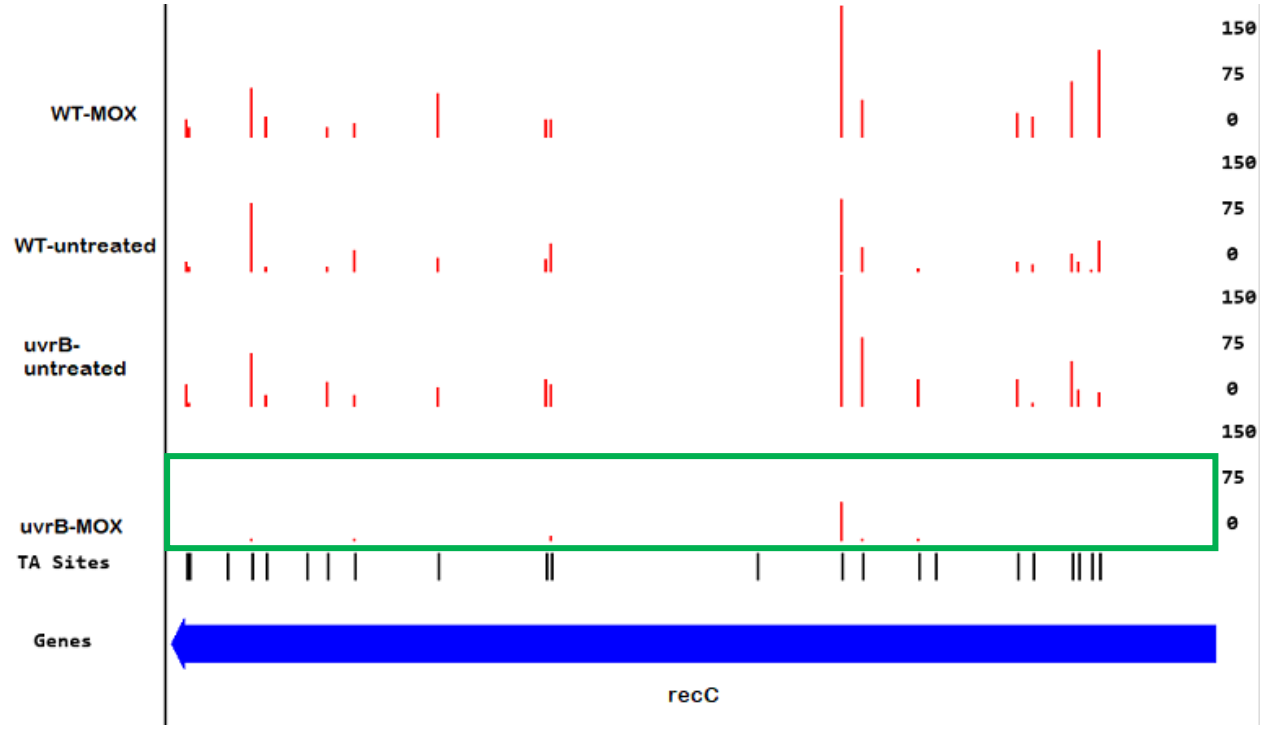


Figure 4. 23 Distribution of Tn insertions in selected genes with aggravating genetic interactions with *uvrB* during MOX treatment. Insertions in the four libraries (WT-untreated, WT-MOX, *uvrB*-untreated and *uvrB*-MOX) that were analysed together are shown for every gene shown. Thick blue arrows show the genes and the red lines indicate Tn insertion frequencies in the different -TA- sites (short black lines). Green outlined rectangles show Tn insertions in the gene of interest in a *uvrB*-MOX library. Numbers on the right represent the magnitude of the Tn insertions. Images were generated using TRANSIT software.

4.3.9. Comparison of aggravating GI genes in Δ*uvrB* and Δ*dnaE2* for both MMC and MOX treatments

When a list of genes that exhibited aggravating genetic interactions with *dnaE2* and *uvrB* under MMC treatment were compared, none was found to be common to both libraries (**Figure 4.24**). Performing the same comparison for MOX treatment revealed 6 genes with aggravating interactions with both *uvrB* and *dnaE2* (**Figure 4.25**). These include MSMEG_3820, a hypothetical protein; MSMEG_6012, a putative acyl-CoA dehydrogenase; MSMEG_6338, a phosphoglycerate mutase family protein; and MSMEG_6235, a thiopurine *S*-transferase superfamily protein (**Figure 4.25**). Visual inspection of the distribution of Tn insertions in these genes showed that MSMEG_6235 had few to no insertions in both Δ*uvrB*-MOX and Δ*dnaE2*-MOX libraries (**Figure 4.26**).

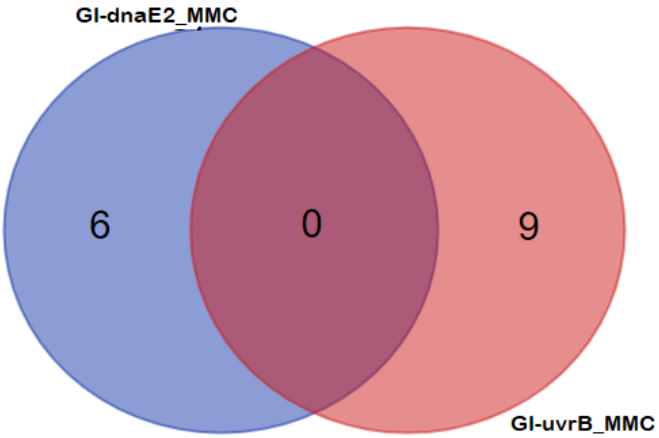
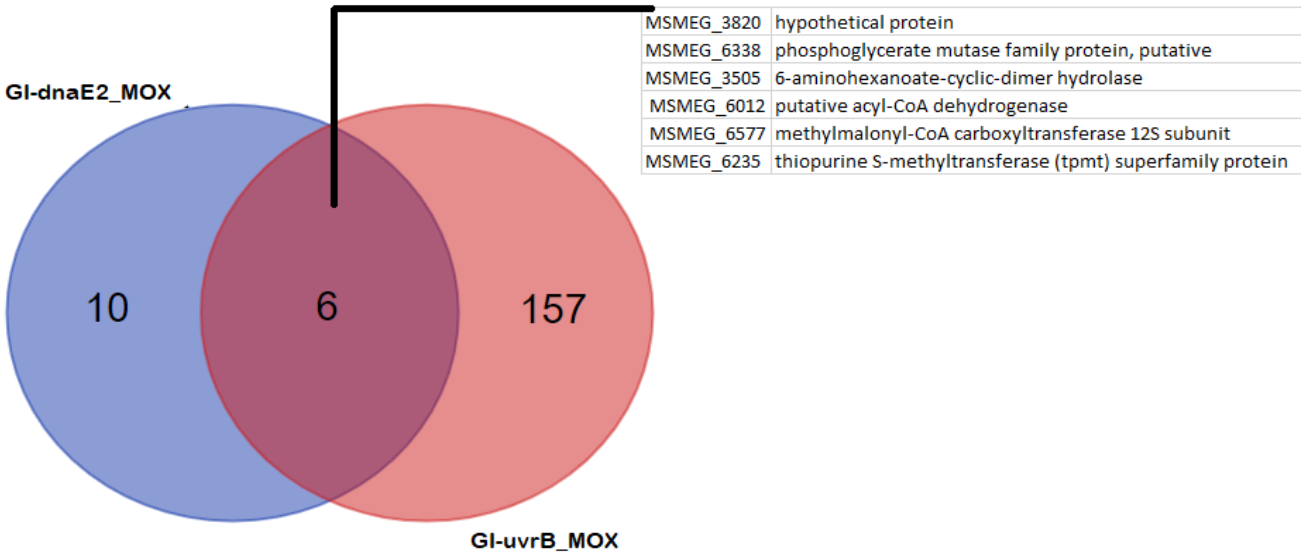


Figure 4. 24 Venn diagram comparing gene sets predicted to have aggravating genetic interactions (GI) with *dnaE2* and *uvrB* during MMC treatment.



MSMEG_3820	hypothetical protein
MSMEG_6338	phosphoglycerate mutase family protein, putative
MSMEG_3505	6-aminohexanoate-cyclic-dimer hydrolase
MSMEG_6012	putative acyl-CoA dehydrogenase
MSMEG_6577	methylmalonyl-CoA carboxyltransferase 12S subunit
MSMEG_6235	thiopurine S-methyltransferase (tpmt) superfamily protein

Figure 4. 25 Venn diagram comparing gene sets predicted to have aggravating genetic interactions with *dnaE2* and *uvrB* during MOX treatment.

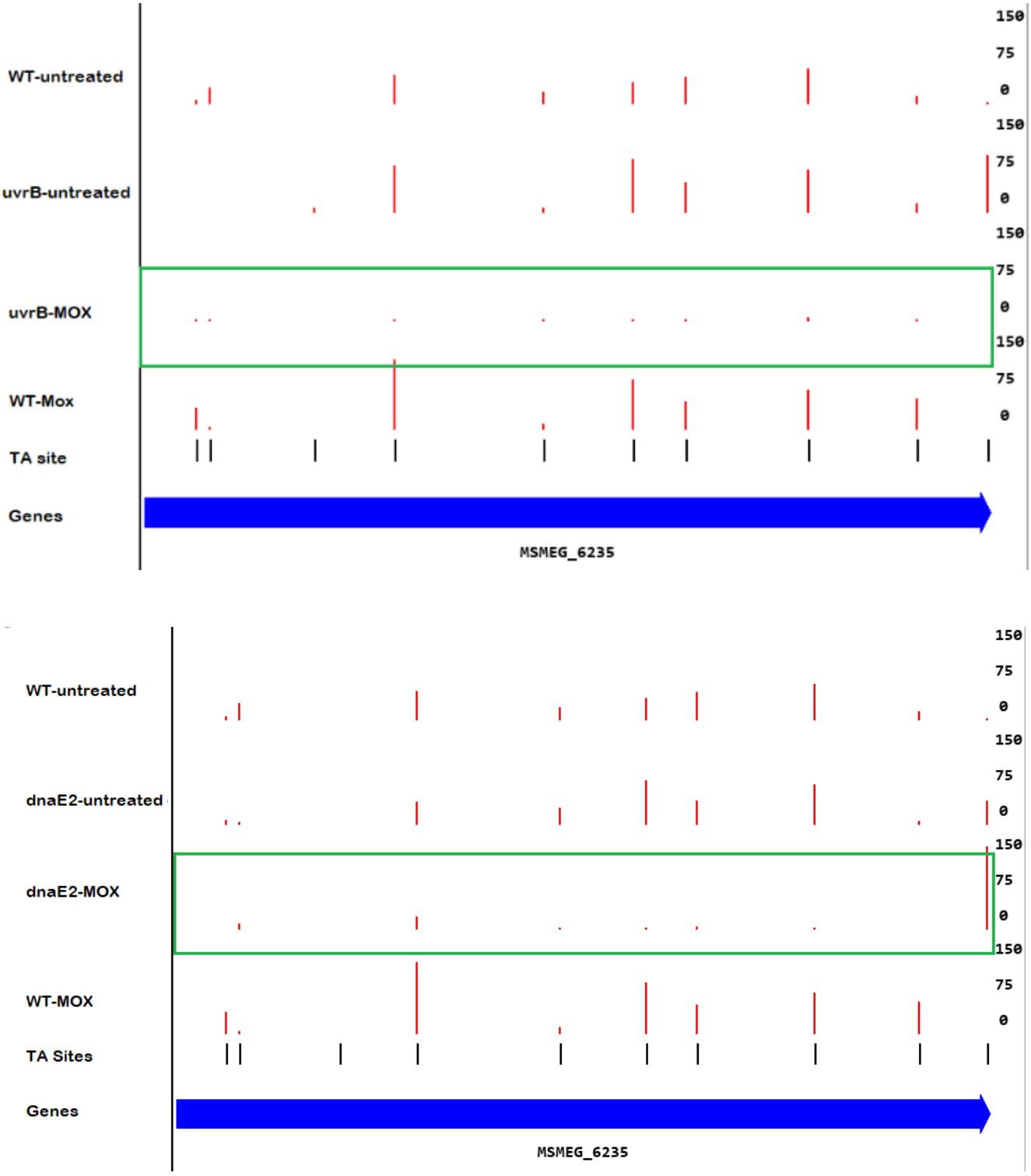
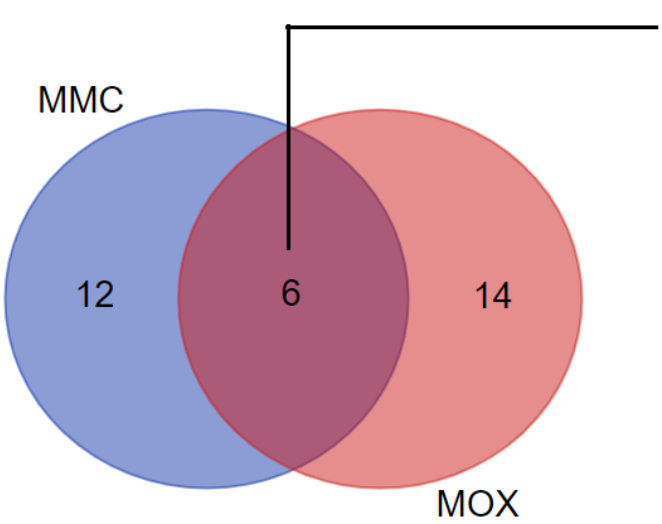


Figure 4. 26 Distribution of Tn insertions in MSMEG_6235, a gene exhibiting aggravating genetic interactions with *uvrB* (upper panel) and *dnaE2* (bottom panel) during MOX treatment. Thick blue arrows show the gene and the red lines indicate Tn insertion frequencies in the different -TA- sites (short black lines). Green outlined rectangles show Tn insertions in MSMEG_6235 in each library of interest. Numbers on the right represent the magnitudes of the Tn insertion frequencies. Images were generated using TRANSIT software.

4.3.10. Genes which are conditionally essential in MMC and MOX treatment independent of bacterial genotype

We found 18 genes that were conditionally essential during MMC treatment independent of the bacterial genotype, while 20 such genes were identified for MOX treatment (**Figure 4.15**). **Figure 4.27** presents a comparison of the respective gene lists which was conducted to show genes which are conditionally essential in both treatments and irrespective of the bacterial strain used. We found 6 such genes which included *pimE*, a mannosyltransferase protein which functions in phospholipid biosynthesis; *asnB*, an asparagine synthase (glutamine hydrolysing) which is required for biosynthesis of asparagine; and *pcrA*, an ATP-dependent DNA helicase predicted to operate in DNA repair and replication (**Figure 4.27**). Among these genes, asnB was almost devoid on Tn insertions in both conditions and in all strains. The distributions of Tn insertions in *asnB* in all the libraries are shown in **Figure 4.28**.



ORF	Name	Description
MSMEG_5149	pimE	mannosyltransferase
MSMEG_6398		antigen 85-A
MSMEG_4269	asnB	asparagine synthase (glutamine-hydrolyzing)
MSMEG_5534	pcrA	ATP-dependent DNA helicase PcrA
MSMEG_1654		isocitrate dehydrogenase, NADP-dependent
MSMEG_5041		acyltransferase

Figure 4. 27 A Venn diagram comparing gene sets predicted to be conditionally essential in MMC and MOX treatments independent of the strain used

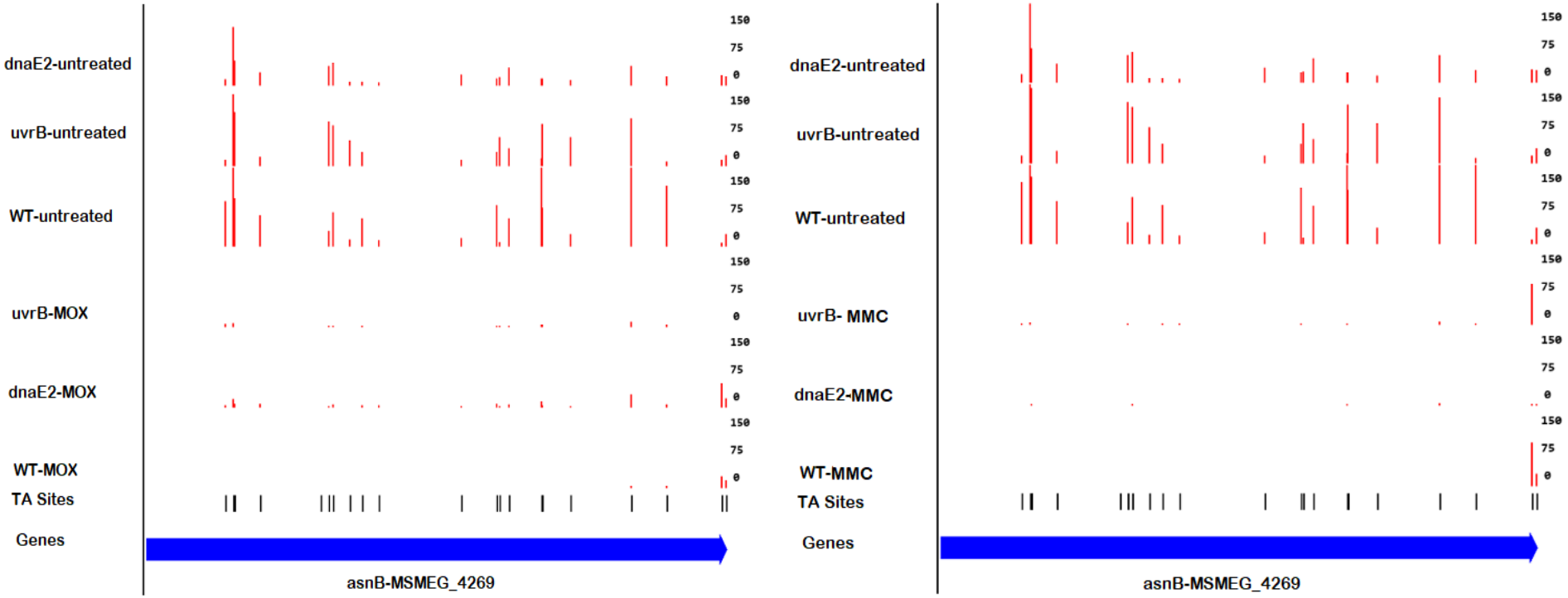


Figure 4. 28 Distribution of Tn insertions in MSMEG_4269 (*asnB*) under MOX (upper panel) and MMC (lower panel) treatments in WT, Δ*dnaE2* and Δ*uvrB* strains. Thick blue arrows show the gene and the red lines indicate Tn insertion frequencies in the different -TA- sites (short black lines). Green outlined rectangles show Tn insertions in MSMEG_6235 in a library of interest. Numbers on the right represent the magnitude of the Tn insertion frequencies. Few or no insertions in a condition indicate essentiality. Images were generated using TRANSIT software.

4.4 Discussion

The DNA damage response in mycobacteria consists of the classical RecA/LexA SOS response and the PafBC-regulated response (Müller et al., 2018) . DnaE2 is part of the RecA/LexA-dependent response whereas UvrB, an NER enzyme, falls into the PafBC-regulated pathway. While these two DNA damage response proteins have been proposed as potential drug targets (Merrikh and Kohli, 2020) putative compensatory and interacting pathways have not been elucidated. This study applied random whole-genome mutagenesis of defined strains to identify such pathways following exposure of cells to MMC, a DNA cross-linking agent and MOX, a DNA gyrase inhibitor.

Phenotypic characterization of *dnaE2* and *uvrB* deletion mutants

The Δ*uvrB* mutant was hypersusceptible to MMC compared to WT and the Δ*dnaE2* mutant, an observation consistent with a review by Noll *et al*. (who reported that in both prokaryotes and eukaryotes repair of inter-strand crosslinks involves the NER and homologous recombination pathways. Furthermore, it has been suggested that UvrB in particular is responsible for the recognition of the inter-strand cross-link (Noll et al., 2006), hence the hypersensitivity of the Δ*uvrB* mutant to MMC despite possessing an intact homologous recombination repair pathway. The *dnaE2* mutant, on the other hand, was two-fold more susceptible to MMC treatment compared to WT, suggesting that TLS is important but not critical for tolerance of MMC DNA damage. WT, Δ*uvrB* and Δ*dnaE2* all had comparable MICs for MOX. On its own, this result implied that the damage induced by MOX was not primarily withstood by DnaE2-dependent TLS nor UvrB-dependent NER; however, the very large number of conditionally essential genes inferred from the MOX-treated Δ*uvrB* libraries was difficult to reconcile with a minor role for NER in reversing MOX genotoxicity. Instead, this result suggests that NER is a major contributor to mycobacterial survival under MOX. Moreover, by analogy with the minimum duration for kill (MDK) metric recently proposed for tolerant/persister populations(Brauner et al., 2016), it suggests that an analysis of the rate of kill under MOX treatment, rather than a simplistic MIC determination, might be more revealing.

TnSeq library generation and selection

According to DeJesus *et al.*, a good library should have genome saturation /complexity of >50%. This gives statistical confidence as predictions of essentiality from high-density libraries are more reliable owing to the fact that missing insertions are more likely a result of selection than chance (DeJesus et al., 2015). Hence, it is important to have a good library complexity even before selection. Furthermore, one of the assumptions of the TnSeq method is that, during library selection, only mutants with attenuated survival should decrease in frequency. Therefore, it is important to maintain the complexity of the library during selection in order to avoid unspecific loss of mutants (Chao et al., 2016). In this study, 2 million bacterial cells per library were exposed to selective pressure. According to Long *et al*., applying selective pressure on a population of >1million cells is sufficient to maintain complexity in a library that contains at least 50,000 independent mutants (Long et al., 2015). Even though we treated about 2 million cells from each library with >50 % genome coverage, we observed some differences in library complexity between untreated and treated libraries; *e.g*., an untreated Δ*dnaE2* library had 57.7% genome coverage but, after treatment with MMC, the coverage was reduced to 39.9% (Table 3.11 and Table 4.0). Unspecific (stochastic) loss of mutants during selection can occur during library preparation for sequencing. Processes which include biased PCR reactions, excessive primer dimers, vector and adapter contamination, and poor sequencing can all affect genome complexity (DeJesus, 2019). Perhaps more importantly, especially in the case of drug treatment, the degree of selection depends on the strength of drug pressure on the library mutants and on the amount of time the pressure was applied (Long et al., 2015). In this study, the libraries were treated at ¼×MIC for each in order to minimise loss of mutants owing to high drug concentration.

Conditionally essential genes in MMC and MOX treatments

Identifying conditionally essential genes with TnSeq is enabled by comparing Tn libraries in two different conditions. In this case, conditional essentiality of a region is depicted by lack of Tn insertions in one condition but not in another. This analysis is important as it is a way of identifying drug targets or working out pathways (DeJesus et al., 2015). The results of this study show that for both MOX and MMC treatments in all strains, genes that function in cell wall and cell processes and those in intermediary metabolism and respiration are abundant (**Figure 4.7** and **Figure 4.19**). The "cell wall and cell processes" functional category includes proteins involved in cell division and cell cycle, cell wall and cell membrane biogenesis,

transmembrane transporters etc. (https://mycobrowser.epfl.ch/). These represent proteins that offer the protective barrier of the cell and those involved in transport of substances into and outside the cell. The complexity of the cell wall is a major feature of mycobacterial physiology, including its contribution to intrinsic resistance. Therefore, enzymes involved in the biosynthesis, remodelling, recycling and degradation of the cell wall are potential drug targets for anti-TB treatment (Maitra et al., 2019). Intrinsic resistance owing to the impermeability of the cell wall usually works in synergy with active efflux of drugs (Viveiros et al., 2012). It then follows that disruption of these functions will either make it easy for drugs to enter the cell or difficult for the cell to extrude them via efflux.

The "intermediary metabolism and respiration" functional group involves pathways that function in anabolic and catabolic reactions. In bacteria, metabolism is a key determinant of pathogenicity as it includes pathways that enable the bacteria to respond to stressful conditions which may include nutrient unavailability and exposure to drugs; susceptibility to specific drugs can be affected by shifts in metabolic states – *e.g*., growth arrest in MTB has been associated with tolerance to a number of antibiotics (Warner, 2015). Furthermore, some of the metabolism enzymes are involved in biotransformation reactions which may play a role in drug modifications and inactivation. This reactions include hydrolysis, N-alkylations, amidations and nitro reductions (Awasthi and Freundlich, 2017).

This study identified, among others, genes that belong to the phosphate specific transporter (Pst), *pstC* and *pstS* as conditionally essential in all the strains under MMC treatment. The *pimE* gene, which encodes a mannosyltransferase, was however conditionally essential in both MMC and MOX in all the strains used. These genes belong to the cell wall and cell processes functional category: phosphate specific transporters (Pst) belong to the ATP binding cassette family of permeases. While MTB has several copies of the Pst system, *M.smegmatis* only has a single copy organised as *pstSCAB* operon (Gebhard et al., 2006). In mycobacteria, the transcriptional response to phosphate limitation is mediated by the SenX3-RegX3 two-component system which requires a functional Pst phosphate uptake system. The two component system becomes functional in phosphate limiting conditions (Tischler et al., 2013). MTB requires its phosphate sensing signal transduction system for virulence and antibiotic tolerance (Namugenyi et al., 2017). Our results suggest that inactivating components of the Pst system disrupts the phosphate sensing system and leads to hypersensitivity to MMC treatment. It is worth noting that the 7H10 agar medium used in our experiments contains inorganic

phosphates: disodium phosphate and monopotassium phosphate at 1.50 g/ml, hence the cells were phosphate replete. It then follows that the Pst system is required to negatively regulate activity of the SenX3-RegX3 two-component system in response to available phosphate, failure of which results in the cells becoming more susceptible to environmental stress.

Another cell wall gene that was identified in both MMC and MOX treatments irrespective of strain background was Ag-85A. Proteins of the Ag-85A complex are crucial in formation of cord factor (Huygen, 2014). Cord factor is one of the major glycolipids that facilitates formation of the mycobacterial cell wall as formidable permeation barrier. MTB cell wall lipids prevent entry of solutes including antitubercular drugs which must penetrate the cell wall to reach their targets (Favrot and Ronning, 2012). Consistent with our results, deactivation of Ag-85A alters the cell integrity and makes the bacteria more susceptible to drug treatments.

Furthermore, Phosphate ABC transporters are well known for their role in efflux-dependent drug resistance in both prokaryotes and eukaryotes. Inactivation of *pstB* in *M.smegmatis* not only diminished phosphate uptake but also resulted in hypersensitivity to fluoroquinolones (Bhatt et al., 2000). The phosphate ABC transporter was not identified as essential in our MOX data. However, a closer look at the raw data showed a ~4-fold reduction in Tn insertions from 261 in untreated WT to 68 in moxifloxacin treated WT, suggesting that the statistical analysis was too stringent or its statistical significance was uncertain due to noise/variability in the insertion counts. Nonetheless, the data show that the phosphate ABC transporter is also likely to be essential for the response of MSM to the DNA damage caused by MOX treatment.

PimE is involved in phospholipid metabolism as part of the phosphatidylinositol (PI) pathway. PI is a precursor of phosphatidylinositol mannosides (PIMS) among other glycosylphosphatidylinositols and. accounts for about 25% of total phospholipids (Morita et al., 2006). PIM biosynthesis is achieved by sequential additions of mannoses onto PI, with PimE mediating the fifth mannose transfer in this sequence (Morita et al., 2006). These reactions play an important role in the generation of the mycobacterial cell wall. In MSM, deletion of *pimE* lead to changes in the cell wall hydrophobicity and plasma membrane organisation (Morita et al., 2006). The present study adds to this body of knowledge in showing that disruption of *pimE* renders MSM highly sensitive to MMC and MOX treatments, potentially because it enables enhanced drug penetration.

The asparagine synthase-encoding gene, *asnB*, was also identified as conditionally essential in both MMC and MOX treatments in all library strains investigated. AsnB has been implicated in conferring natural resistance in mycobacteria to multiple drugs: inactivation of *asnB* sensitizes MSM to rifampicin, novobiocin, erythromycin and fusidic acid (Ren and Liu, 2006). MSM contains two copies of *asnB*, MSMEG_4269 and MSMEG_2594c. Both are non-essential for *in vitro* growth in this study. However, their counterpart in MTB, Rv2201, is essential for *in vitro* growth (DeJesus et al., 2017a). Our results show that only MSMEG_4269 was conditionally essential in MOX and MMC. This is interesting because, although they are orthologs (which explains why they are both non-essential *in vitro*), exposure to drugs differentiates their response vis-à-vis conditional essentiality, implying they might have distinct functions and play different roles which suggests non-redundant roles in MSM.

Conditional essentiality of *uvrB* and *dnaE2* under MMC treatment

As expected, both *dnaE2* and *uvrB* were identified as conditionally essential in the MMC-treated WT library. Inter-strand crosslinks caused by MMC have been shown to be repaired and/or tolerated by both error-free and error-prone processes which draw heavily on proteins involved in NER and TLS (Noll et al., 2006). It has further been inferred that UvrB plays a role in recognising inter-strand crosslinks (Noll et al., 2006). Many studies in mycobacteria have shown that both *dnaE2* and *uvrB* are upregulated during MMC treatment (Rand et al., 2003, Boshoff et al., 2003, Namouchi et al., 2016, Müller et al., 2018). The hypothesis that NER might compensate for the loss of DnaE2 was confirmed by the significant decrease in Tn insertion counts in the *uvrB* gene in the Δ*dnaE2*-MMC library (**Figure 4.6 UPPER**). Surprisingly, however, the converse was not true: the analyses revealed that *dnaE2* does not appear to compensate for the loss of *uvrB*. This result might indicate a potential role for *dnaE2* in NER which is not essential, whereas the involvement of UvrB in *dnaE2*-dependent TLS is. Or perhaps UvrB enables DnaE2? Nonetheless this finding indicates that there are other compensatory mechanisms of MMC-mediated DNA damage repair in the absence of UvrB which are separate from TLS. In this context, it is interesting to note that *recG*, a HR gene, was essential in Δ*uvrB* under MMC treatment.

In general, the results reported here showed that most information pathways genes that were conditionally essential during MMC treatment in all library strains were those with helicase activity, *e.g.*, MSMEG_1930, DEAD/DEATH box helicase, and ATP-dependent DNA

helicase, PcrA, both of which were essential in MMC independent of strain background. DNA helicases catalyse the unwinding of duplex DNA molecules. They are important in transcription, replication, recombination, and DNA repair. Since MMC causes inter-strand cross links, and given the function of UvrB as NER helicase, it is not surprising that there is an over-reliance on DNA helicases during treatment with this agent.

Conditional essentiality of *uvrB* and *dnaE2* under MOX treatment

Contrary to what was observed under MMC treatment, *dnaE2* and *uvrB* were both non-essential during treatment of WT with MOX. This finding highlights the core differences in the type of DNA damage caused by MOX compared to MMC. While MMC causes DNA crosslinks (Tomasz and Palom, 1997), MOX causes chromosomal fragmentation and accumulation of dsDNA fragments (Drlica et al., 2008). Double-strand breaks (DSBs) in DNA are mostly repaired by HR, NHEJ and SSA (Gupta et al., 2011). However, members of these pathways were not conditionally essential in WT_MOX. This observation led to the speculation that the concentration of MOX we used might have been too low to elicit an appropriate damage response. Nonetheless, PcrA and Pol I were identified as conditionally essential in WT_MOX library. The orthologue of PcrA in MTB is Rv0949, a probable ATP-dependent DNA helicase which is thought to be involved in nucleotide excision repair (https://mycobrowser.epfl.ch/). Furthermore, *pcrA* was also found to be conditionally essential in MMC and MOX treatments in all strains highlighting PcrA as a critical helicase during DNA repair in *MSM*. Notably, expression data also show that PcrA is overexpressed during MMC treatment of WT *MSM* (Müller et al., 2018).

The analysis of conditionally essential genes under MOX treatment also showed that *uvrB* does not compensate the absence of *dnaE2*, an observation which is contrary to what was observed under MMC treatment. Like MMC, *dnaE2* does not appear to compensate the absence of *uvrB* under MOX. However, the distribution of Tn insertions in *dnaE2* (**Figure 4.18**) in the Δ*uvrB*-MOX library was also noteworthy for the apparent bias towards the C- terminus which might indicate a domain-specific functional requirement.

The results reported here indicate that the NER (*uvrB*) and TLS (*dnaE2*) may not be required during MOX treatment. This observation is supported by our MIC data for MOX which showed comparable values for WT, Δ*dnaE2* and Δ*uvrB*. Nonetheless, a very large number of conditionally essential genes were identified in the Δ*uvrB*-MOX library compared to the WT-

MOX and Δ*dnaE2*-MOX libraries (889 genes *versus* 29 and 64 for WT-MOX and *dnaE2*-MOX, respectively). This observation was puzzling given that the MIC for these strains was comparable. The large number of conditionally essential genes in uvrB-MOX may indicate several things: the Δ*uvrB* Tn library maybe less tolerant to ¼×MIC MOX treatment than WT and Δ*dnaE2* libraries, *i.e*., the Δ*uvrB* library is less tolerant to MOX than the parental strain; It could also indicate that MOX tolerance is different between the strains with the Δ*uvrB* strain being less tolerant compared to the other strains. Tolerant bacteria exhibit the same MIC as susceptible bacteria, hence MIC is not a suitable metric to evaluate tolerance. In the current work, a basic damage tolerance 'spotting' assay was carried out and the results between the strains were comparable. A more specific measure might be to determine the minimum duration of kill (MDK) (Brauner et al., 2016); It could also be due to higher noise in these samples; lower saturation, or imperfect normalization. Hence, further validations of these conditionally essential genes will be critical to verify such inconsistencies.

It was also interesting to find that, in both Δ*uvrB*-MOX and Δ*dnaE2*-MOX libraries, the *uvrD*-encoded helicase (MSMEG_1941) was conditionally essential. A study by Epshtein *et al*. showed that, in *E. coli*, UvrD induces backtracking of stalled RNA polymerase at lesions to allow the NER enzymes to gain access to damaged sites. These authors concluded that UvrD was important in resolving conflicts between transcription and DNA repair complexes (Epshtein et al., 2014). Our results clearly separate UvrD from the NER pathway, at least in *MSM*, because the enzyme is essential even in the absence of the NER pathway. Furthermore, the essentiality of *pcrA* (*uvrD1* in MTB) also emphasises the separation of *uvrD* from the NER pathway. There is also the possibility that UvrD resolves conflicts between transcription and DNA repair complexes from various repair pathways.

Another DNA repair gene identified in the Δ*dnaE2*-MOX library was *recO*, which has been shown to be involved in both HR and the SSA pathways (Gupta et al., 2015). These pathways were also conditionally essential in Δ*uvrB*-MOX library as indicated by the essentiality of recC, *recO* and *recG*, hence strengthening the observation that DNA damage elicited by MOX does not require NER and TLS but it is instead repaired by HR and SSA.

It is commonly known that active efflux and reduced influx play a critical role in fluroquinolone resistance (Alvarez et al., 2008). ABC transporters are among the major efflux mechanisms that facilitates fluroquinolone resistance (Alvarez et al., 2008). Numerous ABC transporter genes were found to be conditionally essential in the Δ*uvrB*-MOX library. This indicates that

inactivation of these transporters renders the bacteria more susceptible to MOX treatment because of lack of efflux of the drug. However, in the WT-MOX library, no ABC transporter-encoding genes were found to be essential, instead the *mmpL4* transmembrane transport gene was conditionally essential. In *dnaE2*-MOX, an ABC type transporter, MSMEG_4702 and a potassium transporting ATPase, *kdpA* were conditionally essential, thus implicating these genes in possible transport of MOX into or out of the bacterial cells.

CRISPRi validation of conditionally essential genes

TnSeq is a screening method whose accuracy and predictive power can be affected by several factors. These include library saturation, sequencing depth and normalisation of reads counts (Chao et al., 2016). Therefore, gene function leads must be followed up with other, often lower throughput approaches, *e.g*., by creating gene knockouts using allelic exchange mutagenesis. To validate our TnSeq data, we employed the recently published anhydrotetracycline (ATc) inducible CRISPR-dCas9 technology for mycobacteria (Rock et al., 2017). Unlike the conventional allelic exchange method, CRISPRi dCas9 is a simple and fast platform for regulated gene silencing (Rock et al., 2017). Transposon mutagenesis inserts a transposon element which is >1353 bp in a reading frame of a gene hence leading to non-functional alleles (DeJesus et al., 2017b) in a process known as insertional mutagenesis. Insertion mutations impact protein function in a variety of ways: a premature stop codon may be established leading to a truncated protein, transcripts maybe degraded, if translated, the protein may have no proper function due to frameshifting and so on. CRISPRi, on the other hand, interferes with transcription of the target gene by obstructing RNA polymerase from the target promoter or by blocking transcription elongation(Rock et al., 2017).

In validating some of the TnSeq leads, MSMEG_4259, a probable DnaQ-UvrC fusion protein (https://mycobrowser.epfl.ch/) was selected first. In unpublished results, this gene has been shown to be essential during MMC treatment by other members of the MMRU (Z. Ditse & D. Griffault, personal commuication) . This gene therefore served as a an positive control for this study. The TnSeq analysis showed this gene to be conditionally essential under MMC treatment in a Δ*dnaE2* strain but not in the parental WT or a Δ*uvrB* mutant. However, CRISPRi showed the gene to be essential in both WT and Δ*dnaE2* strains but not in the Δ*uvrB* strain. It is not clear why the WT TnSeq data differ from CRISPRi and allelic gene deletion data, however, visual analysis of the Tn insertions in this gene showed a reduction in insertions from untreated WT to WT-MMC (175 to73 insertions, respectively) (Figure 4.8). Importantly, the

results implicate MSMEG_4259 as part of the NER pathway because our data shows that deletion of *uvrB* renders all the known NER proteins non-essential, including MSMEG_4259. The protein encoded by MSMEG_4259 is a PafBC regulated protein which was found to be overexpressed during MMC treatment of *MSM* (Müller et al., 2018). A selection of other genes, namely *pimE*, *pstA* and MSMEG_2868 were successfully validated by CRISPRi in this study. Together, these results highlight the utility of this technology method for validating hits identified TnSeq.

Genetic interaction analysis

Analysing TnSeq data using the resampling permutation test identified conditionally essential genes between two conditions usually. To find strain specific conditionally essential genes we carried out GI screens. Three common types of GI exist: (i) an alleviating mutation occurs when an additional mutation does not further affect the phenotype; this indicates that the two genes could be in the same pathway (Beyer et al., 2007)); (ii) suppressive interactions are those in which the double mutant survives better; *i.e*., an additional mutation reverses the initial growth defect - among other things, this could mean that the relationship between the two genes include detoxification or regulation (DeJesus et al., 2017b); and (iii) aggravating interactions are those in which the fitness of a double mutant is more severe than that of the single mutant (DeJesus et al., 2017b); these interactions are also called synergistic interactions because they could imply that the two genes function in parallel or redundant pathways (Beyer et al., 2007). The data generated in this study were analysed for genes exhibiting aggravating interactions with *dnaE2* or *uvrB* under MMC or MOX treatments. This analysis is important as it identifies potential combination therapy targets which the bacteria could use in order to evade therapies targeting UvrB or DnaE2.

Genes that antagonize loss of *uvrB* during MMC and MOX treatment

Genes that antagonize loss of *uvrB* during MMC treatment were identified. Among these were *glgB*, which encodes a glycogen branching enzyme, the mycothiol biosynthesis genes, MSMEG_0933 and MSMEG_5261c, and a conserved hypothetical gene, MSMEG_2788. GlgB plays a crucial role in the biosynthesis of alpha glucans (Dkhar et al., 2015) which produce cytosolic and capsular glucans involved in host-pathogen interactions and also function as energy sources during latent MTB infections (Kalscheuer and Jacobs Jr, 2010). In MTB, *glgB* is essential both *in vitro* and *in vivo* (Sambou et al., 2008). Essentiality of this gene

in *MSM* is uncertain, with this study indicating that it is non-essential while Dragset et.al indicate that it is essential(Dragset et al., 2019).It is not clear why and how loss of *glgB* exacerbates the sensitivity of the Δ*uvrB* mutant to MMC; however, this analysis exposes the limited knowledge of gene networks and highlights the need to study such interactions.

In mycobacteria, mycothiol functions to protect cells against damage caused by oxidative, nitrosative and xenobiotic stress, as well as maintaining an intracellular reducing environment (Hernick, 2013b). As a primary reducing agent, mycothiol protects against oxidative damage, a function which is served by glutathione in eukaryotes. In MTB, decreased mycothiol levels have been associated with enhanced sensitivity to oxidising agents and antibiotics including rifampicin and erythromycin, suggesting that targeting mycothiol might potentiate the action of rifampicin (Hernick, 2013a). On the other hand, the resistance of MTB to reactive nitrogen and oxygen species in mice was shown to be facilitated by the NER pathway through the action of UvrB (Darwin and Nathan, 2005). It is perhaps not surprising, therefore, that our data point to mycothiol biosynthesis as compensating the absence of UvrB. In turn, this implies that any therapy targeting UvrB should be combined with a mycothiol inhibitor to minimise the ability of mycobacteria to tolerate the drug treatment.

The analysis of aggravating genetic interactions for *uvrB* during MOX treatment identified 163 genes, ten times the number (16 genes) identified in the *dnaE2*-MOX library. The Δ*uvrB* library was more sensitive to ¼ ×MIC MOX treatment than WT and Δ*dnaE2*, as evident from the finding of ~800 conditionally essential genes under this treatment for Δ*uvrB* compared to 29 and 64 for WT and Δ*dnaE2*, respectively. As noted earlier, this was despite the fact that the MIC were the same for the parental strains though the absence of survival data here is considered a limitation that should be addressed in future work.

Most of the genes that had aggravating genetic interactions with Δ*uvrB* during MOX treatment were involved in intermediary metabolism and respiration. Interestingly, most of these function in redox reactions. It is not clear how these could be interacting with Δ*uvrB*; however, redox enzymes are known to facilitate detoxification of xenobiotics in mammalian cells (Wright, 2005). In bacteria, a well-known example is the oxidation of tetracycline by TetX enzyme in *Bacteroides fragilis* (Wright, 2005). Furthermore, new evidence suggests that interaction of antibiotics to their target does not fully account for pathogen killing. Instead, complex

metabolic perturbations results that involves the generation of ROS and other damaging molecules (Dwyer et al., 2015). In their recent review, Baquero and Levin further suggest that there are proximate and ultimate mechanisms of antibiotic killing which involves both metabolic changes and cellular responses, including the SOS response and ROS production (Baquero and Levin, 2020). Since ROS facilitates antibiotic induced bacterial cell death, antioxidant systems becomes important mechanisms which contributes to antibiotic resistance(Ren et al., 2020). In this context, the results reported here are not surprising. However, these observations do not explain the relationship between redox mechanisms and *uvrB*.

Interestingly, *recC*, a DNA repair gene was also identified as having aggravating interaction with Δ*uvrB* during MOX treatment. It has been shown that mycobacteria possess three DNA DSB repair pathways: HR, NHEJ and SSA (Gupta et al., 2011). RecC functions in SSA and is located in an operon encoding the RecA-independent helicase nuclease, RecBCD (Gupta et al., 2011). The results from this study indicate that SSA might compensate the absence of NER during MOX treatment, an interpretation reinforced by the observation that *recC* was non-essential in WT and Δ*dnaE2* libraries under both MMC and MOX treatments, and non-essential during MMC treatment in Δ*uvrB*. Hence, this study suggests that the interaction between Δ*uvrB* and *recC* is specific to MOX. It is surprising, however, that *recB* and *recD* did not show the same interaction with Δ*uvrB*. Perhaps this highlights the essential role of *recC* in the RecBCD helicase nuclease operon. Mutagenic repair/tolerance is generally considered a last resort during DNA damage: for example, inactivation of faithful HR does not correlate with increased mutagenic NHEJ (Gupta et al., 2011). Also, while inactivation of mutagenic SSA leads to increased HR, the converse is not true (Gupta et al., 2011). It is interesting, therefore, that inactivation of a non-mutagenic NER/UvrB is compensated by a mutagenic RecC/SSA pathway.

Genes that antagonize loss of *dnaE2* during MMC and MOX treatment

Genes that interact with Δ*dnaE2* during MMC and MOX treatment were also identified in this study. Under MMC treatment, severe aggravating interactions were observed for MSMEG_2868, a PadR family transcriptional regulator; MSMEG_0487c, an ABC transporter permease; and MSMEG_4025, a transcriptional regulator LysR family protein. The phenotype of MSMEG_2868 was validated by CRISPRi. We further observed that the phenotype might

be specific to the Δ*dnaE2* background under MMC treatment since we did not discern the same effect during MOX, GRY or NOV treatments. The first described PadR proteins were involved in regulating genes encoding phenolic acid decarboxylase. Further research has found that PadR family transcriptional regulators from *Lactococcus lactis*, *Bacillus cereus* and *Vibrio cholerae* function in antibiotic efflux mechanisms and regulation of virulence (Isom et al., 2016). In *Clostridium difficile*, PadR transcriptional regulator is part of a broad regulatory network which is involved in virulence, stress response and antibiotic resistance (Isom et al., 2016). A study carried out in *MSM* found that the bacillus is unable to grow on C-19 steroids which are intermediates of cholesterol catabolic pathway. However, mutations in MSMEG_2868 enabled *MSM* to grow on these steroids. It was further shown that MSMEG_2868 represses C-19+ gene cluster (MSMEG_2851-MSMEG_2901), which facilitates growth on C-19 steroids. Hence, deletion of MSMEG_2868 de-represses this cluster, enabling growth on C-19 steroids (Fernández-Cabezón et al., 2017). The C-19 gene cluster is widely distributed in environmental mycobacteria but it is absent in pathogenic mycobacteria (Fernández-Cabezón et al., 2017). However, Rv3488, a conserved non-essential hypothetical protein, has recently been assigned to the PadR family of transcriptional regulators. Expression of Rv3488 in *MSM* has been shown to increase intracellular survival and it is hypothesised to be involved in metal detoxification (Kumari et al., 2018). It is not clear why MSMEG_2868 was specifically identified as conditionally essential in the Δ*dnaE2* mutant under MMC treatment but the data imply that this protein might regulate genes that compensate the absence of *dnaE2* during MMC treatment.

Sixteen genes involved in aggravating genetic interactions with *dnaE2* during MOX treatment were also identified. However, visual inspection of Tn insertions in these genes showed that most of these interactions were probably due to background noise in the analysis highlighting the need for further validation of the TnSeq data. Nonetheless, a few genes, including MSMEG_6235, which encodes a thiopurine-S-methyltransferase superfamily conserved hypothetical protein and MSMEG_6012c, a putative acyl-CoA dehydrogenase, were observed to have genuine aggravating genetic interactions. While the orthologue of MSMEG_6012 *in* MTB is Rv3560c and is involved in lipid degradation (mycobrowser), Rv3699, the MTB orthologue of MSMEG_6235is a conserved protein of unknown function. Furthermore, it was found that MSMEG_6235 has aggravating interactions with both Δ*uvrB* and Δ*dnaE2* during MOX treatment (**Figure 4.26**). This highlights the importance of genetic interaction analysis in facilitating the annotation of genes of unknown or only predicted function.

Chapter 5 Concluding remarks

The inferred importance of DNA damage repair mechanisms in the adaptation of MTB as obligate human pathogen, including the development of drug resistance, supports their potential as drug targets (Reiche et al., 2017). In microbiological experiments and in animal infection models, DnaE2-dependent TLS and NER have consistently been identified as major DNA repair mechanisms, suggesting a likely role during MTB pathogenesis (Warner, 2010). Owing to the dominance of these mechanisms in studies to date, the doctoral work presented here sought to elucidate repair mechanisms which might buffer the loss of DnaE2 and UvrB in bacilli exposed to genotoxic stress. By exposing genome-wide Tn mutagenesis libraries of MSM mutants lacking either *dnaE2* or *uvrB,* conditionally essential repair pathways were identified under MMC and MOX treatments.

Baseline essentiality *versus* treated conditions.

In standard (unstressed) growth conditions, baseline gene essentiality calls in the three strains assayed - WT, Δ*dnaE2* and Δ*uvrB* - were mostly comparable. Notably, a potential discrepancy was observed in the inferred essentiality of the mycobacterial F_1F_0 ATP synthase in *dnaE2 versus* WT Tn libraries; however, this was ascribed to loss of a duplicated chromosomal region in the Δ*dnaE2* mutant. Intermediary metabolism and respiration, together with the information pathways categories, represented a large proportion of the essential gene list. These corresponded to genes that were involved in translation and amino acids biosynthesis, and are consistent with those found in MTB essential gene list (Sassetti et al., 2003). While metabolic pathways remained essential upon treatment with either MMC or MOX , cell wall and cell processes pathways also became very important unlike in untreated conditions. These represent proteins that constitute the protective barrier of the cell and are involved in transport of substances into and outside the cell (Maitra et al., 2019). It is tempting to speculate, therefore, that disruption of these functions will either make it easy for drugs to enter the mycobacterial cell or difficult for the cell to extrude the compounds via efflux. On the other hand, metabolic pathways could enable the bacteria to respond to stressful conditions; susceptibility to specific drugs can be affected by shifts in metabolic states (Warner, 2015). Metabolic pathways are also involved in biotransformation reactions which may play a role in drug modifications and inactivation (Awasthi and Freundlich, 2017). The study identified among others (**Figure 4.27**), PimE and AsnB, as conditionally essential under genotoxic stress. The results are consistent with other studies showing that both PimE and AsnB facilitate intrinsic antibiotic resistance to

multiple drugs in mycobacteria (Xu et al., 2017, Ren and Liu, 2006). These proteins too might be considered potential drug targets that can be used in combination therapy.

Does UvrB enable DnaE2 function under some conditions?

The results presented here confirm that TLS and NER are necessary for surviving MMC-mediated genotoxicity. It was notable, however, that while *uvrB* was identified as essential in both WT and *dnaE2* backgrounds, the opposite did not hold true: *dnaE2* was conditionally essential in WT but not the *uvrB* deletion mutant during MMC treatment. This was a surprising result. An explanation for this observation might be that DnaE2 and UvrB are functionally connected, such that loss of UvrB affects DnaE2 activity - *i.e.*, UvrB enables DnaE2 activity. Although there are no other data supporting this hypothesis currently, recent work in *Caulobacter* suggests an NER-coupled DnaE2 function that is active in non-replicating bacteria (Joseph et al., 2021). Further work is required to investigate the potential existence of an equivalent mechanism in mycobacteria; nevertheless, the possible connection between NER (UvrB) and TLS (DnaE2) components is tantalizing and reinforces the utility of unbiased Tn screens to elucidate previously cryptic functional interactions.

UvrB and DnaE2 as potential co-drug targets

The study uncovered a list of genes that interact with Δ*uvrB* and Δ*dnaE2* under both MMC and MOX treatments. These are genes whose deletion renders the resulting loss-of-function mutants more susceptible to genotoxic stress. Pathways that exacerbated the sensitivity of the Δ*uvrB* mutant to MMC included glycogen biosynthesis (*glgB*) and mycothiol biosynthesis (MSMEG_0933 and MSMEG_5261c). On the other hand, gene pathways involved in redox reactions were also found to make Δ*uvrB* more susceptible to MOX. While not necessarily predictable *a priori*, the potential that these (and other) metabolic pathways might interact with DNA repair highlights our limited knowledge of gene networks in mycobacteria as well as the utility of this and related whole-genome Tn screens to identify unpredicted gene-gene interactions. In this context, it is worth noting that the possible role of apparently unrelated metabolic pathways in susceptibility to genotoxic stress is putatively consistent with growing evidence that initial interactions of drugs with their targets cannot fully account for the antibiotic lethality (Dwyer et al., 2015) – and with very recent work implicating mutations in core metabolic pathways in antibiotic resistance (Lopatkin et al., 2021) . Albeit that validation of the Tn screen results is outstanding, the observations presented here imply that these

pathways are associated with NER and that mycothiol and *glgB* inhibitors could be used in combination therapy with UvrB targeting drugs.

The only DNA repair gene to compensate inactivation of NER/*uvrB* during MOX was *recC*. RecC is part of a mycobacterial RecBCD-dependent SSA pathway that repairs DSBs (Gupta et al., 2011). One implication of this observation is that inactivating a non-mutagenic pathway (NER) can be compensated by a mutagenic pathway (SSA), with potentially severe consequences for mutagenesis, including the development of drug-resistance mutations. While theoretical, this knowledge is important, too, for triaging potential DNA repair targets – the risk that inhibition of one pathway might be compensated by another, mutagenic pathway, should not be ignored.

In the context of TLS by DnaE2, this study showed that, under MMC treatment, severe aggravating interactions were observed with a PadR family transcriptional regulator, MSMEG_2868, and a transcriptional regulator from LysR family of proteins, MSMEG_4025, among other pathways (**Table 4.8**). These transcriptional regulators are involved in the stress response and virulence in bacteria (Isom et al., 2016, Maddocks and Oyston, 2008) and it was unexpected to find them interacting with TLS mechanisms. Although the mechanistic basis underlying the detected interaction remains unclear, follow-up work (**Figure 4.13**) indicated that targeting these pathways together with DnaE2 might limit resistant development.

Other key findings

Consistent with previous reports, our observations reiterated that TnSeq essentiality data do not necessarily correlate with transcriptional responsiveness; that is, there can be limited correlation between gene expression and requirement for survival (Rengarajan et al., 2005, Jensen et al., 2017): Of the 270 genes which were upregulated under MMC in MSM *(*Müller *et al.,* 2018*)*, only eight were found to be conditionally essential in our WT_MMC data (**Table 4.6**). Among other interactions suggested by this analysis, we identified MSMEG_4259, a probable *dnaQ-uvrC* fusion protein, as potential member of the NER pathway: MSMEG_4259 was 'essential' in both WT and Δ*dnaE2* strains but not in the Δ*uvrB* strain; deletion of *uvrB* was also shown to render all the known NER proteins non-essential, including MSMEG_4259. Furthermore, the results clearly separate UvrD from the NER pathway because *uvrD* remained essential even in the *uvrB* mutant, in which NER is crippled. A study by Epshtein *et al*. showed that, in *E. coli*, UvrD induces backtracking of stalled RNA polymerase at lesions to allow the NER enzymes to gain access to damaged sites. These authors concluded that UvrD was

important in resolving conflicts between transcription and DNA repair complexes (Epshtein et al., 2014). The data from our study points to the possibility that UvrD resolves conflicts between transcription and DNA repair complexes not only during the NER mechanism but from various other repair pathways. This study also highlighted the utility of TnSeq in guiding functional analyses of putative and hypothetical proteins as numerous genes with unknown functions were identified under the different conditions. The approach used in this study is also useful in studying genes that are essential in MTB but non-essential in MSM; *e.g.*, *asnB* encodes an asparagine biosynthesis protein that is essential in MTB (DeJesus et al., 2017a) but non-essential in MSM yet was identified as conditionally essential under MMC and MOX treatments.

TnSeq method considerations and limitations

It was interesting to observe that MIC results do not necessarily correspond to TnSeq results, *e.g.*, the *uvrB* mutant was more susceptible to MMC than the *dnaE2* mutant by MIC (**Figure 4.1**) whereas the TnSeq analysis indicated that, compared to NER, DnaE2-mediated TLS might be more critical in WT under this condition (**Figure 4.5**). On the same note, the *uvrB* mutant library was also more sensitive to MOX treatment than either WT or *dnaE2* libraries even though the MICs of all strains to MOX was the same. Further validations of these conditionally essential genes will be critical to verify the apparent discrepancies.

TnSeq is a screening method whose accuracy and predictive power can be affected by several factors including library saturation, sequencing depth and normalization of read counts (Chao et al., 2016). Hence, inferred gene functions/essentialities must be followed up with other, often lower-throughput, approaches – for example, targeted gene inactivation via allelic exchange mutagenesis. However, this study highlighted the utility of CRISPRi for the rapid validation of genes predicted essential via TnSeq. It was interesting, though, that this utility seemed to be confined only to essential genes (*i.e.*, those with no/almost no Tn insertion in a particular condition compared to another) but not to genes whose disruption led to reduced fitness (*i.e.*, those with only reduced Tn insertion in one condition compared to another). The reason for this discrepancy is not clear but might be ascribed to the difference between permanent inactivation of a gene via Tn disruption *versus* transient inhibition of gene expression owing to CRISPRi.

It is also worth noting that central carbon metabolism plays an important part in the physiology and virulence of mycobacteria (Rhee et al., 2011). Culture media (*e.g*., 7H10 supplemented with glycerol), were developed for optimal propagation of bacilli *in vitro*. The difference in carbon metabolism between bacteria *in vivo* compared to standard *in vitro* conditions is one of the reasons why target-based assays have limited predictive values to test novel compounds (Pethe et al., 2010). This means that *in vitro* gene essentiality is influenced by the carbon source used in the experiment and this essentiality might not be reproducible *in vivo*. Furthermore, it has been shown that potency of some drugs is modulated by carbon source (Kalia et al., 2019); *e.g*., Q203, a cytochrome bc_1:aa_3 inhibitor, was unable to inhibit mycobacterial growth on 7H10 agar supplemented with glycerol whereas it was potent when the same agar was supplemented with pyruvate instead (Kalia et al., 2019). For this reason, it is important to carry out follow-up studies to validate some of the leads discovered not only in MTB but also in a mouse model.

Conclusion and future studies

The results described here offer some new insights into the complexity of gene networks that point to the fact that initial interactions of drugs with their targets cannot fully account for antibiotic lethality. Most importantly, the study highlighted potential pathways that may buffer the loss of UvrB and DnaE2 in MSM (**Figure 5.1**). Furthermore, potential (potentiating) drug targets have also been uncovered by studying WT MSM. Since MSM is a close relative of MTB, we hypothesize that these drug targets can potentially be useful in MTB treatment. Therefore, follow-up work on the hits uncovered in this thesis is critical. Key areas for future studies should include: (1) Validation of MSM conditionally essential genes in the pathogenic mycobacterium, MTB; (2) Determination of gene pathways that interact and compensate both *uvrB* and *dnaE2* in a double KO Tn library of MTB (or using a combinatorial CRISPRi approach); (3) Determine the infectivity of MTB double KO in appropriate experimental models (e.g., ex vivo macrophage infection assays); and (4) Identification of small molecules that target MTB UvrB, DnaE2, AsnB, PimE and other potential drug targets.

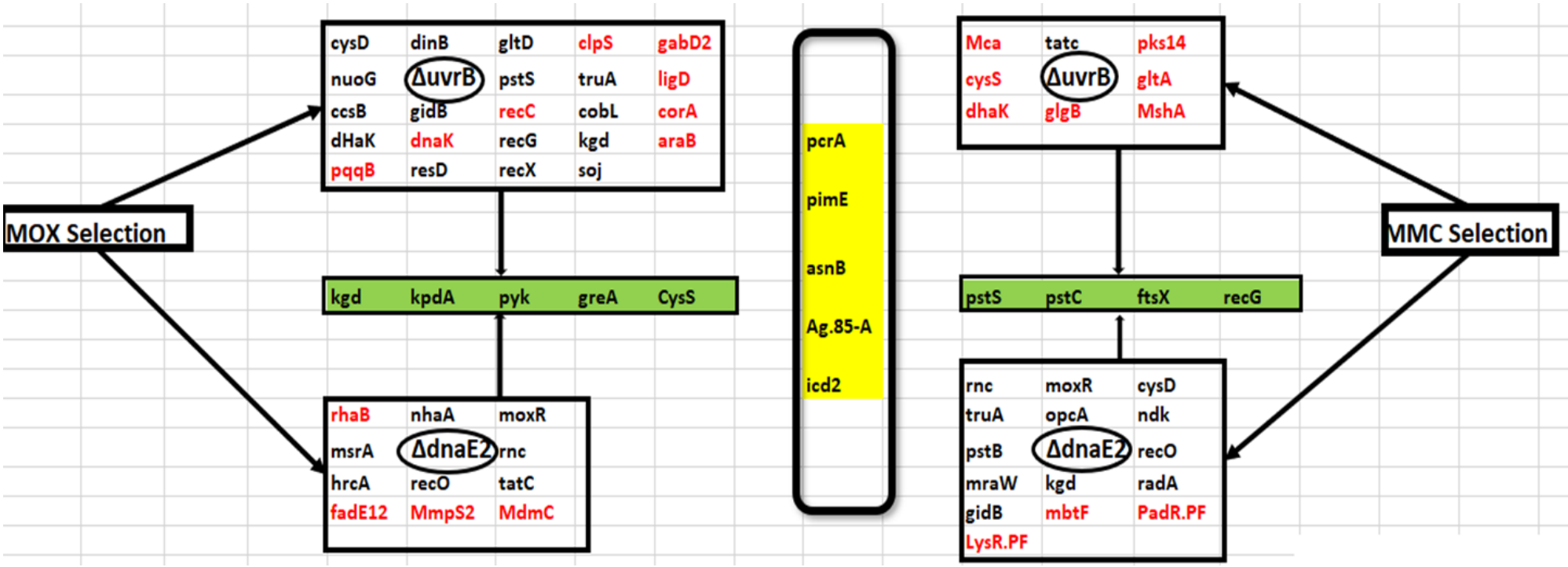


Figure 5. 1 Summary of key findings from the thesis. Genes that were CE in either MMC or MOX are show shown in black. Genes highlighted in green were CE is both strains under the respective treatments. Genes that had aggravating GI with either Δ*uvrB* or Δ*dnaE2* are shown in red. Genes highlighted in yellow were CE in both MMC and MOX treatments irrespective of strain background.

Supplementary information online

Supplementary data 1: https://figshare.com/s/1faa5544d2b6fc955061

Supplementary data 2: https://figshare.com/s/553c2193ee9836b7e4dc

Supplementary data 3: https://figshare.com/s/7aeda7a055cf14f43858

Supplementary data 4: https://figshare.com/s/0bf1f22a42e9384c2861

References

ABRAHAMS, K. A. & BESRA, G. S. 2018. Mycobacterial cell wall biosynthesis: a multifaceted antibiotic target. *Parasitology,* 145**,** 116-133.

ALDERWICK, L. J., HARRISON, J., LLOYD, G. S. & BIRCH, H. L. 2015. The Mycobacterial cell wall—peptidoglycan and Arabinogalactan. *Cold Spring Harbor perspectives in medicine,* 5**,** a021113.

ALI, M. K., LI, X., TANG, Q., LIU, X., CHEN, F., XIAO, J., ALI, M., CHOU, S.-H. & HE, J. 2017. Regulation of inducible potassium transporter KdpFABC by the KdpD/KdpE two-component system in Mycobacterium smegmatis. *Frontiers in microbiology,* 8**,** 570.

ALVAREZ, A. I., PÉREZ, M., PRIETO, J. G., MOLINA, A. J., REAL, R. & MERINO, G. 2008. Fluoroquinolone efflux mediated by ABC transporters. *Journal of pharmaceutical sciences,* 97**,** 3483-3493.

AMATO, S. M. & BRYNILDSEN, M. P. 2014. Nutrient transitions are a source of persisters in Escherichia coli biofilms. *PloS one,* 9**,** e93110.

ANDERSSON, D. I. & HUGHES, D. 2014. Microbiological effects of sublethal levels of antibiotics. *Nature Reviews Microbiology,* 12**,** 465-478.

ANDERSSON, D. I., NICOLOFF, H. & HJORT, K. 2019. Mechanisms and clinical relevance of bacterial heteroresistance. *Nature Reviews Microbiology***,** 1.

AWASTHI, D. & FREUNDLICH, J. S. 2017. Antimycobacterial metabolism: Illuminating Mycobacterium tuberculosis biology and drug discovery. *Trends in microbiology,* 25**,** 756-767.

BAHAROGLU, Z. & MAZEL, D. 2014. SOS, the formidable strategy of bacteria against aggressions. *FEMS microbiology reviews,* 38**,** 1126-1145.

BALABAN, N. Q., HELAINE, S., LEWIS, K., ACKERMANN, M., ALDRIDGE, B., ANDERSSON, D. I., BRYNILDSEN, M. P., BUMANN, D., CAMILLI, A. & COLLINS, J. J. 2019. Definitions and guidelines for research on antibiotic persistence. *Nature Reviews Microbiology,* 17**,** 441-448.

BAQUERO, F. & LEVIN, B. R. 2020. Proximate and ultimate causes of the bactericidal action of antibiotics. *Nature Reviews Microbiology***,** 1-10.

BARRETT, T. C., MOK, W. W., MURAWSKI, A. M. & BRYNILDSEN, M. P. 2019. Enhanced antibiotic resistance development from fluoroquinolone persisters after a single exposure to antibiotic. *Nature communications,* 10**,** 1177.

BAYSAROWICH, J., KOTEVA, K., HUGHES, D. W., EJIM, L., GRIFFITHS, E., ZHANG, K., JUNOP, M. & WRIGHT, G. D. 2008. Rifamycin antibiotic resistance by ADP-ribosylation: structure and diversity of Arr. *Proceedings of the National Academy of Sciences,* 105**,** 4886-4891.

BERCOVIER, H. & VINCENT, V. 2001. Mycobacterial infections in domestic and wild animals due to Mycobacterium marinum, M. fortuitum, M. chelonae, M. porcinum, M. farcinogenes, M. smegmatis, M. scrofulaceum, M. xenopi, M. kansasii, M. simiae and M. genavense. *Revue scientifique et technique (International Office of Epizootics),* 20**,** 265-290.

BERNEY, M. & COOK, G. M. 2010. Unique flexibility in energy metabolism allows mycobacteria to combat starvation and hypoxia. *PloS one,* 5**,** e8614.

BEYER, A., BANDYOPADHYAY, S. & IDEKER, T. 2007. Integrating physical and genetic maps: from genomes to interaction networks. *Nature reviews genetics,* 8**,** 699.

BHATT, K., BANERJEE, S. K. & CHAKRABORTI, P. K. 2000. Evidence that phosphate specific transporter is amplified in a fluoroquinolone resistant Mycobacterium smegmatis. *European journal of biochemistry,* 267**,** 4028-4032.

BIGGER, J. 1944. Treatment of staphylococcal infections with penicillin by intermittent sterilisation. *The Lancet,* 244**,** 497-500.

BOSHOFF, H. I., REED, M. B., BARRY, C. E. & MIZRAHI, V. 2003. DnaE2 polymerase contributes to in vivo survival and the emergence of drug resistance in Mycobacterium tuberculosis. *Cell,* 113**,** 183-193.

BRADNER, W. 2001. Mitomycin C: a clinical update. *Cancer treatment reviews,* 27**,** 35-50.

BRANUM, M. E., REARDON, J. T. & SANCAR, A. 2001. DNA Repair Excision Nuclease Attacks Undamaged DNA A POTENTIAL SOURCE OF SPONTANEOUS MUTATIONS. *Journal of Biological Chemistry,* 276**,** 25421-25426.

BRAUNER, A., FRIDMAN, O., GEFEN, O. & BALABAN, N. Q. 2016. Distinguishing between resistance, tolerance and persistence to antibiotic treatment. *Nature Reviews Microbiology,* 14**,** 320-330.

BROOKS, P. C., MOVAHEDZADEH, F. & DAVIS, E. O. 2001. Identification of some DNA damage-inducible genes of Mycobacterium tuberculosis: apparent lack of correlation with LexA binding. *Journal of bacteriology,* 183**,** 4459-4467.

BYRNE, R. T., CHEN, S. H., WOOD, E. A., CABOT, E. L. & COX, M. M. 2014. Escherichia coli genes and pathways involved in surviving extreme exposure to ionizing radiation. *Journal of bacteriology,* 196**,** 3534-3545.

CADENA, A. M., FORTUNE, S. M. & FLYNN, J. L. 2017. Heterogeneity in tuberculosis. *Nature reviews Immunology,* 17**,** 691.

CAMBAU, E. & DRANCOURT, M. 2014. Steps towards the discovery of Mycobacterium tuberculosis by Robert Koch, 1882. *Clinical Microbiology and Infection,* 20**,** 196-201.

CAREY, A. F., ROCK, J. M., KRIEGER, I. V., CHASE, M. R., FERNANDEZ-SUAREZ, M., GAGNEUX, S., SACCHETTINI, J. C., IOERGER, T. R. & FORTUNE, S. M. 2018. TnSeq of Mycobacterium tuberculosis clinical isolates reveals strain-specific antibiotic liabilities. *PLoS pathogens,* 14**,** e1006939.

CASTAÑEDA-GARCÍA, A., BLÁZQUEZ, J. & RODRÍGUEZ-ROJAS, A. 2013. Molecular mechanisms and clinical impact of acquired and intrinsic fosfomycin resistance. *Antibiotics,* 2**,** 217-236.

CASTAÑEDA-GARCÍA, A., MARTÍN-BLECUA, I., CEBRIÁN-SASTRE, E., CHINER-OMS, A., TORRES-PUENTE, M., COMAS, I. & BLÁZQUEZ, J. 2020. Specificity and mutagenesis bias of the mycobacterial alternative mismatch repair analyzed by mutation accumulation studies. *Science advances,* 6**,** eaay4453.

CASTANEDA-GARCIA, A., PRIETO, A., RODRIGUEZ-BELTRAN, J., ALONSO, N., CANTILLON, D., COSTAS, C., PÉREZ-LAGO, L., ZEGEYE, E., HERRANZ, M. & PLOCIŃSKI, P. 2017. A non-canonical mismatch repair pathway in prokaryotes. *Nature communications,* 8**,** 14246.

CECCHINI, G. 2013. Respiratory complex II: role in cellular physiology and disease. *BBA-Bioenergetics,* 5**,** 541-542.

CHAO, M. C., ABEL, S., DAVIS, B. M. & WALDOR, M. K. 2016. The design and analysis of transposon insertion sequencing experiments. *Nature Reviews Microbiology,* 14**,** 119-128.

COLE, S., BROSCH, R., PARKHILL, J., GARNIER, T., CHURCHER, C., HARRIS, D., GORDON, S., EIGLMEIER, K., GAS, S. & BARRY III, C. 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature,* 393**,** 537.

COSTANZO, M., BARYSHNIKOVA, A., BELLAY, J., KIM, Y., SPEAR, E. D., SEVIER, C. S., DING, H., KOH, J. L., TOUFIGHI, K. & MOSTAFAVI, S. 2010. The genetic landscape of a cell. *Science,* 327**,** 425-431.

COX, G. & WRIGHT, G. D. 2013. Intrinsic antibiotic resistance: mechanisms, origins, challenges and solutions. *International Journal of Medical Microbiology,* 303**,** 287-292.

CRAIG, N. L. 1997. Target site selection in transposition. *Annual review of biochemistry,* 66**,** 437-474.

DABBS, E. R., YAZAWA, K., MIKAMI, Y., MIYAJI, M., MORISAKI, N., IWASAKI, S. & FURIHATA, K. 1995. Ribosylation by mycobacterial strains as a new mechanism of rifampin inactivation. *Antimicrobial Agents and Chemotherapy,* 39**,** 1007-1009.

DARTOIS, V. 2014. The path of anti-tuberculosis drugs: from blood to lesions to mycobacterial cells. *Nature Reviews Microbiology,* 12**,** 159-167.

DARWIN, K. H., EHRT, S., GUTIERREZ-RAMOS, J.-C., WEICH, N. & NATHAN, C. F. 2003. The proteasome of Mycobacterium tuberculosis is required for resistance to nitric oxide. *Science,* 302**,** 1963-1966.

DARWIN, K. H. & NATHAN, C. F. 2005. Role for nucleotide excision repair in virulence of Mycobacterium tuberculosis. *Infection and immunity,* 73**,** 4581-4587.
DATTA, M., VIA, L. E., KAMOUN, W. S., LIU, C., CHEN, W., SEANO, G., WEINER, D. M., SCHIMEL, D., ENGLAND, K. & MARTIN, J. D. 2015. Anti-vascular endothelial growth factor treatment normalizes tuberculosis granuloma vasculature and improves small molecule delivery. *Proceedings of the National Academy of Sciences,* 112**,** 1827-1832.
DAVIS, E. O. & FORSE, L. N. 2009. DNA repair: key to survival. *Mycobacterium: Genomics and Molecular Biology***,** 79-119.
DAVIS, J. M. & RAMAKRISHNAN, L. 2009. The role of the granuloma in expansion and dissemination of early tuberculous infection. *Cell,* 136**,** 37-49.
DE WET, T. J., GOBE, I., MHLANGA, M. M. & WARNER, D. F. 2018. CRISPRi-Seq for the Identification and Characterisation of Essential Mycobacterial Genes and Transcriptional Units. *bioRxiv***,** 358275.
DEJESUS, M. A. 2019. transit Documentation.
DEJESUS, M. A., AMBADIPUDI, C., BAKER, R., SASSETTI, C. & IOERGER, T. R. 2015. TRANSIT-a software tool for Himar1 TnSeq analysis. *PLoS computational biology,* 11**,** e1004401.
DEJESUS, M. A., GERRICK, E. R., XU, W., PARK, S. W., LONG, J. E., BOUTTE, C. C., RUBIN, E. J., SCHNAPPINGER, D., EHRT, S. & FORTUNE, S. M. 2017a. Comprehensive essentiality analysis of the Mycobacterium tuberculosis genome via saturating transposon mutagenesis. *MBio,* 8**,** e02133-16.
DEJESUS, M. A. & IOERGER, T. R. 2013. A Hidden Markov Model for identifying essential and growth-defect regions in bacterial genomes from transposon insertion sequencing data. *BMC bioinformatics,* 14**,** 303.
DEJESUS, M. A. & IOERGER, T. R. Reducing type I errors in Tn-Seq experiments by correcting the skew in read count distributions. 7th International Conference on Bioinformatics and Computational Biology (BICoB 2015), 2015.
DEJESUS, M. A., NAMBI, S., SMITH, C. M., BAKER, R. E., SASSETTI, C. M. & IOERGER, T. R. 2017b. Statistical analysis of genetic interactions in Tn-Seq data. *Nucleic acids research,* 45**,** e93-e93.
DEWACHTER, L., FAUVART, M. & MICHIELS, J. 2019. Bacterial heterogeneity and antibiotic survival: understanding and combatting persistence and heteroresistance. *Molecular Cell,* 76**,** 255-267.
DITSE, Z., LAMERS, M. H. & WARNER, D. F. 2017a. DNA replication in Mycobacterium tuberculosis. *Microbiology spectrum,* 5.
DITSE, Z., LAMERS, M. H. & WARNER, D. F. 2017b. DNA Replication in Mycobacterium tuberculosis. *Tuberculosis and the Tubercle Bacillus***,** 581-606.
DKHAR, H. K., GOPALSAMY, A., LOHARCH, S., KAUR, A., BHUTANI, I., SAMINATHAN, K., BHAGYARAJ, E., CHANDRA, V., SWAMINATHAN, K. & AGRAWAL, P. 2015. Discovery of Mycobacterium tuberculosis α-1, 4-glucan branching enzyme (GlgB) inhibitors by structure-and ligand-based virtual screening. *Journal of Biological Chemistry,* 290**,** 76-89.
DOMENECH, P., KOLLY, G. S., LEON-SOLIS, L., FALLOW, A. & REED, M. B. 2010. Massive gene duplication event among clinical isolates of the Mycobacterium tuberculosis W/Beijing family. *Journal of bacteriology,* 192**,** 4562-4570.
DÖRR, T., VULIĆ, M. & LEWIS, K. 2010. Ciprofloxacin causes persister formation by inducing the TisB toxin in Escherichia coli. *PLoS biology,* 8**,** e1000317.
DOS VULTOS, T., BLÁZQUEZ, J., RAUZIER, J., MATIC, I. & GICQUEL, B. 2006. Identification of Nudix Hydrolase Family Members with an Antimutator Role in Mycobacterium tuberculosis and Mycobacterium smegmatis. *Journal of Bacteriology,* 188**,** 3159-3161.
DRAGSET, M. S., IOERGER, T. R., ZHANG, Y. J., MÆRK, M., GINBOT, Z., SACCHETTINI, J. C., FLO, T. H., RUBIN, E. J. & STEIGEDAL, M. 2019. Genome-wide Phenotypic Profiling Identifies and Categorizes Genes Required for Mycobacterial Low Iron Fitness. *Scientific reports,* 9**,** 1-11.

DRAIN, P. K., BAJEMA, K. L., DOWDY, D., DHEDA, K., NAIDOO, K., SCHUMACHER, S. G., MA, S., MEERMEIER, E., LEWINSOHN, D. M. & SHERMAN, D. R. 2018. Incipient and subclinical tuberculosis: a clinical review of early stages and progression of infection. *Clinical microbiology reviews,* 31.

DRLICA, K., MALIK, M., KERNS, R. J. & ZHAO, X. 2008. Quinolone-mediated bacterial death. *Antimicrobial agents and chemotherapy,* 52**,** 385-392.

DUPUY, P., HOWLADER, M. & GLICKMAN, M. S. 2020. A multilayered repair system protects the mycobacterial chromosome from endogenous and antibiotic-induced oxidative damage. *Proceedings of the National Academy of Sciences,* 117**,** 19517-19527.

DUPUY. A, S. A. 2015. DNA damage and gene therapy for xeroderma pigmentosum, a human DNA repair- deficient disease. *Mutation Research,* 776**,** 2-8.

DWYER, D. J., BELENKY, P. A., YANG, J. H., MACDONALD, I. C., MARTELL, J. D., TAKAHASHI, N., CHAN, C. T., LOBRITZ, M. A., BRAFF, D. & SCHWARZ, E. G. 2014. Antibiotics induce redox-related physiological alterations as part of their lethality. *Proceedings of the National Academy of Sciences,* 111**,** E2100-E2109.

DWYER, D. J., COLLINS, J. J. & WALKER, G. C. 2015. Unraveling the physiological complexities of antibiotic lethality. *Annual review of pharmacology and toxicology,* 55**,** 313-332.

EHRT, S. & SCHNAPPINGER, D. 2009. Mycobacterial survival strategies in the phagosome: defence against host stresses. *Cellular microbiology,* 11**,** 1170-1178.

ELDHOLM, V. & BALLOUX, F. 2016. Antimicrobial Resistance in Mycobacterium tuberculosis: The Odd One Out. *Trends in microbiology*.

ENGELS, S., LUDWIG, C., SCHWEITZER, J. E., MACK, C., BOTT, M. & SCHAFFER, S. 2005. The transcriptional activator ClgR controls transcription of genes involved in proteolysis and DNA repair in Corynebacterium glutamicum. *Molecular microbiology,* 57**,** 576-591.

EPSHTEIN, V., KAMARTHAPU, V., MCGARY, K., SVETLOV, V., UEBERHEIDE, B., PROSHKIN, S., MIRONOV, A. & NUDLER, E. 2014. UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature,* 505**,** 372-377.

ERILL, I., CAMPOY, S., MAZON, G. & BARBÉ, J. 2006. Dispersal and regulation of an adaptive mutagenesis cassette in the bacteria domain. *Nucleic acids research,* 34**,** 66-77.

FALZON, D., SCHÜNEMANN, H. J., HARAUSZ, E., GONZÁLEZ-ANGULO, L., LIENHARDT, C., JARAMILLO, E. & WEYER, K. 2017. World Health Organization treatment guidelines for drug-resistant tuberculosis, 2016 update. *European Respiratory Journal,* 49**,** 1602308.

FANGE, D., NILSSON, K., TENSON, T. & EHRENBERG, M. 2009. Drug efflux pump deficiency and drug target resistance masking in growing bacteria. *Proceedings of the National Academy of Sciences,* 106**,** 8215-8220.

FAVROT, L. & RONNING, D. R. 2012. Targeting the mycobacterial envelope for tuberculosis drug development. *Expert review of anti-infective therapy,* 10**,** 1023-1036.

FEDOROFF, N. 2001. How jumping genes were discovered. *Nature Structural & Molecular Biology,* 8**,** 300-301.

FEDOROFF, N. V. 2012. Transposable elements, epigenetics, and genome evolution. *Science,* 338**,** 758-767.

FERNÁNDEZ-CABEZÓN, L., GARCÍA-FERNÁNDEZ, E., GALÁN, B. & GARCÍA, J. L. 2017. Molecular characterization of a new gene cluster for steroid degradation in Mycobacterium smegmatis. *Environmental microbiology,* 19**,** 2546-2563.

FISHER, R. A., GOLLAN, B. & HELAINE, S. 2017. Persistent bacterial infections and persister cells. *Nature Reviews Microbiology,* 15**,** 453.

FONSECA, J., KNIGHT, G. & MCHUGH, T. 2015. The complex evolution of antibiotic resistance in Mycobacterium tuberculosis. *International Journal of Infectious Diseases,* 32**,** 94-100.

FORD, C. B., SHAH, R. R., MAEDA, M. K., GAGNEUX, S., MURRAY, M. B., COHEN, T., JOHNSTON, J. C., GARDY, J., LIPSITCH, M. & FORTUNE, S. M. 2013. Mycobacterium tuberculosis mutation rate

estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nature genetics,* 45**,** 784-790.
FRIEDBERG, E. C. 2005. Suffering in silence: the tolerance of DNA damage. *Nature Reviews Molecular Cell Biology,* 6**,** 943-953.
FRIEDBERG, E. C., AGUILERA, A., GELLERT, M., HANAWALT, P. C., HAYS, J. B., LEHMANN, A. R., LINDAHL, T., LOWNDES, N., SARASIN, A. & WOOD, R. D. 2006. DNA repair: from molecular mechanism to human disease. *DNA repair,* 5**,** 986-996.
FRIEDLI, M. & TRONO, D. 2015. The Developmental Control of Transposable Elements and the Evolution of Higher Species. *Annual review of cell and developmental biology,* 31**,** 429-451.
FUX, C., SHIRTLIFF, M., STOODLEY, P. & COSTERTON, J. 2005. Can laboratory reference strains mirror 'real-world'pathogenesis? *Trends in microbiology,* 13**,** 58-63.
GALAMBA, A., SOETAERT, K., WANG, X.-M., DE BRUYN, J., JACOBS, P. & CONTENT, J. 2001. Disruption of adhC reveals a large duplication in the Mycobacterium smegmatis mc2155 genome. *Microbiology,* 147**,** 3281-3294.
GAMULIN, V., CETKOVIC, H. & AHEL, I. 2004. Identification of a promoter motif regulating the major DNA damage response mechanism of Mycobacterium tuberculosis. *FEMS microbiology letters,* 238**,** 57-63.
GANDHI, N. R., NUNN, P., DHEDA, K., SCHAAF, H. S., ZIGNOL, M., VAN SOOLINGEN, D., JENSEN, P. & BAYONA, J. 2010. Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis. *The Lancet,* 375**,** 1830-1843.
GCEBE, N., MICHEL, A., GEY VAN PITTIUS, N. C. & RUTTEN, V. 2016. Comparative genomics and proteomic analysis of four non-tuberculous Mycobacterium species and Mycobacterium tuberculosis complex: occurrence of shared immunogenic proteins. *Frontiers in microbiology,* 7**,** 795.
GEBHARD, S., TRAN, S. L. & COOK, G. M. 2006. The Phn system of Mycobacterium smegmatis: a second high-affinity ABC-transporter for phosphate. *Microbiology,* 152**,** 3453-3465.
GEIMAN, D. E., RAGHUNAND, T. R., AGARWAL, N. & BISHAI, W. R. 2006. Differential gene expression in response to exposure to antimycobacterial agents and other stress conditions among seven Mycobacterium tuberculosis whiB-like genes. *Antimicrobial agents and chemotherapy,* 50**,** 2836-2841.
GLOVER, R. T., KRIAKOV, J., GARFORTH, S. J., BAUGHN, A. D. & JACOBS, W. R. 2007. The two-component regulatory system senX3-regX3 regulates phosphate-dependent gene expression in Mycobacterium smegmatis. *Journal of bacteriology,* 189**,** 5495-5503.
GOODMAN, M. F. 2002. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annual review of biochemistry,* 71**,** 17-50.
GOODMAN, M. F. & WOODGATE, R. 2013. Translesion DNA polymerases. *Cold Spring Harbor perspectives in biology,* 5**,** a010363.
GRAHAM, J. E. & CLARK-CURTISS, J. E. 1999. Identification of Mycobacterium tuberculosis RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). *Proceedings of the National Academy of Sciences,* 96**,** 11554-11559.
GRANT, S. S., KAUFMANN, B. B., CHAND, N. S., HASELEY, N. & HUNG, D. T. 2012. Eradication of bacterial persisters with antibiotic-generated hydroxyl radicals. *Proceedings of the National Academy of Sciences,* 109**,** 12147-12152.
GRAY, A. N., KOO, B.-M., SHIVER, A. L., PETERS, J. M., OSADNIK, H. & GROSS, C. A. 2015. High-throughput bacterial functional genomics in the sequencing era. *Current opinion in microbiology,* 27**,** 86-95.
GROMAN, N. B. & BOBB, D. 1955. The inhibition of adsorption of Corynebacterium diphtheriae phage by tween 80. *Virology,* 1**,** 313-323.
GUINN, K. M., HICKEY, M. J., MATHUR, S. K., ZAKEL, K. L., GROTZKE, J. E., LEWINSOHN, D. M., SMITH, S. & SHERMAN, D. R. 2004. Individual RD1-region genes are required for export of ESAT-

6/CFP-10 and for virulence of Mycobacterium tuberculosis. *Molecular microbiology,* 51**,** 359-370.
GUIRADO, E., SCHLESINGER, L. S. & KAPLAN, G. Macrophages in tuberculosis: friend or foe. Seminars in immunopathology, 2013. Springer, 563-583.
GUPTA, A., KAUL, A., TSOLAKI, A. G., KISHORE, U. & BHAKTA, S. 2012. Mycobacterium tuberculosis: immune evasion, latency and reactivation. *Immunobiology,* 217**,** 363-374.
GUPTA, R., BARKAN, D., REDELMAN-SIDI, G., SHUMAN, S. & GLICKMAN, M. S. 2011. Mycobacteria exploit three genetically distinct DNA double-strand break repair pathways. *Molecular microbiology,* 79**,** 316-330.
GUPTA, R., SHUMAN, S. & GLICKMAN, M. S. 2015. RecF and RecR play critical roles in the homologous recombination and single-strand annealing pathways of mycobacteria. *Journal of bacteriology,* 197**,** 3121-3132.
GUPTA, R., UNCIULEAC, M.-C., SHUMAN, S. & GLICKMAN, M. S. 2016. Homologous recombination mediated by the mycobacterial AdnAB helicase without end resection by the AdnAB nucleases. *Nucleic acids research,* 45**,** 762-774.
GÜTHLEIN, C., WANNER, R. M., SANDER, P., DAVIS, E. O., BOSSHARD, M., JIRICNY, J., BÖTTGER, E. C. & SPRINGER, B. 2009. Characterization of the mycobacterial NER system reveals novel functions of the uvrD1 helicase. *Journal of bacteriology,* 191**,** 555-562.
HARBOTTLE, J., MOSAEI, H., ALLENBY, N. & ZENKIN, N. 2021. Kanglemycin A can overcome rifamycin resistance caused by ADP-ribosylation by Arr protein. *Antimicrobial Agents and Chemotherapy***,** AAC. 00864-21.
HATA, T., SANO, Y., SUGAWARA, R., MATSUMAE, A., KANAMORI, K., SHIMA, T. & HOSHI, T. 1956. Mitomycin, a new antibiotic from Streptomyces. I. *The Journal of Antibiotics, Series A,* 9**,** 141-146.
HAYES, F. 2003. Transposon-based strategies for microbial functional genomics and proteomics. *Annual review of genetics,* 37**,** 3-29.
HAYNES, B., SAADAT, N., MYUNG, B. & SHEKHAR, M. P. 2015. Crosstalk between translesion synthesis, Fanconi anemia network, and homologous recombination repair pathways in interstrand DNA crosslink repair and development of chemoresistance. *Mutation Research/Reviews in Mutation Research,* 763**,** 258-266.
HEATON, B. E., BARKAN, D., BONGIORNO, P., KARAKOUSIS, P. C. & GLICKMAN, M. S. 2014. Deficiency of double-strand DNA break repair does not impair Mycobacterium tuberculosis virulence in multiple animal models of infection. *Infection and immunity,* 82**,** 3177-3185.
HERNANDO-AMADO, S., BLANCO, P., ALCALDE-RICO, M., CORONA, F., REALES-CALDERÓN, J. A., SÁNCHEZ, M. B. & MARTÍNEZ, J. L. 2016. Multidrug efflux pumps as main players in intrinsic and acquired resistance to antimicrobials. *Drug Resistance Updates,* 28**,** 13-27.
HERNICK, M. 2013a. Mycothiol: a target for potentiation of rifampin and other antibiotics against Mycobacterium tuberculosis. *Expert review of anti-infective therapy,* 11**,** 49-67.
HERNICK, M. 2013b. Targeting Mycothiol Biosynthesis and Mycothiol-Dependent Detoxification for the Treatment of Tuberculosis. *J Anc Dis Prev Rem,* 1**,** e105.
HICKS, N. D., YANG, J., ZHANG, X., ZHAO, B., GRAD, Y. H., LIU, L., OU, X., CHANG, Z., XIA, H. & ZHOU, Y. 2018. Clinically prevalent mutations in Mycobacterium tuberculosis alter propionate metabolism and mediate multidrug tolerance. *Nature microbiology,* 3**,** 1032-1042.
HOOPER, D. C. & JACOBY, G. A. 2016. Topoisomerase inhibitors: fluoroquinolone mechanisms of action and resistance. *Cold Spring Harbor perspectives in medicine,* 6**,** a025320.
HORSBURGH JR, C. R., BARRY III, C. E. & LANGE, C. 2015. Treatment of tuberculosis. *New England Journal of Medicine,* 373**,** 2149-2160.
HOUBEN, D., DEMANGEL, C., VAN INGEN, J., PEREZ, J., BALDEÓN, L., ABDALLAH, A. M., CALEECHURN, L., BOTTAI, D., VAN ZON, M. & DE PUNDER, K. 2012. ESX-1-mediated translocation to the cytosol controls virulence of mycobacteria. *Cellular microbiology,* 14**,** 1287-1298.

HUYGEN, K. 2014. The immunodominant T-cell epitopes of the mycolyl-transferases of the antigen 85 complex of M. tuberculosis. *Frontiers in immunology,* 5**,** 321.

IOERGER, T. R., FENG, Y., GANESULA, K., CHEN, X., DOBOS, K. M., FORTUNE, S., JACOBS, W. R., MIZRAHI, V., PARISH, T. & RUBIN, E. 2010. Variation among genome sequences of H37Rv strains of Mycobacterium tuberculosis from multiple laboratories. *Journal of bacteriology,* 192**,** 3645-3653.

ISOM, C. E., MENON, S. K., THOMAS, L. M., WEST, A. H., RICHTER-ADDO, G. B. & KARR, E. A. 2016. Crystal structure and DNA binding activity of a PadR family transcription regulator from hypervirulent Clostridium difficile R20291. *BMC microbiology,* 16**,** 231.

JAISHANKAR, J. & SRIVASTAVA, P. 2017. Molecular basis of stationary phase survival and applications. *Frontiers in microbiology,* 8**,** 2000.

JANEL-BINTZ, R., NAPOLITANO, R. L., ISOGAWA, A., FUJII, S. & FUCHS, R. P. 2017. Processing closely spaced lesions during Nucleotide Excision Repair triggers mutagenesis in E. coli. *PLoS genetics,* 13**,** e1006881.

JANION, C. 2008. Inducible SOS response system of DNA repair and mutagenesis in Escherichia coli. *International journal of biological sciences,* 4**,** 338.

JENSEN, P. A., ZHU, Z. & VAN OPIJNEN, T. 2017. Antibiotics disrupt coordination between transcriptional and phenotypic stress responses in pathogenic bacteria. *Cell reports,* 20**,** 1705-1716.

JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *science***,** 1225829.

JOLIVET-GOUGEON, A., KOVACS, B., LE GALL-DAVID, S., LE BARS, H., BOUSARGHIN, L., BONNAURE-MALLET, M., LOBEL, B., GUILLÉ, F., SOUSSY, C.-J. & TENKE, P. 2011. Bacterial hypermutation: clinical implications. *Journal of medical microbiology,* 60**,** 563-573.

JOSEPH, A. M., DAW, S., SADHIR, I. & BADRINARAYANAN, A. 2021. Coordination between nucleotide excision repair and specialized polymerase DnaE2 action enables DNA damage survival in non-replicating bacteria. *bioRxiv***,** 2021.02.15.431208.

JOSHI, S. M., PANDEY, A. K., CAPITE, N., FORTUNE, S. M., RUBIN, E. J. & SASSETTI, C. M. 2006. Characterization of mycobacterial virulence genes through genetic interaction mapping. *Proceedings of the National Academy of Sciences,* 103**,** 11760-11765.

KAD, N. M., WANG, H., KENNEDY, G. G., WARSHAW, D. M. & VAN HOUTEN, B. 2010. Collaborative dynamic DNA scanning by nucleotide excision repair proteins investigated by single-molecule imaging of quantum-dot-labeled proteins. *Molecular cell,* 37**,** 702-713.

KALIA, N. P., LEE, B. S., AB RAHMAN, N. B., MORASKI, G. C., MILLER, M. J. & PETHE, K. 2019. Carbon metabolism modulates the efficacy of drugs targeting the cytochrome bc 1: aa 3 in Mycobacterium tuberculosis. *Scientific reports,* 9**,** 1-9.

KALSCHEUER, R. & JACOBS JR, W. R. 2010. The significance of GlgE as a new target for tuberculosis. *Drug News Perspect,* 23**,** 619-24.

KANA, B. D., ABRAHAMS, G. L., SUNG, N., WARNER, D. F., GORDHAN, B. G., MACHOWSKI, E. E., TSENOVA, L., SACCHETTINI, J. C., STOKER, N. G. & KAPLAN, G. 2010. Role of the DinB homologs Rv1537 and Rv3056 in Mycobacterium tuberculosis. *Journal of bacteriology,* 192**,** 2220-2227.

KAPLAN, G. 2020. Tuberculosis control in crisis-causes and solutions. *Progress in Biophysics and Molecular Biology,* 152**,** 6-9.

KHAN, A., SINGH, V. K., HUNTER, R. L. & JAGANNATH, C. 2019. Macrophage heterogeneity and plasticity in tuberculosis. *Journal of leukocyte biology,* 106**,** 275-282.

KIPKORIR, T., MASHABELA, G. T., DE WET, T. J., KOCH, A. S., WIESNER, L., MIZRAHI, V. & WARNER, D. F. 2020. De novo cobalamin biosynthesis, transport and assimilation and cobalamin-mediated regulation of methionine biosynthesis in Mycobacterium smegmatis. *bioRxiv*.

KISKER, C., KUPER, J. & VAN HOUTEN, B. 2013. Prokaryotic nucleotide excision repair. *Cold Spring Harbor perspectives in biology,* 5**,** a012591.
KLECKNER, N., CHAN, R. K., TYE, B.-K. & BOTSTEIN, D. 1975. Mutagenesis by insertion of a drug-resistance element carrying an inverted repetition. *Journal of molecular biology,* 97**,** 561-575.
KLING, A. 2015. Mode of action and resistance mechanism of griselimycin.
KOHANSKI, M. A., DEPRISTO, M. A. & COLLINS, J. J. 2010. Sublethal antibiotic treatment leads to multidrug resistance via radical-induced mutagenesis. *Molecular cell,* 37**,** 311-320.
KOONIN, E. V. 2003. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nature Reviews Microbiology,* 1**,** 127-136.
KOORITS, L., TEGOVA, R., TARK, M., TARASSOVA, K., TOVER, A. & KIVISAAR, M. 2007. Study of involvement of ImuB and DnaE2 in stationary-phase mutagenesis in Pseudomonas putida. *DNA Repair,* 6**,** 863-868.
KORB, V. C., CHUTURGOON, A. A. & MOODLEY, D. 2016. Mycobacterium tuberculosis: manipulator of protective immunity. *International journal of molecular sciences,* 17**,** 131.
KRAEMER, K. H., PATRONAS, N. J., SCHIFFMANN, R., BROOKS, B. P., TAMURA, D. & DIGIOVANNA, J. J. 2007. Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: a complex genotype–phenotype relationship. *Neuroscience,* 145**,** 1388-1396.
KREUZER, K. N. 2013. DNA damage responses in prokaryotes: regulating gene expression, modulating growth patterns, and manipulating replication forks. *Cold Spring Harbor perspectives in biology,* 5**,** a012674.
KUMAR, R., MADHUMATHI, B. S. & NAGARAJA, V. 2014. Molecular basis for the differential quinolone susceptibility of mycobacterial DNA gyrase. *Antimicrobial agents and chemotherapy,* 58**,** 2013-2020.
KUMARI, M., PAL, R. K., MISHRA, A. K., TRIPATHI, S., BISWAL, B. K., SRIVASTAVA, K. K. & ARORA, A. 2018. Structural and functional characterization of the transcriptional regulator Rv3488 of Mycobacterium tuberculosis H37Rv. *Biochemical Journal,* 475**,** 3393-3416.
KURTHKOTI, K. & VARSHNEY, U. 2012. Distinct mechanisms of DNA repair in mycobacteria and their implications in attenuation of the pathogen growth. *Mechanisms of ageing and development,* 133**,** 138-146.
LAMPE, D. J., GRANT, T. E. & ROBERTSON, H. M. 1998. Factors affecting transposition of the Himar1 mariner transposon in vitro. *Genetics,* 149**,** 179-187.
LENAERTS, A., BARRY, C. E. & DARTOIS, V. 2015. Heterogeneity in tuberculosis pathology, microenvironments and therapeutic responses. *Immunological reviews,* 264**,** 288-307.
LEVIN-REISMAN, I., RONIN, I., GEFEN, O., BRANISS, I., SHORESH, N. & BALABAN, N. Q. 2017. Antibiotic tolerance facilitates the evolution of resistance. *Science,* 355**,** 826-830.
LI, G., ZHANG, J., GUO, Q., JIANG, Y., WEI, J., ZHAO, L.-L., ZHAO, X., LU, J. & WAN, K. 2015. Efflux pump gene expression in multidrug-resistant Mycobacterium tuberculosis clinical isolates. *PLoS One,* 10**,** e0119013.
LIN, P. L. & FLYNN, J. L. 2018. The end of the binary era: revisiting the spectrum of tuberculosis. *The Journal of Immunology,* 201**,** 2541-2548.
LIPWORTH, S., HAMMOND, R., BARON, V., HU, Y., COATES, A. & GILLESPIE, S. 2016. Defining dormancy in mycobacterial disease. *Tuberculosis*.
LOBRITZ, M. A., BELENKY, P., PORTER, C. B., GUTIERREZ, A., YANG, J. H., SCHWARZ, E. G., DWYER, D. J., KHALIL, A. S. & COLLINS, J. J. 2015. Antibiotic efficacy is linked to bacterial cellular respiration. *Proceedings of the National Academy of Sciences,* 112**,** 8173-8180.
LONG, J. E., DEJESUS, M., WARD, D., BAKER, R. E., IOERGER, T. & SASSETTI, C. M. 2015. Identifying essential genes in Mycobacterium tuberculosis by global phenotypic profiling. *Gene Essentiality: Methods and Protocols***,** 79-95.

LOPATKIN, A. J., BENING, S. C., MANSON, A. L., STOKES, J. M., KOHANSKI, M. A., BADRAN, A. H., EARL, A. M., CHENEY, N. J., YANG, J. H. & COLLINS, J. J. 2021. Clinically relevant mutations in core metabolic genes confer antibiotic resistance. *Science,* 371.

MADDOCKS, S. E. & OYSTON, P. C. 2008. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology,* 154**,** 3609-3623.

MAITRA, A., MUNSHI, T., HEALY, J., MARTIN, L. T., VOLLMER, W., KEEP, N. H. & BHAKTA, S. 2019. Cell wall peptidoglycan in Mycobacterium tuberculosis: An Achilles' heel for the TB-causing pathogen. *FEMS microbiology reviews,* 43**,** 548-575.

MAJUMDAR, G., MBAU, R., SINGH, V., WARNER, D. F., DRAGSET, M. S. & MUKHERJEE, R. 2017. Genome-Wide Transposon Mutagenesis in Mycobacterium tuberculosis and Mycobacterium smegmatis. *In Vitro Mutagenesis: Methods and Protocols***,** 321-335.

MALIK, M., CHAVDA, K., ZHAO, X., SHAH, N., HUSSAIN, S., KUREPINA, N., KREISWIRTH, B. N., KERNS, R. J. & DRLICA, K. 2012. Induction of mycobacterial resistance to quinolone class antimicrobials. *Antimicrobial agents and chemotherapy,* 56**,** 3879-3887.

MARTIN, C. J., CAREY, A. F. & FORTUNE, S. M. A bug's life in the granuloma. Seminars in immunopathology, 2016. Springer, 213-220.

MASCART, F. & LOCHT, C. 2015. Integrating knowledge of Mycobacterium tuberculosis pathogenesis for the design of better vaccines. *Expert review of vaccines,* 14**,** 1573-1585.

MCGRATH, M., VAN PITTIUS, N. G., VAN HELDEN, P., WARREN, R. & WARNER, D. 2014. Mutation rate and the emergence of drug resistance in Mycobacterium tuberculosis. *Journal of Antimicrobial Chemotherapy,* 69**,** 292-302.

MERRIKH, H. & KOHLI, R. M. 2020. Targeting evolution to inhibit antibiotic resistance. *The FEBS journal,* 287**,** 4341-4353.

MINIAS, A., BRZOSTEK, A. & DZIADEK, J. 2019. Targeting DNA Repair Systems in Antitubercular Drug Development. *Current medicinal chemistry,* 26**,** 1494-1505.

MIZRAHI, V. & ANDERSEN, S. J. 1998. DNA repair in Mycobacterium tuberculosis. What have we learnt from the genome sequence? *Molecular microbiology,* 29**,** 1331-1339.

MOHAN, A., PADIADPU, J., BALONI, P. & CHANDRA, N. 2015. Complete genome sequences of a Mycobacterium smegmatis laboratory strain (MC2 155) and isoniazid-resistant (4XR1/R2) mutant strains. *Genome announcements,* 3**,** e01520-14.

MOLEIRINHO, A., CARNEIRO, J., MATTHIESEN, R., SILVA, R. M., AMORIM, A. & AZEVEDO, L. 2011. Gains, losses and changes of function after gene duplication: study of the metallothionein family. *PloS one,* 6**,** e18487.

MORITA, Y. S., SENA, C. B., WALLER, R. F., KUROKAWA, K., SERNEE, M. F., NAKATANI, F., HAITES, R. E., BILLMAN-JACOBE, H., MCCONVILLE, M. J. & MAEDA, Y. 2006. PimE is a polyprenol-phosphate-mannose-dependent mannosyltransferase that transfers the fifth mannose of phosphatidylinositol mannoside in mycobacteria. *Journal of Biological Chemistry,* 281**,** 25143-25155.

MORRIS, R. P., NGUYEN, L., GATFIELD, J., VISCONTI, K., NGUYEN, K., SCHNAPPINGER, D., EHRT, S., LIU, Y., HEIFETS, L. & PIETERS, J. 2005. Ancestral antibiotic resistance in Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences,* 102**,** 12200-12205.

MOWA, M. B., WARNER, D. F., KAPLAN, G., KANA, B. D. & MIZRAHI, V. 2009. Function and regulation of class I ribonucleotide reductase-encoding genes in mycobacteria. *Journal of bacteriology,* 191**,** 985-995.

MÜLLER, A. U., IMKAMP, F. & WEBER-BAN, E. 2018. The Mycobacterial LexA/RecA-Independent DNA Damage Response Is Controlled by PafBC and the Pup-Proteasome System. *Cell Reports,* 23**,** 3551-3564.

MÜLLER, B., BORRELL, S., ROSE, G. & GAGNEUX, S. 2013. The heterogeneous evolution of multidrug-resistant Mycobacterium tuberculosis. *Trends in Genetics,* 29**,** 160-169.

MURRY, J. P., SASSETTI, C. M., LANE, J. M., XIE, Z. & RUBIN, E. J. 2008. Transposon site hybridization in Mycobacterium tuberculosis. *Microbial Gene Essentiality: Protocols and Bioinformatics.* Springer.

NAMOUCHI, A., GÓMEZ-MUÑOZ, M., FRYE, S. A., MOEN, L. V., ROGNES, T., TØNJUM, T. & BALASINGHAM, S. V. 2016. The Mycobacterium tuberculosis transcriptional landscape under genotoxic stress. *BMC genomics,* 17**,** 791.

NASH, K. A. 2016. Multidrug Resistance in Mycobacteria. *Current Clinical Microbiology Reports,* 3**,** 53-61.

NASIRI, M. J., HAEILI, M., GHAZI, M., GOUDARZI, H., PORMOHAMMAD, A., IMANI FOOLADI, A. A. & FEIZABADI, M. M. 2017. New insights in to the intrinsic and acquired drug resistance mechanisms in mycobacteria. *Frontiers in microbiology,* 8**,** 681.

NGUYEN, L. & THOMPSON, C. J. 2006. Foundations of antibiotic resistance in bacterial physiology: the mycobacterial paradigm. *Trends in microbiology,* 14**,** 304-312.

NIMMO, C., BRIEN, K., MILLARD, J., GRANT, A. D., PADAYATCHI, N., PYM, A. S., O'DONNELL, M., GOLDSTEIN, R., BREUER, J. & BALLOUX, F. 2020. Dynamics of within-host Mycobacterium tuberculosis diversity and heteroresistance during treatment. *EBioMedicine,* 55**,** 102747.

NISHIOKA, H. & DOUDNEY, C. 1969. Different modes of loss of photoreversibility of mutation and lethal damage in ultraviolet-light resistant and sensitive bacteria. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 8**,** 215-228.

NOLL, D. M., MASON, T. M. & MILLER, P. S. 2006. Formation and repair of interstrand cross-links in DNA. *Chemical reviews,* 106**,** 277-301.

OLIVENCIA, B. F., MÜLLER, A. U., ROSCHITZKI, B., BURGER, S., WEBER-BAN, E. & IMKAMP, F. 2017. Mycobacterium smegmatis PafBC is involved in regulation of DNA damage response. *Scientific reports,* 7**,** 13987.

PAGÁN, A. J. & RAMAKRISHNAN, L. 2018. The formation and function of granulomas. *Annual review of immunology,* 36**,** 639-665.

PALOMINO, J.-C., MARTIN, A., CAMACHO, M., GUERRA, H., SWINGS, J. & PORTAELS, F. 2002. Resazurin microtiter assay plate: simple and inexpensive method for detection of drug resistance in Mycobacterium tuberculosis. *Antimicrobial agents and chemotherapy,* 46**,** 2720-2722.

PALOMINO, J. & PORTAELS, F. 1999. Simple procedure for drug susceptibility testing of Mycobacterium tuberculosis using a commercial colorimetic assay. *European Journal of Clinical Microbiology and Infectious Diseases,* 18**,** 380-383.

PARK, K.-H., KIM, E. S., KIM, H. S., PARK, S.-J., BANG, K. M., PARK, H. J., PARK, S.-Y., MOON, S. M., CHONG, Y. P. & KIM, S.-H. 2012. Comparison of the clinical features, bacterial genotypes and outcomes of patients with bacteraemia due to heteroresistant vancomycin-intermediate Staphylococcus aureus and vancomycin-susceptible S. aureus. *Journal of antimicrobial chemotherapy,* 67**,** 1843-1849.

PASIPANODYA, J. G. & GUMBO, T. 2011. A new evolutionary and pharmacokinetic–pharmacodynamic scenario for rapid emergence of resistance to single and multiple anti-tuberculosis drugs. *Current opinion in pharmacology,* 11**,** 457-463.

PECSI, I., HARDS, K., EKANAYAKA, N., BERNEY, M., HARTMAN, T., JACOBS, W. R. & COOK, G. M. 2014. Essentiality of succinate dehydrogenase in Mycobacterium smegmatis and its role in the generation of the membrane potential under hypoxia. *MBio,* 5**,** e01093-14.

PENG, R., CHEN, J.-H., FENG, W.-W., ZHANG, Z., YIN, J., LI, Z.-S. & LI, Y.-Z. 2017. Error-prone DnaE2 Balances the Genome Mutation Rates in Myxococcus xanthus DK1622. *Frontiers in Microbiology,* 8.

PEÑUELAS-URQUIDES, K., VILLARREAL-TREVIÑO, L., SILVA-RAMÍREZ, B., RIVADENEYRA-ESPINOZA, L., SAID-FERNÁNDEZ, S. & LEÓN, M. B. D. 2013. Measuring of Mycobacterium tuberculosis growth: a correlation of the optical measurements with colony forming units. *Brazilian journal of microbiology,* 44**,** 287-290.

PESTOVA, E., MILLICHAP, J. J., NOSKIN, G. A. & PETERSON, L. R. 2000. Intracellular targets of moxifloxacin: a comparison with other fluoroquinolones. *Journal of Antimicrobial Chemotherapy,* 45**,** 583-590.

PETHE, K., SEQUEIRA, P. C., AGARWALLA, S., RHEE, K., KUHEN, K., PHONG, W. Y., PATEL, V., BEER, D., WALKER, J. R. & DURAISWAMY, J. 2010. A chemical genetic screen in Mycobacterium tuberculosis identifies carbon-source-dependent growth inhibitors devoid of in vivo efficacy. *Nature communications,* 1**,** 1-8.

PHILLIPS, P. C. 1998. The language of gene interaction. *Genetics,* 149**,** 1167-1171.

PRAMMANANAN, T., PHUNPRUCH, S., JAITRONG, S. & PALITTAPONGARNPIM, P. 2012. Mycobacterium tuberculosis uvrC essentiality in response to uv-induced cell damage. *Southeast Asian Journal of Tropical Medicine & Public Health,* 43**,** 370-375.

PRIDEAUX, B., VIA, L. E., ZIMMERMAN, M. D., EUM, S., SARATHY, J., O'BRIEN, P., CHEN, C., KAYA, F., WEINER, D. M. & CHEN, P.-Y. 2015. The association between sterilizing activity and drug distribution into tuberculosis lesions. *Nature medicine,* 21**,** 1223-1227.

PULE, C. M., SAMPSON, S. L., WARREN, R. M., BLACK, P. A., VAN HELDEN, P. D., VICTOR, T. C. & LOUW, G. E. 2015. Efflux pump inhibitors: targeting mycobacterial efflux systems to enhance TB therapy. *Journal of Antimicrobial Chemotherapy,* 71**,** 17-26.

QI, L. S., LARSON, M. H., GILBERT, L. A., DOUDNA, J. A., WEISSMAN, J. S., ARKIN, A. P. & LIM, W. A. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell,* 152**,** 1173-1183.

QUENARD, F., FOURNIER, P. E., DRANCOURT, M. & BROUQUI, P. 2017. Role of second-line injectable antituberculosis drugs in the treatment of MDR/XDR tuberculosis. *International journal of antimicrobial agents,* 50**,** 252-254.

RACHMAN, H., STRONG, M., ULRICHS, T., GRODE, L., SCHUCHHARDT, J., MOLLENKOPF, H., KOSMIADI, G. A., EISENBERG, D. & KAUFMANN, S. H. 2006. Unique transcriptome signature of Mycobacterium tuberculosis in pulmonary tuberculosis. *Infection and immunity,* 74**,** 1233-1242.

RAD, M. E., BIFANI, P., MARTIN, C., KREMER, K., SAMPER, S., RAUZIER, J., KREISWIRTH, B., BLAZQUEZ, J., JOUAN, M. & VAN SOOLINGEN, D. 2003. Mutations in putative mutator genes of Mycobacterium tuberculosis strains of the W-Beijing family. *Emerging infectious diseases,* 9**,** 838.

RAMAKRISHNAN, L. 2012. Revisiting the role of the granuloma in tuberculosis. *Nature Reviews Immunology,* 12**,** 352-366.

RAND, L., HINDS, J., SPRINGER, B., SANDER, P., BUXTON, R. S. & DAVIS, E. O. 2003. The majority of inducible DNA repair genes in Mycobacterium tuberculosis are induced independently of RecA. *Molecular microbiology,* 50**,** 1031-1042.

REICHE, M. A., WARNER, D. F. & MIZRAHI, V. 2017. Targeting DNA replication and repair for the development of novel therapeutics against tuberculosis. *Frontiers in molecular biosciences,* 4**,** 75.

REN, H. & LIU, J. 2006. AsnB is involved in natural resistance of Mycobacterium smegmatis to multiple drugs. *Antimicrobial agents and chemotherapy,* 50**,** 250-255.

REN, X., ZOU, L. & HOLMGREN, A. 2020. Targeting bacterial antioxidant systems for antibiotics development. *Current Medicinal Chemistry,* 27**,** 1922-1939.

RENGARAJAN, J., BLOOM, B. R. & RUBIN, E. J. 2005. Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 8327-8332.

RHEE, K. Y., DE CARVALHO, L. P. S., BRYK, R., EHRT, S., MARRERO, J., PARK, S. W., SCHNAPPINGER, D., VENUGOPAL, A. & NATHAN, C. 2011. Central carbon metabolism in Mycobacterium tuberculosis: an unexpected frontier. *Trends in microbiology,* 19**,** 307-314.

ROBASKY, K., LEWIS, N. E. & CHURCH, G. M. 2014. The role of replicates for error mitigation in next-generation sequencing. *Nature Reviews Genetics,* 15**,** 56.

ROCK, J. M., HOPKINS, F. F., CHAVEZ, A., DIALLO, M., CHASE, M. R., GERRICK, E. R., PRITCHARD, J. R., CHURCH, G. M., RUBIN, E. J. & SASSETTI, C. M. 2017. Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. *Nature microbiology,* 2**,** 16274.

ROCK, J. M., LANG, U. F., CHASE, M. R., FORD, C. B., GERRICK, E. R., GAWANDE, R., COSCOLLA, M., GAGNEUX, S., FORTUNE, S. M. & LAMERS, M. H. 2015. DNA replication fidelity in Mycobacterium tuberculosis is mediated by an ancestral prokaryotic proofreader. *Nature genetics,* 47**,** 677-681.

ROY, U. & SCHÄRER, O. D. 2016. Involvement of translesion synthesis DNA polymerases in DNA interstrand crosslink repair. *DNA repair,* 44**,** 33-41.

RUSTAD, T. R., HARRELL, M. I., LIAO, R. & SHERMAN, D. R. 2008. The enduring hypoxic response of Mycobacterium tuberculosis. *PloS one,* 3**,** e1502.

SAMBOU, T., DINADAYALA, P., STADTHAGEN, G., BARILONE, N., BORDAT, Y., CONSTANT, P., LEVILLAIN, F., NEYROLLES, O., GICQUEL, B. & LEMASSU, A. 2008. Capsular glucan and intracellular glycogen of Mycobacterium tuberculosis: biosynthesis and impact on the persistence in mice. *Molecular microbiology,* 70**,** 762-774.

SASSETTI, C. M., BOYD, D. H. & RUBIN, E. J. 2003. Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular microbiology,* 48**,** 77-84.

SCHLOSSER-SILVERMAN, E., ELGRABLY-WEISS, M., ROSENSHINE, I., KOHEN, R. & ALTUVIA, S. 2000. Characterization of Escherichia coli DNA lesions generated within J774 macrophages. *Journal of bacteriology,* 182**,** 5225-5230.

SCHOONMAKER, M. K., BISHAI, W. R. & LAMICHHANE, G. 2014. Nonclassical transpeptidases of Mycobacterium tuberculosis alter cell size, morphology, the cytosolic matrix, protein localization, virulence, and resistance to β-lactams. *Journal of bacteriology,* 196**,** 1394-1402.

SCOTLAND, M. K., HELTZEL, J. M., KATH, J. E., CHOI, J.-S., BERDIS, A. J., LOPARO, J. J. & SUTTON, M. D. 2015. A genetic selection for dinB mutants reveals an interaction between DNA polymerase IV and the replicative polymerase that is required for translesion synthesis. *PLoS genetics,* 11**,** e1005507.

SGARAGLI, G. & FROSINI, M. 2016. Human tuberculosis I. Epidemiology, diagnosis and pathogenetic mechanisms. *Current medicinal chemistry,* 23**,** 2836-2873.

SHILOH, M. U. & CHAMPION, P. A. D. 2010. To catch a killer. What can mycobacterial models teach us about Mycobacterium tuberculosis pathogenesis? *Current opinion in microbiology,* 13**,** 86-92.

SINGH, A. 2017. Guardians of the mycobacterial genome: A review on DNA repair systems in Mycobacterium tuberculosis. *Microbiology,* 163**,** 1740-1758.

SINGH, R., DWIVEDI, S. P., GAHARWAR, U. S., MEENA, R., RAJAMANI, P. & PRASAD, T. 2020. Recent updates on drug resistance in Mycobacterium tuberculosis. *Journal of applied microbiology,* 128**,** 1547-1567.

SMOLLETT, K. L., SMITH, K. M., KAHRAMANOGLOU, C., ARNVIG, K. B., BUXTON, R. S. & DAVIS, E. O. 2012. Global analysis of the regulon of the transcriptional repressor LexA, a key component of SOS response in Mycobacterium tuberculosis. *Journal of Biological Chemistry,* 287**,** 22004-22014.

SNAPPER, S., MELTON, R., MUSTAFA, S., KIESER, T. & WR JR, J. 1990. Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis. *Molecular microbiology,* 4**,** 1911-1919.

SPIVAK, G. 2016. Transcription-coupled repair: an update. *Archives of toxicology,* 90**,** 2583-2594.

STEVENS, P. 2017. Diseases of poverty and the 10/90 gap. *Fighting the diseases of poverty.* Routledge.

STURM, M., SCHROEDER, C. & BAUER, P. 2016. SeqPurge: highly-sensitive adapter trimming for paired-end NGS data. *BMC Bioinformatics,* 17**,** 208.

SU, W., FENG, J., CHIU, Y., HUANG, S. & LEE, Y. 2011. Role of 2-month sputum smears in predicting culture conversion in pulmonary tuberculosis. *European Respiratory Journal,* 37**,** 376-383.
TARK, M., TOVER, A., KOORITS, L., TEGOVA, R. & KIVISAAR, M. 2008. Dual role of NER in mutagenesis in Pseudomonas putida. *DNA Repair,* 7**,** 20-30.
TIMINSKAS, K., BALVOČIŪTĖ, M., TIMINSKAS, A. & VENCLOVAS, Č. 2014. Comprehensive analysis of DNA polymerase III α subunits and their homologs in bacterial genomes. *Nucleic acids research,* 42**,** 1393-1413.
TIPPIN, B., PHAM, P. & GOODMAN, M. F. 2004. Error-prone replication for better or worse. *Trends in microbiology,* 12**,** 288-295.
TOMASZ, M. & PALOM, Y. 1997. The mitomycin bioreductive antitumor agents: cross-linking and alkylation of DNA as the molecular basis of their activity. *Pharmacology & therapeutics,* 76**,** 73-87.
TRAN, S. L. & COOK, G. M. 2005. The F1Fo-ATP synthase of Mycobacterium smegmatis is essential for growth. *Journal of bacteriology,* 187**,** 5023-5028.
TRUGLIO, J. J., CROTEAU, D. L., VAN HOUTEN, B. & KISKER, C. 2006. Prokaryotic nucleotide excision repair: the UvrABC system. *Chemical reviews,* 106**,** 233-252.
VAN DEN BERGH, B., MICHIELS, J. E., WENSELEERS, T., WINDELS, E. M., BOER, P. V., KESTEMONT, D., DE MEESTER, L., VERSTREPEN, K. J., VERSTRAETEN, N. & FAUVART, M. 2016. Frequency of antibiotic application drives rapid evolutionary adaptation of Escherichia coli persistence. *Nature microbiology,* 1**,** 1-7.
VAN HAL, S. J., JONES, M., GOSBELL, I. B. & PATERSON, D. L. 2011. Vancomycin heteroresistance is associated with reduced mortality in ST239 methicillin-resistant Staphylococcus aureus blood stream infections. *PloS one,* 6**,** e21217.
VAN LOON, B., MARKKANEN, E. & HÜBSCHER, U. 2010. Oxygen as a friend and enemy: How to combat the mutational potential of 8-oxo-guanine. *DNA repair,* 9**,** 604-616.
VAN OPIJNEN, T., BODI, K. L. & CAMILLI, A. 2009. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature methods,* 6**,** 767-772.
VAN OPIJNEN, T. & CAMILLI, A. 2013. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nature Reviews Microbiology,* 11**,** 435-442.
VELAYATI, A. A., MASJEDI, M. R., FARNIA, P., TABARSI, P., GHANAVI, J., ZIAZARIFI, A. H. & HOFFNER, S. E. 2009. Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in Iran. *Chest Journal,* 136**,** 420-425.
VILCHÈZE, C. & JACOBS JR, W. R. 2019. The Isoniazid Paradigm of Killing, Resistance, and Persistence in Mycobacterium tuberculosis. *Journal of molecular biology*.
VIVEIROS, M., MARTINS, M., RODRIGUES, L., MACHADO, D., COUTO, I., AINSA, J. & AMARAL, L. 2012. Inhibitors of mycobacterial efflux pumps as potential boosters for anti-tubercular drugs. *Expert review of anti-infective therapy,* 10**,** 983-998.
VOLMINK, J. & GARNER, P. 2007. Directly observed therapy for treating tuberculosis. *Cochrane Database of systematic reviews*.
WANG, Y., HUANG, Y., XUE, C., HE, Y. & HE, Z.-G. 2011. ClpR protein-like regulator specifically recognizes RecA protein-independent promoter motif and broadly regulates expression of DNA damage-inducible genes in mycobacteria. *Journal of Biological Chemistry,* 286**,** 31159-31167.
WARNER, D. F. 2010. The role of DNA repair in M. tuberculosis pathogenesis. *Drug Discovery Today: Disease Mechanisms,* 7**,** e5-e11.
WARNER, D. F. 2015. Mycobacterium tuberculosis metabolism. *Cold Spring Harbor perspectives in medicine,* 5**,** a021121.
WARNER, D. F., ETIENNE, G., WANG, X.-M., MATSOSO, L. G., DAWES, S. S., SOETAERT, K., STOKER, N. G., CONTENT, J. & MIZRAHI, V. 2006. A derivative of Mycobacterium smegmatis mc2155 that lacks the duplicated chromosomal region. *Tuberculosis,* 86**,** 438-444.

WARNER, D. F. & MIZRAHI, V. 2006. Tuberculosis chemotherapy: the influence of bacillary stress and damage response pathways on drug efficacy. *Clinical microbiology reviews,* 19**,** 558-570.
WARNER, D. F., NDWANDWE, D. E., ABRAHAMS, G. L., KANA, B. D., MACHOWSKI, E. E., VENCLOVAS, Č. & MIZRAHI, V. 2010. Essential roles for imuA'-and imuB-encoded accessory factors in DnaE2-dependent mutagenesis in Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences,* 107**,** 13093-13098.
WARSINSKE, H. C., DIFAZIO, R. M., LINDERMAN, J. J., FLYNN, J. L. & KIRSCHNER, D. E. 2017. Identifying mechanisms driving formation of granuloma-associated fibrosis during Mycobacterium tuberculosis infection. *Journal of theoretical biology,* 429**,** 1-17.
WERNGREN, J. & HOFFNER, S. E. 2003. Drug-susceptible Mycobacterium tuberculosis Beijing genotype does not develop mutation-conferred resistance to rifampin at an elevated rate. *Journal of clinical microbiology,* 41**,** 1520-1524.
WHO 2017. Guidelines for treatment of drug-susceptible tuberculosis and patient care.
WHO 2018. Global tuberculosis report 2017. 2017. *Google Scholar*.
WHO 2019. Global tuberculosis report 2018. 2018. *Geneva: World Health Organization*.
WHO 2020. WHO consolidated guidelines on tuberculosis: module 4: treatment: drug-resistant tuberculosis treatment: online annexes.
WRIGHT, G. D. 2005. Bacterial resistance to antibiotics: enzymatic degradation and modification. *Advanced drug delivery reviews,* 57**,** 1451-1470.
WU, H., FANG, Y., YU, J. & ZHANG, Z. 2014. The quest for a unified view of bacterial land colonization. *The Isme Journal,* 8**,** 1358.
XU, W., DEJESUS, M. A., RÜCKER, N., ENGELHART, C. A., WRIGHT, M. G., HEALY, C., LIN, K., WANG, R., PARK, S. W. & IOERGER, T. R. 2017. Chemical genetic interaction profiling reveals determinants of intrinsic antibiotic resistance in Mycobacterium tuberculosis. *Antimicrobial agents and chemotherapy,* 61**,** e01334-17.
YAMCHI, J. K., HAEILI, M., FEYISA, S. G., KAZEMIAN, H., SHAHRAKI, A. H., ZAHEDNAMAZI, F., FOOLADI, A. A. I. & FEIZABADI, M. M. 2015. Evaluation of efflux pump gene expression among drug susceptible and drug resistant strains of Mycobacterium tuberculosis from Iran. *Infection, Genetics and Evolution,* 36**,** 23-26.
YOUNG, D. B., COMAS, I. & DE CARVALHO, L. P. 2015. Phylogenetic analysis of vitamin B12-related metabolism in Mycobacterium tuberculosis. *Frontiers in molecular biosciences,* 2**,** 6.
ZAKHAM, F., SIRONEN, T., VAPALAHTI, O. & KANT, R. 2021. Pan and Core Genome Analysis of 183 Mycobacterium tuberculosis Strains Revealed a High Inter-Species Diversity among the Human Adapted Strains. *Antibiotics,* 10**,** 500.
ZHANG, J. & WALTER, J. C. 2014. Mechanism and regulation of incisions during DNA interstrand cross-link repair. *DNA repair,* 19**,** 135-142.
ZHANG, Y., ZBORNÍKOVÁ, E., REJMAN, D. & GERDES, K. 2018. Novel (p) ppGpp binding and metabolizing proteins of Escherichia coli. *MBio,* 9**,** e02188-17.