REPLICATION FIDELITY IN THE MICROEVOLUTION OF MYCOBACTERIA

Zanele Ditse

DTSZAN001

Division of Medical Microbiology,

Department of Clinical and Laboratory Sciences.

A dissertation submitted to the Faculty of Health Sciences,

for the Degree of DOCTOR OF PHILOSOPHY,

University of Cape Town

February, 2015.

The Road Not Taken

Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;

Then took the other, as just as fair,
And having perhaps the better claim
Because it was grassy and wanted wear,
Though as for that the passing there
Had worn them really about the same

And both that morning equally lay
In leaves no step had trodden black.
Oh, I marked the first for another day!
Yet knowing how way leads on to way
I doubted if I should ever come back.

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I,
I took the one less traveled by,
And that has made all the difference.

By **Robert Frost**

DECLARATION

I declare that this thesis is my own unaided work, both in concept and execution, and that apart from the normal guidance from my supervisors, I have received no assistance. It is being submitted for the degree of Doctor of Philosophy at the University of Cape Town, Cape Town. It has not been submitted for any degree or examination at any other university.

Signed by candidate

Signature Removed

_____

Zanele Ditse **Date:** 19 February 2015

ABSTRACT

This thesis aimed to elucidate the structure-function relationships determining the differential fidelities of the *dnaE1*- and *dnaE2*-encoded mycobacterial PolIIIα subunits under conditions of genotoxic stress. To this end, the role in DnaE1 intrinsic fidelity of highly conserved PHP domain residues was explored by site-directed replacement of targeted amino acids, resulting in a panel of *Mycobacterium smegmatis* mutants carrying selected *dnaE1* alleles. A complementary approach investigated the contribution of the mycobacterial proofreading DnaQ subunit homolog to the maintenance of DnaE1-dependent replicative fidelity by generating a targeted *dnaQ* knockout mutant. The third component of this study focused on the inferred role of a highly-conserved N-terminal extension and C-terminal pentapeptide motif in the function of the alternative, error-prone DNA PolIIIα subunit, DnaE2.

Replacement of wild-type *dnaE1* with mutant *dnaE1*^{E133A} and *dnaE1*^{D228N} alleles resulted in reproducible increases in the spontaneous mutation rate of 3-fold and 10-fold, respectively. This result confirmed the predicted role of highly conserved PHP-domain residues in DnaE1 intrinsic fidelity and, significantly, provided the first evidence of mutator alleles in mycobacteria. A reproducible, but not significant, loss in fidelity (~1.4-fold increase in mutation rate) was observed for a *dnaQ* knock-out mutant compared with the wild-type strain; moreover, DnaQ was shown to be dispensable for DNA damage-induced mutagenesis and damage tolerance. In contrast, a second *dnaQ* homolog, *dnaQ-uvrC*, which comprises an N-terminal 3'-5' exonuclease domain and a C-terminal UvrC-like endonuclease domain, was shown to be required for DNA damage survival, suggesting a role in SOS-mediated DNA repair. Finally, targeted deletion of conserved N- and C-terminal regions in DnaE2 had no effect on DNA damage tolerance or induced mutagenesis, indicating that these domains are not crucial for the function of the error-prone polymerase under conditions of genotoxic stress. In summary, these results reinforced the notion that the mycobacterial replisome differs in key respects from the well-characterised *E. coli* model, and so urge further work to elucidate the composition and regulation of the protein complex which governs DNA replication and repair in a major human pathogen.

ACKNOWLEDGEMENTS

I would like to thank the **Almighty Lord** for blessing me every day and taking me this far.

I am grateful to, the **Carnegie Corporation of New York**, the German Academic Exchange Service (**DAAD**) in partnership with the National Research Foundation (**NRF**), the Department of Science of Technology (**DST**), the Medical Research Council of South Africa (**MRC**), the University of Cape Town (**UCT**) and other funders of the Molecular Mycobacteriology Research Unit (**MMRU**) for financial support, and for contributing to local and international travel expenses.

To my supervisor, **Associate Professor Digby Warner**, thank you for your guidance, advice and academic support throughout my PhD training. When I decided to switch fields, I felt like I was in a foreign country, but through you support and making everything look so easy, my journey was challenging but very exciting and rewarding. You have instilled in me an appreciation of scientific research.

To my co-supervisor, **Professor Valerie Mizrahi**, thank you for your invaluable scientific advice. It has been a privilege working in your lab.

I am grateful to **Dr Christopher Ford** and the **Fortune lab** (Harvard University) for the kind donation of the *dnaQ* mutants.

I would like to thank **Professor Česlovas Venclovas** (Vilnius University) for advice and helpful discussions regarding structural information/modelling of DnaE1 and DnaE2 proteins.

To my colleagues, thank you for your support, friendship and technical advice when I was experiencing challenges in the lab.

To my husband, **Bafana Vumasi**, thank you for your patience, love, sacrifice and support throughout my studies. To my siblings, **Dudu, Palesa** and **Mpho**, thank you for your love,

support and encouragement when everything seemed impossible. And finally, to my parents, **Mr Andrew Ditse** and **Mrs Tiny Ditse**, who have always been my rock and pillar of strength. Thank you for believing in me and teaching me that the sky is not the limit. Without all of you, I would have not gone this far.

PUBLICATIONS FROM THIS THESIS

Ditse Z., Venclovas C., Mizrahi V., and Warner D.F. Elucidating the molecular features contributing to the differential fidelities of the *dnaE*-encoded α subunits in mycobacteria (in preparation)

Ditse Z., Martin Z., Mizrahi V., and Warner D.F. Interplay between DNA replication and repair pathways in *Mycobacterium tuberculosis* (in preparation)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: Literature Review

1.1 Introduction

Mycobacterium tuberculosis (Mtb), the causative agent of tuberculosis (TB), is exposed to multiple host-derived reactive oxygen and nitrogen species which are genotoxic and pose a major threat to the integrity of the genome (Darwin and Nathan, 2005, Dutta *et al.*, 2010, Sassetti and Rubin, 2003, Gorna *et al.*, 2010). Moreover, Mtb has the ability to persist for decades in a poorly understood subclinical state, in some cases reactivating decades later to cause post-primary TB (Lillebaek *et al.*, 2002). This suggests that active DNA repair is critical to maintain genome integrity and bacterial viability during long-term infection. Furthermore, unlike many other bacterial pathogens, drug resistance in Mtb arises exclusively from mutations in chromosomal genes that are associated with drug action (Almeida Da Silva and Palomino, 2011, Sandgren *et al.*, 2009). This implies that chromosomal mutagenesis drives the microevolution of this pathogen within the human host and, in turn, suggests that a balance between DNA repair and damage tolerance pathways is critical for genome maintenance.

Although crucial for bacillary survival and strain evolution, the mechanisms governing mycobacterial genome maintenance and DNA damage tolerance remain poorly understood. The fact that there is no evidence of a single polymorphism in an essential component of the DNA replication machinery that significantly impacts replication fidelity (that is, there are no known mycobacterial "mutator" alleles) further suggests that multiple functional and regulatory interactions interact to enable DNA replication and genome maintenance in Mtb (Werngren and Hoffner, 2003). For example, mycobacteria have developed several mechanisms to enable adaptation to hostile and fluctuating host environments; these include the regulation of deoxyribonucleoside-5'-triphosphates (dNTP) (Mathews, 2014) and ribonucleoside triphosphates (rNTP) levels (Clausen *et al.*, 2013, Makarova *et al.*, 2014, Nick McElhinny *et al.*, 2010) during DNA replication, as well as co-ordination of chromosomal replication with mycobacterial cell segregation and division (Aldridge *et al.*, 2012, Kieser and Rubin, 2014, Joyce *et al.*, 2012, Santi *et al.*, 2013). In support of this notion, the unique mode of mycobacterial cell elongation and division has been implicated in the phenotypic

heterogeneity observed in mycobacterial populations and, by implication, the emergence of multidrug-resistant (MDR) and extensively drug-resistant (XDR) strains (Aldridge *et al.*, 2012, Joyce *et al.*, 2012, Kieser and Rubin, 2014, Santi *et al.*, 2013, Singh *et al.*, 2013). However, further research is required to determine accurate estimates of intracellular dNTP concentrations in the different phases of the mycobacterial cell cycle, as well to determine the speed and processivity of the mycobacterial replicase. Also, the relationship between dNTP pools and the replication rate of *M. tuberculosis* at different stages of infection remains to be established. Among numerous factors, an improved understanding of the molecular mechanisms by which M/XDR strains emerge and propagate (Warner and Mizrahi, 2006) could aid the design of control programmes aimed at preserving the efficacy of new and existing drugs.

1.2 Drug resistance in Mtb

TB remains a global health problem. In 2013, the World Health Organisation (WHO) reported 9 million new cases and 1.5 million deaths due to this pathogen (WHO, 2014). MDR TB – defined as resistance to two of the first-line drugs, rifampicin (RIF) and isoniazid (INH) – was identified in 3.5% of new cases and 20.5% of previously treated TB cases (WHO, 2014). On average, 9.0% of patients with MDR-TB had XDR-TB, which is defined as MDR-TB plus resistance to a fluoroquinolone and any second-line injectable agent. So-called “totally drug-resistant (TDR)” cases have also been reported, however this form of resistance is still to be defined by WHO (WHO, 2012).

Infection with Mtb follows inhalation of aerosolized bacteria, and leads to either active disease or, in most individuals, latent infection (Cosma *et al.*, 2004, Barry *et al.*, 2009). Patients who are actively infected with TB present with a large bacterial burden at diagnosis and are treated with multiple antibiotics to prevent the emergence of drug resistance mutations. In contrast, latently infected individuals, when diagnosed or as part of INH preventative therapy (IPT) (Balcells *et al.*, 2006, Lincoln, 1954), are treated with a single antibiotic since latent infection is thought to have little capacity for mutation (Lincoln, 1954, Balcells *et al.*, 2006).

The emergence of strains that are resistant to first- and second-line treatment has made the treatment of TB complex. To compensate for the loss of the two most potent drugs, RIF and INH, the treatment regimen for MDR- and XDR-TB requires prolonged administration of second- and third-line agents that are expensive, difficult to administer, and are often associated with severe toxicities and increased side effects (WHO, 2010). Additionally, in comparison with the 6 months required to treat drug-susceptible TB, drug-resistant TB generally requires prolonged treatment of at least 18 months. These factors contribute to patient non-adherence or premature discontinuation of therapy, leading to treatment failure and the emergence of strains with additional drug resistance.

The common mechanisms by which most bacteria achieve antimicrobial resistance are: (i) barrier mechanisms (decreased permeability or efflux), (ii) degrading or inactivating enzymes, (iii) modification of pathways involved in drug activation or metabolism, and (iv) drug target modification or target amplification. Mtb uses all of these mechanisms to achieve antimicrobial resistance. Efflux mechanisms only allow the bacteria to tolerate higher levels of drugs but do not necessarily result in clinically relevant levels of resistance to multiple antibiotics (Schmalstieg *et al.*, 2012, Machado *et al.*, 2012, Li and Nikaido, 2009, Borrell and Gagneux, 2011). In most bacteria, genetically-encoded drug resistance can be mediated by plasmids, horizontal gene transfer and chromosomal mutations (Gillespie, 2001). However, in Mtb, there is little evidence of horizontal gene transfer and no epigenetic information in the form of plasmids; instead, genetic diversity is principally driven by chromosomal mutations and rearrangements (Hershberg *et al.*, 2008). All drug resistance in Mtb is associated with mutations in target or related genes, the majority of which are single nucleotide polymorphisms (SNPs) (Sandgren *et al.*, 2009) – suggesting that strain variation and the emergence of antibiotic resistance in Mtb are dependent on mutagenesis. For a pathogen probably exposed to multiple genotoxic stresses during infection (Darwin and Nathan, 2005, Dutta *et al.*, 2010, Sassetti and Rubin, 2003), this observation suggests that DNA repair pathways might be crucial to both bacillary survival and the adaptive evolution of Mtb within its obligate human host (Gorna *et al.*, 2010, Warner, 2010). Furthermore, by implication, it also highlights the importance of developing compounds to

inhibit repair pathways in Mtb, especially those associated with induced mutagenesis (Smith and Romesberg, 2007).

Recent studies have provided insight into individual genes in which mutations confer resistance in Mtb (Farhat *et al.*, 2013, Safi *et al.*, 2013, Sandgren *et al.*, 2009, Zhang *et al.*, 2013). However, some resistance cannot be accounted for by the current models, suggesting that sequential acquisition of multiple small-to-moderate effect drug-resistance and compensatory mutations can reduce the innate susceptibility of the organism to antibiotics while maintaining strain fitness (Warner and Mizrahi, 2013). This notion is supported by the work of Safi and colleagues who applied a combination of *in vitro* selection and molecular genetic techniques to demonstrate that the acquisition of high-level ethambutol resistance occurs as a multi-step process, contradicting the notion that drug resistance in Mtb arises exclusively through single-step mutations (Safi *et al.*, 2013). Moreover, recent studies involving drug-resistant Mtb strains used whole-genome sequencing (WGS) to identify genetic loci that are associated with resistance, including genes involved in the synthesis or regulation of surface exposed lipids (Farhat *et al.*, 2013, Zhang *et al.*, 2013). Both studies detected resistance-associated mutations that were also present in drug-sensitive strains, suggesting that these might represent early mutation events in the sequential acquisition of resistance. Most importantly, they also reinforced the notion that XDR-TB arises through the accumulation of non-synonymous mutations that are associated with resistance to second-line drugs rather than mutations in a small number of genes conferring resistance. These observations, combined with those of Safi *et al.*, imply that, while prevalent, low-level resistance might be undetected by current diagnostic methods. This could have serious clinical implications, especially in TB endemic regions, and so reinforces the idea that replacing current clinical microbiology methods for routine diagnosis and drug-susceptibility testing with high-throughput WGS approaches might be of significant benefit (Koser *et al.*, 2013).

As sequencing of Mtb strains becomes more common and affordable, it is hoped that the expansion of genotype-phenotype databases will enable the identification and definition of additional drug-resistance signatures with improved potential to inform new tools for diagnosing drug-resistant TB. Understanding the factors which drive the emergence and

transmission of Mtb strains that are resistant to one or more frontline anti-tubercular drugs will, however, require elucidation of molecular mechanisms governing mycobacterial DNA replication fidelity and DNA damage tolerance at every stage of the infection process.

1.3 DNA replication in prokaryotes and genomic distribution of C-family polymerases

DNA polymerase III (PolIII) holoenzyme (HE) is a tripartite protein machine responsible for replication of the bacterial genome (Kornberg and Baker, 2005, Kurth and O'Donnell, 2013, McHenry, 2011b, McHenry, 2011a, Robinson *et al.*, 2012). It consists of the PolIII core, the β_2 processivity factor, and a clamp loader complex. The PolIII core is further divided into three subunits: a) *dnaE*-encoded α subunit, b) *dnaQ*-encoded ε subunit possessing 3'-5' exonuclease proofreading activity, and c) *holE*-encoded θ subunit, whose function is to stabilize the ε subunit. DNA synthesis is performed by the α subunit (PolIIIα). PolIIIα subunits are classified into the C-family of DNA polymerases (Ito and Braithwaite, 1991), and come in two major forms: DnaE-type (Bailey *et al.*, 2006, Lamers *et al.*, 2006) and PolC-type (Evans *et al.*, 2008). PolC is present in low-GC Gram-positive bacteria such as *Bacillus subtilis*, whereas DnaE is PolIIIα of the widely studied Gram-negative model organism, *E. coli*.

In a major recent study, Venclovas and colleagues analysed the distribution of C-family polymerases among ~2000 sequenced bacterial genomes, demonstrating that the majority of bacterial genomes sequenced to date contain two, three or even four putative C-family polymerases (Timinskas *et al.*, 2014). According to their results, most genomes encode a single replicative polymerase of the DnaE1 type – which is the sole high-fidelity replicative C-family DNA polymerase in the cell (Zhao *et al.*, 2006, Timinskas *et al.*, 2014). DnaE1 was found to be the only polymerase that could exist either alone or in combination with DnaE2 and PolC (Zhao *et al.*, 2006), whereas the other members always occur in combination with representatives from at least one of the other different groups. PolC was found to co-occur with either DnaE1 or the third DnaE-type polymerase, DnaE3, with DnaE3 found in *Bacilli* and DnaE1 in *Clostridia* and *Negativicutes*. DnaE3 always co-occurs with PolC, whereas the second DnaE-type polymerase, DnaE2, does not conform to phylogenetic boundaries and

can co-exist with DnaE1 or PolC (Timinskas *et al.*, 2014). Interestingly, the analysis also suggested that non-essential DnaE2 polymerases are common among oxygen-using bacteria with large GC-rich genomes. Consistent with this observation, Mtb contains two DnaE-type PolIIIα subunits, DnaE1 and DnaE2 instead of a PolC-type polymerase (Boshoff *et al.*, 2003, Timinskas *et al.*, 2014).

1.4 The composition and the arrangement of structural/functional domains in distinct C-family polymerases

Members of the C-family of DNA polymerases contain a set of four highly conserved and ordered domains within a single polypeptide in the following sequence: Polymerase and Histidinol Phosphatase (PHP) domain, Palm domain, Thumb domain, and Fingers domain. DnaE and PolC subfamilies can be distinguished within the C-family of DNA polymerases depending on the arrangement of additional accessory domains such as the oligonucleotide/oligosaccharide binding (OB) domain that binds single-stranded DNA (Huang and Ito, 1999, Lamers and O'Donnell, 2008, McHenry, 2011b). DnaE and PolC share some degree of sequence homology with other DNA polymerases, in both bacterial and eukaryotic replicative DNA polymerases. Interestingly, the crystal structures of *E. coli* PolIII (Lamers *et al.*, 2006), *T. aquaticus* PolIII (Bailey *et al.*, 2006) and *G. kaustophilus* PolC (Evans *et al.*, 2008) have shown that the active sites of these polymerases are structurally related to the X-family of DNA polymerases, which are typically slow and exhibit low fidelity and processivity. This is in contrast to the high-fidelity replicative C-family polymerases which are amongst the fastest polymerases known. In addition, DnaE and PolC polymerases are unique in that they contain a PHP domain that is not found in other polymerases, except for some bacterial PolX family members (Aravind and Koonin, 1998).

The role of the PHP domain in DnaE and PolC polymerases remains unclear. The PHP domains of *T. thermophilus* PolIII and *T. aquaticus* PolIII are functional exonucleases possessing a complete set of metal co-ordinating residues [H H D/H H E H C/H D/N H] that appear to be essential for this activity (Stano *et al.*, 2006, Wing *et al.*, 2008). Similarly, the PHP domains of PolX from both *B. subtilis* (Banos *et al.*, 2008) and *T. thermophilus* (Nakane

et al., 2009) have exonuclease activity. In contrast, no exonuclease activity could be detected for the PHP domain of *G. kaustophilus* PolC (Evans *et al.*, 2008) which contains an almost intact active site with the exception of an aspartate to asparagine substitution at position 8 (Barros *et al.*, 2013). Mtb DnaE1 contains a complete set of the metal co-ordinating residues, suggesting that the intrinsic exonuclease is functional in the mycobacterial replicative polymerase. In contrast, similar to *G. kaustophilus* PolC, DnaE2 contains an almost intact active site, with a substitution at position 8, though in this case it is a glycine that replaces the canonical aspartate residue. The PHP domain in *E. coli* PolIII represents a significant variant: five of the metal co-ordinating residues are replaced by amino acids that are not compatible with this function. However, despite the apparent loss of catalytic function, the structural scaffold of the PHP domain is conserved, and has been shown to be critical for the stability and activity of the *E. coli* PolIII (Barros *et al.*, 2013).

A detailed study by Lamers and colleagues demonstrated an apparent correlation between the presence of a variant PHP domain and the possession of separate proofreading exonucleases (Barros *et al.*, 2013); that is, bacterial genomes which do not possess an intact DnaE PHP domain containing all conserved metal co-ordinating residues appear to encode a separate exonuclease subunit which functions as proofreader during replication. The analysis by Timinskas *et al.* (2014) further observed that, in organisms in which PolC occurs with DnaE3, PolC is the main high fidelity polymerase since DnaE3 contains a disordered and degraded PHP domain which may be catalytically inactive. However, *Clostridia* - in which PolC co-occurs with DnaE1 - lack the integral exonuclease domain, suggesting that DnaE1 performs high fidelity synthesis in these organisms (Timinskas *et al.*, 2014). This raises questions about organisms like Mtb whose genome encodes a DnaE1 protein with an intact PHP domain, as well as two separate *dnaQ* homologs: what is the relative contribution of the DnaE1 PHP domain *versus* either (or both) of the *dnaQ* homologs to replication fidelity? Is there an evolutionary advantage associated with the presence of both intrinsic (DnaE1-associated) exonuclease and separate DnaQ-type proofreading subunits? Unlike in organisms such as *E. coli* (Gerdes *et al.*, 2003), *Haemophilus influenza* (Akerley *et al.*, 2002), or *Streptococcus pneumoniae* (van Opijnen *et al.*, 2009), where the ε subunit is essential for viability, both *dnaQ* homologs in Mtb are dispensable for growth (Griffin *et al.*, 2011, Sassetti *et al.*, 2001, Sassetti and Rubin, 2003). This raises additional questions that require

further investigation, including: 1) What is the role of the *dnaQ* homologs in Mtb, and 2) given that DnaE2 contains an almost complete set of metal co-ordinating residues, is the PHP domain in the non-essential PolIIIα subunit active and 3) if so, does DnaE2 PHP contribute to relative α subunit fidelity? In turn, these questions, suggested the need for studies focused on the contribution of intrinsic fidelity and proofreading to overall replication fidelity in Mtb.

In their paper, Venclovas and colleagues showed that PolC and DnaE-type polymerases also diverge in their domain composition and arrangement; for example, the OB domain is located in the C-terminal region of DnaE, while in PolC it is situated in the N-terminus (Timinskas *et al.*, 2014). DnaE2 polymerases differ from both DnaE1 and DnaE3 in that not a single DnaE2 possesses the C-terminal domain whereas the DnaE3-type polymerases are characterised by a similar domain organization as the DnaE1 group, albeit with a totally disordered and smaller (degraded) PHP domain (Timinskas *et al.*, 2014). Instead of the C-terminal τ domain that is critical for connecting DNA to the rest of the replisome during DNA replication, DnaE2 contains a C-terminal pentapeptide tail, SRDF[H/R]), that is conserved amongst 77% of bacterial species containing a DnaE2-type polymerase (Timinskas *et al.*, 2014). It has been hypothesized that this motif is required for mediating protein-protein interactions during function of the mutasome – thought to comprise ImuA', ImuB and DnaE2 proteins (Warner *et al.*, 2010) - however this remains to be demonstrated.

1.5 The mycobacterial DNA replication machinery

PolIII functions in the context of a replisome complex containing the PolIII HE, DnaB helicase, DnaG primase and SSB (single-stranded DNA-binding) proteins (**Figure 1.1**) (Johnson and O'Donnell, 2005, O'Donnell, 2006). Together with numerous interacting partners, these proteins function in bacterial DNA replication (Johnson and O'Donnell, 2005). Most of the replisome components are conserved across different bacteria, but they have been most thoroughly characterised in *E. coli* (Reyers-Lamothe *et al.*, 2010) and *B. subtilis* (Sanders *et al.*, 2010). Based on these organisms, models for the bacterial replisome have been constructed; these consist of the primosome, which includes the replicative

helicase that unwinds the duplex DNA; the primase, which is activated by DnaB to produce short RNA primers for discontinuous lagging-strand synthesis (McHenry, 2011b); and the PolIII HE, which consists of a PolIII core polymerase, the β_2 sliding clamp, and the clamp-loader complex (**Figure 1.1**). Recent studies have shown that three, instead of two, PolIII core polymerases are required for concurrent leading and lagging strand synthesis at a single replication fork to form a tri-polymerase replisome (Georgescu *et al.*, 2011, McInerney *et al.*, 2007). A single PolIII core is required for leading strand synthesis while the other two PolIII cores extend multiple Okazaki fragments on the lagging strand (Georgescu *et al.*, 2011). Three β_2 sliding clamps are present at each replication fork, with the third associated with either the clamp loader or the third PolIII core (Robinson *et al.*, 2012).

Current models of the *E. coli* replisome are based on a PolIII HE that contains three principal subunits: a seven-subunit DnaX clamp loader complex ($\tau_3\delta\delta'\chi\psi$) which binds three PolIII cores through the τ-α interaction, and the β_2 sliding clamp processivity factor which connects the PolIII core to the DNA through a separate interaction with PolIIIα. Of the remaining DnaX subunits, δ and δ' serve to load and unload the β_2 sliding clamp, while χ and ψ stabilize the interactions between DnaX, δ, and δ' (McHenry, 2011b). The genes predicted to be involved in DNA replication in Mtb are shown in **Table 1.1** and, based on this list, it appears that the mycobacterial replisome lacks obvious homologs of several components which perform key functions in the model organism (Warner *et al.*, 2014). Comparative genomic analyses established, however, that this gene complement is typical of many bacteria and, therefore, it was among the genomes which contributed to the definition of a basic bacterial replication module that contains the replication initiator protein, DnaA, the DnaB helicase, DnaG primase, PolIIIα, the β_2 sliding clamp, ε proofreading subunit, τ^3, δ and δ', SSB, DNA ligase, and PolI (McHenry, 2011b, Robinson *et al.*, 2012) (**Figure 1.1)**.

Although *in vitro* studies have provided some insight into the mycobacterial replisome, it is poorly characterised relative to model organisms (Johnson and O'Donnell, 2005, O'Donnell, 2006). The working model of a mycobacterial replisome proposed in this study is therefore based on insights provided by the *E. coli* system (**Figure 1.1**) and a number of studies focused on the mycobacterial replisome components (**Table 1.1**). However, there are still

gaps that require to be addressed. For example, the stoichiometry and architecture of the mycobacterial replisome are still unknown: how many additional copies of specific replisome components are present at the replication fork, for example? It is also unclear where the second *dnaE*-encoded α subunit, DnaE2, "sits" within the replisome. Similarly, further work is required to determine whether the ε subunit forms part of the PolIII core, as observed in *E. coli*. Moreover, additional cellular factors which may affect (or modify) the composition of replisome - such as the relative dNTP/rNTP levels present at the replication fork, as well as the role of protein-protein interactions at the replication fork - are still unclear. Again, much additional work is required to address these gaps.

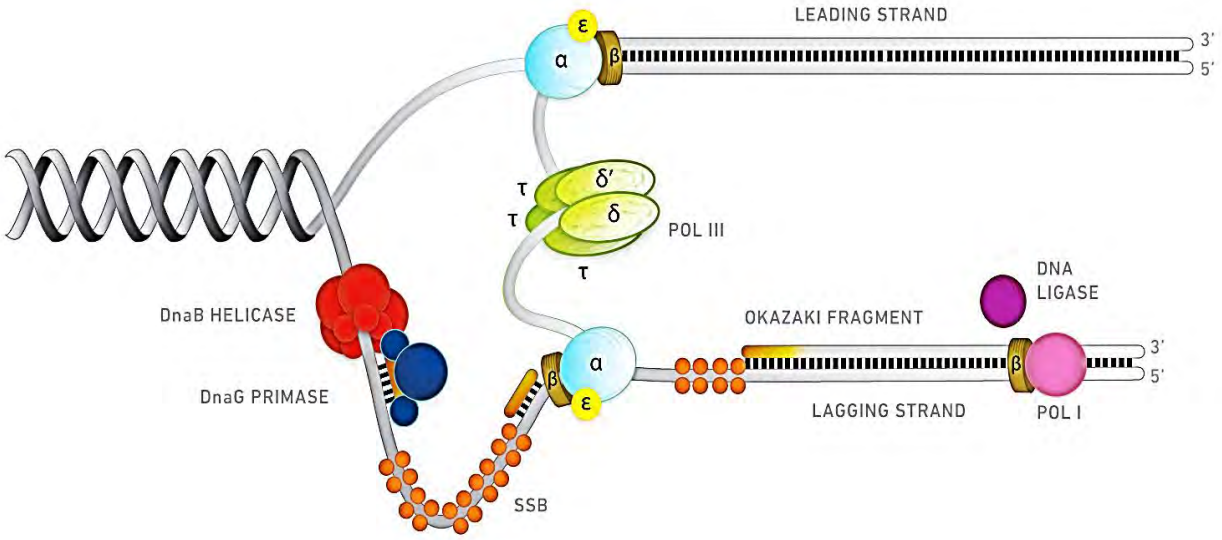


Figure 1.1: Schematic representation of the "model" bacterial replisome consisting of the PolIII core polymerase, the homodimeric β_2 sliding clamp, the τ3δδ' clamp-loader complex, DnaB helicase (red hexamer), DnaG primase (blue), PolI (pink) DNA ligase (purple) and SSB (orange). Stoichiometry and architecture of the mycobacterial replisome have been poorly characterised. It is unclear whether the mycobacterial replisome functions as a di- or tri-polymerase system or where DnaE2 sits within the replisome and also whether the ε subunit forms part of the mycobacterial PolIII core.

1.6 A model for DNA replication in mycobacteria

Despite the genetic diversity between *E. coli* and Mtb, they are assumed to share essentially the same mechanisms of chromosomal replication given that most of the replisome components are conserved (Robinson *et al.*, 2012, Warner *et al.*, 2014). Briefly, the DnaB replicative helicase, which is loaded onto the lagging strand template, separates the two templates that are subsequently copied by the PolIII core. PolIII core activation follows binding of the β_2 sliding clamp, which encircles double-stranded DNA. Sliding clamps are added and removed by the γ/τ complex clamp loader, whose τ component oligomerises the PolIII core. The leading PolIII core-β_2 complex continuously extends DNA in the direction of unwinding. DnaG primase binds to DnaB during cycles of priming and DNA synthesis on the lagging strand template. Discontinuous synthesis on the lagging strand requires PolIII to move from the finished fragment, rapidly dissociate from the β_2 clamp and DNA, and re-associate with a new clamp for extension. The β_2 clamp also interacts with PolI and DNA ligase to convert RNA to DNA and seal the finished fragment (**Figure 1.1**) (Reyers-Lamothe *et al.*, 2010, O'Donnell, 2006). Similarities and deviations between the mycobacterial and *E. coli* replication systems are discussed in detail in a recent review (Warner *et al.*, 2014) and so will be summarized briefly in this section.

Bacteria have evolved rigorous control mechanisms to regulate the initiation of DNA replication, and to ensure that it does not occur at random sequences throughout the chromosome (Reyes-Lamothe *et al.*, 2012). Instead, replication is initiated at a single site (*oriC*) and proceeds bi-directionally around the chromosome until the two replication forks meet in the replication terminus (*ter*), a region located approximately opposite *oriC*. The two strands of the template DNA are separated at the origin, yielding two fork structures. Replicative DNA polymerases and accessory proteins are assembled onto each of these forks, and synthesize new DNA bi-directionally around the circular chromosome until the two replication forks meet in the *ter*, yielding two copies of the bacterial chromosome, each containing one strand from the parental chromosome and one nascent strand. Moreover, since this must occur only once during the cell cycle, a diverse array of regulatory

mechanisms ensures that the assembly of the replication machinery is triggered at the appropriate stage (Warner *et al.*, 2014, Robinson *et al.*, 2012).

As in *E. coli*, key features of bacterial replication are retained in Mtb: the DNA-ATP interaction is critical for replication initiation, since it results in the opening of the DNA duplex to allow loading of DnaB, and the *dnaA* promoter remains active during replication to ensure progression through the cell cycle (Nair *et al.*, 2009). Mtb *oriC* is located in the 527 bp intergenic region between *dnaA* and *dnaN*, and contains multiple predicted and confirmed DnaA-binding sites. Interestingly, this region also serves as a common locus for the insertion of IS*6110* transposable elements. To date, however, there is no evidence to suggest the insertions have any effect on the replication process, including the timing of replication initiation. Instead, these sites have been used as markers for RFLP fingerprinting of clinical Mtb isolates (Turcios *et al.*, 2009).

In *E. coli*, DnaA recruits the hexameric DnaB replicative helicase to the origin to initiate strand separation. Recent work has confirmed the physical interaction of Mtb DnaA and DnaB, and has further implicated DnaB in controlling DnaA complex formation and the interaction with *oriC* (Xie and He, 2009). In contrast, Mtb does not possess a homolog of the DnaC helicase loader, which is required for loading DnaB helicase onto the DNA in *E. coli* (**Table 1.1**). This suggests that the DnaC function is performed by another protein or that, alternatively, DnaA alone might be sufficient for DnaB loading, a possibility that is supported by recent insights into the structure of DnaB from *H. pylori* – another organism lacking a DnaC helicase loader (Stelter *et al.*, 2012).

PolIII HE is loaded onto each of the two replication forks following strand separation and unwinding of the parental chromosomal DNA by the DnaB replicative helicase, with two PolIII HEs acting on the lagging strand (Georgescu *et al.*, 2011). Leading strand synthesis is highly processive and involves the continuous extension of DNA, in contrast to lagging-strand synthesis, which requires discontinuous replication via the extension and ligation of Okazaki fragments. DnaG primase produces short RNA primers for extension by PolIII. Owing to its essential role, Mtb DnaG was recently expressed and applied in a novel high-throughput enzyme assay to identify inhibitors of the priming reaction (Biswas *et al.*, 2013).

Following completion of Okazaki fragment synthesis, a second switch occurs so that PolIII is replaced by PolI, the polymerase which catalyses high-fidelity DNA synthesis across the resulting gap. Consistent with previous predictions, Mtb *Rv2228c* has been shown to be essential for growth of Mtb *in vitro* (Griffin *et al.*, 2011) and encodes a bi-functional protein that fuses RNase HI and CobC-like α-ribazole phosphatase activities in a single polypeptide. The protein is unusual in that, in addition to functioning as a classic RNase HI in cleaving RNA/DNA hybrids, it also has double-stranded RNase activity (Watkins and Baker, 2010, Minias *et al.*, 2015). Replisome components are highly conserved amongst mycobacteria (**Table 1.1**), including *M. leprae*, another obligate pathogen whose genome displays evidence of extensive decay (Cole *et al.*, 2001).

1.7 Targeting replisome components for drug development

The predicted role of DNA metabolism in pathogenesis suggests the possibility of targeting selected mycobacterial pathways with novel chemotherapeutic agents. Existing antimicrobial drugs have been identified from phenotypic screens on the basis of their ability to inhibit pathways or functions that are essential for bacterial growth and survival; for this reason, components of the DNA replication machinery represent the preferred targets (Robinson *et al.*, 2012, Warner, 2010). With the exception of DNA gyrase inhibitors, targeting the replisome components has failed to yield candidate drugs, identifying DNA replication among the most underexploited targets (Robinson *et al.*, 2012) and prompting a call for the application of new technologies including structure-based drug design and fragment-based lead generation to identify replication components for chemical inhibition (Sanyal and Doig, 2012). Recent evidence implicating the lethal incorporation of oxidized guanine into DNA as a major cause of antibiotic-induced bacterial cell death (Foti *et al.*, 2012) suggests that DNA replication and repair pathways might contribute significantly to intrinsic drug resistance and, for that reason, further supports the call for DNA metabolic pathways to be targeted aggressively as potentially novel antimicrobial therapies.

As for most other major bacterial pathogens, DNA replication and repair proteins in Mtb have been investigated for their potential exploitation as targets for antimycobacterial

agents. Type II topoisomerases are well conserved across bacteria, and inhibition of DNA gyrase by fluoroquinolones (FQ) has provided an attractive option for treatment of TB. A number of studies have shown the clinical benefit of FQs such as ofloxacin and moxifloxacin in the treatment of TB (Duong *et al.*, 2009). However, the emergence of Mtb strains resistant to FQs poses a threat to their continued use – emphasizing the need of identification of novel GyrA inhibitors with unique binding mechanisms. A recent study identified a group of novel gyrase inhibitors, benzimidazoles and *N*-linked aminopiperidinyl-based gyrase inhibitors with anti-TB activity (Hameed *et al.*, 2014a, Hameed *et al.*, 2014b). Moreover, Karkare and colleagues (Karkare *et al.*, 2013) demonstrated enhanced activity for the naphthoquinones, diospyrin and 7-methyljuglone, particularly against Mtb. Furthermore, researchers at GlaxoSmithKline (GSK) identified an additional panel of novel Mtb DNA gyrase inhibitors with antimycobacterial activity (Blanco *et al.*, 2015).

Insights from other organisms suggest that Mtb DnaE1 could be a profitable target for drug development, since it is essential for DNA synthesis and conserved across all bacteria. For example, 6-anilinouracils and their derivatives have been shown to inhibit DNA PolIII and have antibacterial activity against low GC Gram positive bacteria (Xu *et al.*, 2011, Zhi *et al.*, 2003). In addressing this gap, a recent study by Harris and colleagues identified a panel of novel imidazoline compounds that have bactericidal activity against both replicating and non-replicating Mtb and other Gram positive cocci (Harris *et al.*, 2014). The bacterial β_2 sliding clamp is also emerging as a potential drug target due to its central role in DNA replication, serving as a protein− protein interaction hub (Yin *et al.*, 2014, Georgescu *et al.*, 2008). Small molecule inhibitors (Georgescu *et al.*, 2008) and tetrahydrocarbazole derivatives have been shown to inhibit DNA replication *in vitro*, in *E. coli*, *A. baylyi*, *B. subtilis* and *S. aureus* (Yin *et al.*, 2014) – emphasizing the need to test these compounds in Mtb.

Consistent with their inferred importance for mycobacterial pathogenesis, tolerance pathways might offer an additional option for novel antibacterial therapies. There has been considerable discussion of the possibility of inhibiting tolerance mechanisms, particularly inducible mutagenesis pathways, in order to protect current drugs by targeting the mechanisms that underlie the evolution of resistance (Smith and Romesberg, 2007). In some respects, this approach can be considered analogous to inhibiting efflux pathways (Adams *et*

al., 2011): on its own, a specific efflux pump(s) is not an attractive target but, in combination with the appropriate frontline drug, its inhibition might be critical to efficacy by ensuring that the active compound is maintained at an elevated intracellular concentration (Adams *et al.*, 2014, Adams *et al.*, 2011, Gupta *et al.*, 2015). Mtb DnaE2 represents a good candidate for this approach since it is not essential for normal growth *in vitro*, yet loss of DnaE2 activity attenuates virulence *in vivo* and reduces the frequency of drug resistance mutations during chronic infection (Boshoff *et al.*, 2003). Furthermore, DnaE2 has been demonstrated to function in association with other DNA damage response proteins as a split 'mutagenic cassette' (Warner *et al.*, 2010), suggesting an alternative strategy of targeting the other pathway components, for example by disrupting the protein–protein interactions that are essential to mutasome function (Georgescu *et al.*, 2008, Yin *et al.*, 2014).

Understanding which parts are conserved when targeting any system for drug development is critical, since this has implications for the spectrum of activity of any new inhibitors against bacterial species, as well as the potential for development of drug resistance (Robinson *et al.*, 2012). This will be challenging in Mtb since there are still knowledge gaps on the conservation and architecture of the mycobacterial replisome, which may further complicate attempts of identifying possible replisome components as potential drug targets. Moreover, there are limited data supporting the creation of comprehensive maps of protein networks and interactions for Mtb (Wang *et al.*, 2010).

Table 1.1: Components of the mycobacterial replisome

Gene	Mtb	Msm	*M. leprae*	Function	Catalytic activity	Essentiality *in vitro*[a] / Comments	References
Initiation complex							
dnaA	*Rv0001*	MSMEG_6947	ML0001	Replication initiator	ATPase	Essential; required for regulation of DNA replication	(Leonard and Grimwade, 2011, Xie and He, 2009)
dnaB	*Rv0058*	MSMEG_6892	ML2680	Replicative helicase	ATPase	Essential; controls DNA complex formation and interaction with *oriC*	(Xie and He, 2009)
Primosome							
dnaG	*Rv2343c*	MSMEG_4482	ML0833	DNA primase	RNA primase	Essential; required for regulation of DNA replication	(Klann *et al.*, 1998)
priA	*Rv1402*	MSMEG_3061	ML0548	Primosome replication factor, replication restart	ATPase	Essential	
DNA PolIII core							
dnaE1	*Rv1547*	MSMEG_3178	ML1207	α subunit, polymerase activity	DNA polymerase	Essential; high fidelity replicative polymerase	(Boshoff *et al.*, 2003)
dnaQ	*Rv3711c*	MSMEG_6275	Absent	ε subunit, proofreading activity	Exonuclease	Non-essential; associated with drug resistance	(Farhat *et al.*, 2013)
dnaQ-uvrC	*Rv2191*	MSMEG_4259	Absent	Fusion/chimeric protein; N-terminal 3'-5' exonuclease domain, C-terminal UvrC-like endonuclease domain	Exonuclease/ Endonuclease	Non-essential; role in bacillary persistence	(Kesavan *et al.*, 2009)
DNA PolIII clamp loader complex							
dnaZX	*Rv3721c*	MSMEG_6285	ML2335	τ and γ subunits	ATPase	Essential	
holA	*Rv2413c*	MSMEG_4572	ML0603	δ subunit		Non-essential	

holB	*Rv3644c*	MSMEG_6153	ML0202	δ′ subunit	ATPase	Essential	
The mycobacterial replication complex							
dnaN	*Rv0002*	MSMEG_0001	ML0002	β_2 sliding clamp		Essential	
ssb	*Rv0054*	MSMEG_6896	ML2684	Single-stranded DNA binding protein		Essential [b]	
polA	*Rv1629*	MSMEG_3839	ML1381	DNA PolI	DNA polymerase	Essential; lacks a proofreading 3'-5' exonuclease activity, *polA* mutant displayed a DNA damage phenotype following UV irradiation and hydrogen peroxide treatment	(Mizrahi and Huberts, 1996, Gordhan *et al.*, 1996)
ligA	*Rv3014c*	MSMEG_2362	ML1705	DNA ligase, DNA replication	NAD-dependent DNA ligase	Essential	(Srivastava *et al.*, 2007)
ligB	*Rv3062*	MSMEG_2277	Absent	DNA ligase	ATP-dependent DNA ligase	Non-essential; role in DNA repair	(Gong *et al.*, 2005)
ligC	*Rv3731*	MSMEG_6304	Absent	DNA ligase	ATP-dependent DNA ligase	Non-essential; role in Ku-dependent non- homologous end joining (NHEJ) DSB repair pathway	(Gong *et al.*, 2005)
ligD	*Rv0938*	MSMEG_5570	Absent	DNA ligase, DSB repair	ATP-dependent DNA ligase	Non-essential; plays a central role in the mutagenic NHEJ pathway of DSB repair	(Gong *et al.*, 2004, Heaton *et al.*, 2014)
gyrA	*Rv0006*	MSMEG_0006	ML0006	DNA gyrase, subunit A (DNA topoisomerase II)	Topoisomerase, ATPase	Essential	(Merens *et al.*, 2009)
gyrB	*Rv0005*	MSMEG_0005	ML0005	DNA gyrase, subunit B (DNA topoisomerase II)	Topoisomerase, ATPase	Essential	(Merens *et al.*, 2009)
topA	*Rv3646c*	MSMEG_6157	ML0200	DNA topoisomerase I	Topoisomerase, ATPase	Essential; role in DNA repair	(Yang *et al.*, 2012)

rnhA-cobC	*Rv2228c*	MSMEG_4305	ML1637	Chimeric protein; N-terminal RNase HI domain, C-terminal CobC-like α-ribazole phosphatase domain	RNase/ α-ribazole phosphatase	Essential; deletion of *rnhB* in Msm does not alter genome stability	(Watkins and Baker, 2010)
rnhB	*Rv2902c*	MSMEG_2442	ML1611	RNase HII	RNase	Non-essential	(Minias *et al.*, 2015)
Other DNA replication proteins							
dnaE2	*Rv3370c*	MSMEG_1633	pseudogene	α subunit, TLS polymerase activity	DNA polymerase	Non-essential; involved in adaptive mutagenesis and contributes to the emergence of drug resistance	(Boshoff *et al.*, 2003, Warner *et al.*, 2010)
dinB1	*Rv1537*	MSMEG_3172	Absent	DNA PolIV.I	DNA polymerase	Non-essential; required for ribonucleotide discrimination during DNA synthesis	(Ordonez *et al.*, 2014)
dinB2	*Rv3056*	MSMEG_2294	Absent	DNA PolIV.II	DNA polymerase	Non-essential; dispensable for DNA damage tolerance in mycobacteria but instead required for ribonucleotide incorporation	(Ordonez *et al.*, 2014)
polX	*Rv3856c*	MSMEG_6445	Absent	DNA PolX	Unknown [c]	Non-essential	

[a] *In vitro* essentiality, as determined by transposon site hybridization (TraSH) (Griffin *et al.*, 2011, Sassetti *et al.*, 2003).

[b] *Rv0054* did not satisfy the strict criterion for essentiality in the study by Griffin *et al*. (Griffin *et al.*, 2011); however, no transposon (Tn) insertions were identified in any of the five possible TA dinucleotides in the ORF, suggesting that it is likely to be essential.

[c] A natural truncation in the polymerase domain of Rv3856c is predicted to preclude catalytic activity (Warner *et al.*, 2010).

1.8 Is there a role for other proteins at the replication fork?

Even though the α subunit has been demonstrated to be the dominant replicase in bacterial DNA replication, there is increasing evidence to suggest that chromosomal replication is characterised by dynamic DNA polymerase exchange (Fijalkowska *et al.*, 2012). For example, of the four accessory DNA polymerases in *E. coli*, PolI (*polA*) and PolII (*polB*) contribute directly to replication fidelity during normal chromosomal replication (Fijalkowska *et al.*, 2012) through their respective roles in high-fidelity maturation of Okazaki fragments during lagging strand synthesis (PolI), and as a back-up replicative polymerase during transient dissociation of the PolIII HE from either leading or lagging strand (PolII). The remaining accessory polymerases, DNA PolIV (*dinB*) and PolV (*umuD$_2$C*), are both members of the Y-family of specialist translesion (TLS) polymerases that function primarily in the DNA damage response, but can also access the lagging strand under conditions of elevated expression (Fijalkowska *et al.*, 2012).

Mtb does not contain a PolII enzyme, but instead has two homologs of PolIV, DinB1 and DinB2 (Kana *et al.*, 2010). Only DinB1 possesses a consensus β_2 clamp binding motif (Kana *et al.*, 2010), suggesting that DinB2 must interact with another protein(s) in order to access the replication fork. Therefore, while studies have demonstrated that the Msm homolog of Mtb DinB2 is a functional DNA polymerase with a tendency to promote G:T and T:G mismatches (Sharma and Nair, 2012), the role of DinB2 in mycobacterial DNA metabolism is unknown. Similarly, although recent work has provided important insight into the activity and function of the DinB-type Y-family polymerases in Mtb (Ordonez *et al.*, 2014), the biological roles of the mycobacterial DinBs remains speculative (discussed in sections below).

1.9 The mycobacterial replication rate

There are limited data on both replication and mutation rates of Mtb bacilli during host infection, especially during latency. Evidence from molecular epidemiological studies demonstrated endogenous reactivation of Mtb after three decades of latent infection (Lillebaek *et al.*, 2002) and this risk increases 10% per year in HIV-infected patients relative

to immune-competent individuals (Nahid and Daley, 2006). Previous models of latent TB infection (LTBI) suggested that, during the latent phase, Mtb enters a very slowly replicating or non-replicating (but perhaps metabolically active) state in which bacilli are insensitive to killing by the host immune effector molecules and anti-TB drugs (Wayne, 1977). In contrast, alternative models by Sherman and colleagues, which are based on the use of a "clock" plasmid which is lost from daughter cells during division, instead propose a stable balance *in vivo* between bacillary replication and death, probably as a consequence of active immune surveillance (Gill *et al.*, 2009). In support of this notion, recent studies suggest that latency involves a disease spectrum that extends from non-replicating persisting organisms, to replicating but asymptomatic infections, to low-level disease with higher numbers of actively replicating bacteria (Ernst, 2012, Gupta *et al.*, 2012, Lin *et al.*, 2014).

To gain insight into LTBI and define the mutational capacity of Mtb during different stages of infection, Ford and colleagues used WGS to measure the mutation rate of Mtb isolates from cynomolgus macaques with active, latent and reactivated disease (Ford *et al.*, 2011). Given that the generation time of Mtb *in vivo* is unknown, the mutation rate in this study was calculated allowing for a broad range of generation times (between 18 and 240 h). Interestingly, the authors observed similar replication and mutation rates in Mtb isolates from latent, reactivated, and actively-infected macaques. In addition, they demonstrated that macaques with clinically latent infection acquire mutations at a similar rate to rapidly replicating bacteria *in vitro* (Ford *et al.*, 2011) - an intriguing finding since these would be expected to differ on the basis that mutation rates determined *in vitro* often involve large bacterial populations either at exponential or stationary phases of growth (Gillespie, 2002) which do not represent *in vivo* conditions. Instead, the findings from this study suggest that Mtb continues dividing actively during the entire course of prolonged clinical latency, and this active replication is balanced by robust killing, at least in this model. The authors further concluded that the mutation rates observed during latency are likely attributable to oxidative DNA damage rather than replicative errors. In summary, these observations were interpreted as suggesting that the mutational capacity of Mtb during latent infection is determined primarily by the length of time the organism spends in the host environment rather than the replicative capacity and replicative errors of the organism during infection (Ford *et al.*, 2011).

In contrast to the results of Ford *et al.* (2011), Alland and colleagues reported different findings in humans: specifically, their data suggested the possibility that the non-human primate model might not appropriately recapitulate latent TB in humans (Colangeli *et al.*, 2014). Again using WGS, Alland and colleagues calculated the replication and mutation rates of latent Mtb by comparing the genome sequence of a single strain that had been transmitted from a single, incident TB case and resulted in TB disease in close contacts over a period of twenty years. In contrast to findings in macaques, these authors observed lower mutation rates during latency, for any given generation time, even after adjusting for the predicted higher mutation rate that was considered likely to occur during the final stage of infection, as the individual progressed to active TB (Colangeli *et al.*, 2014). Moreover, analysis of the mutation spectrum in the human LTBI model did not reveal a higher proportion of mutations associated with oxidative damage (GC>AT or GC>TA mutations) as observed in macaques; instead, there were fewer mutations associated with oxidative damage, suggesting that, during latent infection in humans, the bacterium's mutational capacity is driven by replicative mutagenesis rather than oxidative stress, contrary to the interpretation of Ford *et al*. based on the non-human primate model.

These findings by Colangeli *et al*. also suggest that the mycobacterial mutation rate is altered during different stages of infection – in turn implying that there is possibly less host pressure on the organism during latency, and therefore, a reduced requirement to replicate and mutate to facilitate adaptation. Importantly, the results of this study might counter suggestions that INH monotherapy for latent TB is a risk factor for the emergence of INH resistance (Ford *et al.*, 2011), given that the rate at which clinical drug resistance emerges depends on the number of bacteria in a latently-infected individual and this has been shown to be very low (Colangeli *et al.*, 2014). However, the small number of samples used in this study, coupled with the fact that the replication and mutation rates which occur during latency are difficult to verify experimentally, reinforces the need for further research to address this issue definitively.

1.10 Replication fidelity and the Mtb mutation rate

The *in vitro* mutation rate in Mtb has been estimated at ~2.9 x10^{-10} per base pair per round of replication (Ford *et al.*, 2011). This figure, which was derived from fluctuation analyses that utilized *rpoB* as the target for rifampicin resistance and were corrected for the mutational target size (Ford *et al.*, 2011), is comparable to *E. coli*. In that organism, it is estimated intrinsic replication fidelity by the replicative polymerase, the removal of any mis-incorporated nucleotides by the 3'-5' exonuclease activity of the replicative polymerase itself or its interacting proofreading subunit, and post-replicative mismatch repair (MMR), contribute 10^{-5}, 10^{-2}, and 10^{-3}, respectively, to the overall error rate of 10^{-10} (Fijalkowska *et al.*, 2012). However, Mtb lacks the MMR system, suggesting that intrinsic fidelity and/or proofreading are able to maintain overall mycobacterial replication fidelity, or that alternative mechanisms exist for the correction of replication errors, perhaps including a non-orthologous system for MMR (Mizrahi and Andersen, 1998).

While the possibility exists that mycobacteria encode a non-orthologous alternative to the canonical, and widely spread MutHLS-based MMR system (Mizrahi and Andersen, 1998), multiple lines of evidence instead suggest that other repair components have enabled Mtb to mitigate the lack of MMR function (Springer et al. 2004; Machowski et al. 2007; Wanner et al. 2008). For example, Springer and colleagues demonstrated that the key NER helicase, UvrD1, fulfils a critical role in limiting recombination-associated mismatches in mycobacteria (Guthlein *et al.*, 2009). In addition, intrinsic features of the genome itself might limit the risk of replication errors: although short sequences of nucleotide repeats pose a significant problem to replicative DNA polymerases and can result in frameshift mutations, the GC-rich Mtb genome appears to have been under strong selective pressure to restrict the number of repeat regions through context-dependent codon choice (Wanner *et al.*, 2008). The maintenance of genome stability within the host during long-term persistence is crucial for survival and propagation of Mtb. However, that fundamental must be balanced against the evolutionary need to adapt genetically to the stresses and fluctuating environments encountered during infection, including those imposed by anti-tubercular drug administration (Dos Vultos *et al.*, 2009, Gorna *et al.*, 2010). Acquisition of foreign DNA can accelerate adaptation in bacteria which reproduce asexually; in Mtb, however, evolution

appears to be limited to the selection and maintenance of chromosomal mutations and rearrangements (Hershberg *et al.*, 2008) and is, therefore, dependent on the generation of mutant alleles which provide the genetic diversity for selection. Proof of adaptation in Mtb is demonstrated by the emergence of MDR and XDR strains; moreover, recent epidemiological data indicate a high degree of genetic diversity between clinical strains that correlates well with human demography (Hershberg *et al.*, 2008). These observations suggest that there is a balance between the activity of error-free DNA repair mechanisms in maintaining genomic integrity, and the operation of mutagenic pathways which might generate diversity during host infection (Warner *et al.*, 2010, Warner *et al.*, 2013). This balance is set by the fidelity of the DNA replication, repair, and recombination (or "3R") system.

The high-fidelity operation of 3R pathways favours genome stability, whereas relaxed fidelity - or loss of specific 3R functions (Dos Vultos *et al.*, 2008) - might facilitate genetic adaptation (Warner *et al.*, 2010, Warner *et al.*, 2013). Evidence demonstrating that other bacterial pathogens are associated with strains exhibiting elevated mutation rates as a result of defective DNA repair functions, most commonly the MMR system (Sundin and Weigand, 2007), suggests that increased mutability might facilitate adaptation to dynamic hostile environments. However, consistent with the idea that most mutations are deleterious, mutator strains are not found in all bacterial populations and, in most cases, the benefits of a mutator phenotype for colonization and the development of drug resistance are likely to be outweighed by negative effects on survival and virulence (Warner *et al.*, 2015). Again, this suggests the need to maintain a balance between stability and mutagenesis. In a provocative study, Gicquel and colleagues found high numbers of polymorphisms in genes of the 3R system of Mtb compared with housekeeping genes, thereby identifying strong selection pressure on 3R genes as a common component in the modern evolutionary history of different strain lineages (Dos Vultos *et al.*, 2008). It is tempting to speculate that the identified polymorphisms result in a relaxation of 3R fidelity that facilitates the adaptation of Mtb to genotoxic conditions and, perhaps, drives the expansion of a clonal infecting bacillus into a microdiverse bacillary population (Warner *et al.*, 2015); however, this has not been demonstrated and so remains speculative.

1.11 Using Msm as a model organism for generating mutator strains

M. smegmatis mc^2155 (Msm) was used in this study as the parental strain for generating mycobacterial mutants containing site-directed substitutions in key active-site and protein-binding domains of the relevant replisome components. Msm is a fast-growing, non-pathogenic mycobacterium widely used as a model organism to study the biology of other virulent and slow-growing species like Mtb (Snapper *et al.*, 1990). However, Msm is not an appropriate model for studying virulence or pathogenesis of Mtb. Important observations with regards to virulence factors of this pathogen should, therefore, be confirmed in Mtb (Reyrat and Kahn, 2001). Prior work in the Warner and Mizrahi laboratory has demonstrated the value of using Msm as model mycobacterium, primarily since the DNA replication and repair systems are highly similar (**Table 1.1**). As such, corresponding phenotypes have been observed for mutants generated in Mtb and Msm (Boshoff *et al.*, 2003, Warner *et al.*, 2010). In addition, given legitimate concerns about the potential biosafety implications contingent on generating mutator strains of Mtb, Msm represents an ideal model for studying mutational dynamics and replication fidelity in mycobacteria.

1.12 Mutagenesis in Mtb

Mtb is not a natural mutator (Ford *et al.*, 2011). Moreover, since drug resistance in this organism occurs primarily through chromosomal mutation, it seems fair to conclude that mutational capacity is the key determinant in developing drug resistance. Numerous studies have provided both experimental and clinical evidence demonstrating that Mtb strains from different lineages vary in their capacity to cause disease (de Jong *et al.*, 2008, Coscolla and Gagneux, 2010, Kato-Maeda *et al.*, 2012, Glynn *et al.*, 2002) and to acquire drug resistance (Anh *et al.*, 2000, Huang *et al.*, 2010, Sun *et al.*, 2012, Johnson *et al.*, 2010). However, the evidence for an association between specific Mtb strains and an elevated mutation rate is mixed. Gicquel and colleagues demonstrated that Mtb strains from the Beijing family contain mutations in genes whose disruption in other bacteria confers a mutator phenotype (Ebrahimi-Rad *et al.*, 2003). Moreover, strains from this lineage have also been shown to have polymorphisms in DNA replication, recombination and repair genes, raising the possibility that they have higher mutation rates (Mestre *et al.*, 2011). Also consistent with

these observations, Fortune and colleagues used WGS to demonstrate strain-based differences in mutation rates between lineage 2 (East Asian) and lineage 4 (Euro American) strains (Ford *et al.*, 2013). Specifically, the authors reported that Mtb lineage 2 strains acquire drug resistance *in vitro* more rapidly than lineage 4 strains and, further, that the observed differences were not due to the enhanced ability of lineage 2 strains to adapt to antibiotic pressure but rather due to a higher basal mutation rate in the presence of the drug. Nevertheless, the mechanism underlying the inferred difference in mutation rates between the selected lineages remains to be determined (Ford *et al.*, 2013).

In contrast to these studies, previous *in vitro* analyses observed no differences between the mutation rates of the Beijing *versus* non-Beijing strains (Werngren and Hoffner, 2003), consistent with the idea that multiple factors other than the mutation rate contribute to the apparent success of the Beijing clade (Parwati *et al.*, 2010). It is possible that the differences observed between these studies are due to the representative Beijing *versus* non-Beijing strains that were used. In this respect, whilst findings from Fortune and colleagues offer compelling evidence of strain-based differences in mutation rates, it is worth noting that: 1) the CDC155 strain employed as an exemplar lineage 4 strain represents a minor branch within this lineage, with its own mode of evolution (Mokrousov, 2014); and 2) HN878, as representative lineage 2 strain, has similarly separated from other Beijing family members (Mokrousov, 2014). This suggests that the isolates studied by Fortune and colleagues might not necessarily be truly representative of their lineages and, in turn, implies that further studies are required to address this important question adequately.

In addition to errors made by the replicative polymerase during DNA replication, mutations can also arise as a consequence of DNA damage by endogenous and exogenous DNA-damaging agents. Damage repair and reversal in mycobacteria have been investigated using genetic and biochemical approaches; however, there are relatively few examples of studies which have utilized a combination of both (Gupta *et al.*, 2011, Warner, 2010, Warner and Mizrahi, 2011). Given the threat the DNA lesions pose to viability, mycobacteria contain multiple, conserved mechanisms for detecting, processing and repairing these lesions. This,

again, emphasizes the potential contribution of DNA repair and/or tolerance pathways to the evolution and pathogenesis of Mtb.

1.13 The mycobacterial complement of Y-family polymerases

Specialized TLS DNA polymerases allow bypass of replication-blocking lesions that have escaped detection and repair, and so provide a mechanism for tolerating DNA damage (Yang and Woodgate, 2007). Most TLS polymerases belong to the Y-family of DNA polymerases which comprises a wide range of structurally related proteins present in bacteria, archaea, and eukaryotes (Ohmori *et al.*, 2001). Many Y-family homologs are upregulated in response to stress, suggesting a potential role in induced mutagenesis (Andersson *et al.*, 2010). The *E. coli* genome contains three TLS polymerases, all of which are upregulated in the DNA damage or SOS response: the B-family polymerase PolII, and the Y-family polymerases PolIV and PolV which are encoded by *dinB* and *umuDC*, respectively, (Goodman, 2002). In contrast to *E. coli*, the Mtb genome does not encode a B-family DNA polymerase, suggesting that all specialist bypass function in Mtb depends on the two PolIV polymerase homologs, originally annotated as DinP (DinB2) and DinX (DinB1) (Mizrahi and Andersen, 1998). In contrast to most bacterial systems, neither *dinB1*- nor *dinB2*-encoded PolIV homolog is upregulated in the mycobacterial damage response (Boshoff *et al.*, 2003, Davis *et al.*, 2002, Rand *et al.*, 2003, Warner *et al.*, 2010); instead, both genes are expressed constitutively during logarithmic growth and stationary phase (Kana *et al.*, 2010). Deletion of *dinB1* and *dinB2* had no effect on the sensitivity of Mtb to multiple DNA damaging agents (Kana *et al.*, 2010), suggesting the dispensability of these polymerases for DNA damage tolerance. Furthermore, overexpression of either Mtb *dinB1* or *dinB2* in WT mc^2155 did not increase the spontaneous mutation rate (Kana *et al.*, 2010). This evidence suggests that DinB homologs from mycobacteria do not behave like their counterparts from other organisms.

In addition to DinB1 and DinB2, Msm encodes a third DinB paralog, DinB3, which is not present in the proteome of Mtb. Shuman and colleagues recently demonstrated that Msm DinB1 and DinB3 are distantly related to DinB2 and, further, that both proteins function as typical DNA-dependent DNA polymerases with very poor to no ability to incorporate rNTPs (Ordonez *et al.*, 2014). Further analysis showed that an aromatic steric gate side chain -

absent in DinB2 - enables rNTP discrimination and so enables DinB1 and DinB3 to differentiate between dNTPs and rNTPs during DNA synthesis (Ordonez *et al.*, 2014). All previously characterised DinB polymerases have aromatic steric gates. Phylogenetic analysis of the phylum *Actinobacteria* revealed that this family consists of DinB2-like polymerases that lack the canonical aromatic steric gate residue and instead have a leucine side chain which governs the naturally strong RNA polymerase activity of this enzyme in mycobacteria (Ordonez *et al.*, 2014).

Of the three DinB paralogs, only *dinB3* is upregulated in the Msm SOS response. The fact that DinB2, which is able to incorporate rNTPs, is not included in the Msm SOS regulon seems to imply that, following DNA damage, repair is primarily directed toward dNTP-dependent repair. In addition to rNTP discrimination, the steric gates of Y-family *E. coli* PolIV (Jarosz *et al.*, 2006) and PolV (Shurtleff *et al.*, 2009) as well as human polymerase κ (Niimi *et al.*, 2009) have been shown to be required for bypassing DNA lesions during TLS. In contrast, this function is performed by the C-family polymerase, DnaE2, in mycobacteria, demonstrating further deviations from model organisms.

In *E. coli*, the Y-family DNA PolV is a functional homolog of Mtb DnaE2 and it is encoded by the *umuD$_2$C* gene complex. UmuD' and UmuC proteins play a role as part of a DNA damage check-point control in response to replication-blocking DNA damage (Murli and Opperman, 2000, Opperman *et al.*, 1999). This check-point involves the interactions of UmuD and UmuC with the β_2 clamp and ε proofreading subunit of DNA PolIII (Sutton *et al.*, 2001a, Sutton *et al.*, 2001b). The equivalent function for Mtb would, therefore, include interaction of ImuB with DnaE1 and DnaE2 as well as the β_2 clamp. In support of this notion, ImuB has been shown to be highly expressed under conditions of DNA damage (Warner *et al.*, 2010). Interactions of UmuD' with the α and β subunits of DNA PolIII are crucial for the TLS DNA synthesis activity of DNA PolV *in vivo* (Bridges 2001; Walker 2001) and *in vitro* (Tang *et al.*, 2000). By interacting differently with components of DNA PolIII, the two different forms of the *umuD* gene product play a role in managing the actions of UmuC (Sutton *et al.*, 2001a, Sutton *et al.*, 2001b, Sutton *et al.*, 1999).

In addition to TLS, *E. coli* DinB is involved in adaptive mutagenesis (McKenzie *et al.*, 2001, Tompkins *et al.*, 2003, Yeiser *et al.*, 2002), catalysing efficient and accurate TLS across certain N^2-dG bulky adducts (Jarosz *et al.*, 2006, Jarosz *et al.*, 2007b, Kumari *et al.*, 2008, Minko *et al.*, 2008) as well as tolerance of alkylation damage (Bjedov *et al.*, 2007). Furthermore, overexpression of PolIV significantly increases mutation rates in *E. coli* (Goodman, 2002, Jarosz *et al.*, 2007a). The *dinB* gene is under the transcriptional control of the LexA repressor and is the only SOS-regulated gene required at induced levels for stress-induced mutagenesis in this organism (Courcelle *et al.*, 2001). In addition, PolIV has been shown to catalyse TLS-mediated recovery of stalled replication forks at DNA lesions. Following replication stalling, PolIV replaces PolIII through the β_2 clamp-mediated increase in PolIV-dNTP affinity as a strategy for cell survival during stressful conditions (Wagner *et al.*, 2000, Bertram *et al.*, 2004). This function has been shown to be performed by DinB2 in mycobacteria (Ordonez *et al.*, 2014), however, the exact mechanism still needs to be established.

1.14 Damage tolerance in Mtb

Mtb is unusual in that it contains both SOS-dependent (Smollett *et al.*, 2012) and SOS-independent (Gamulin *et al.*, 2004) DNA damage responses, with some repair components induced by both mechanisms (Rand *et al.*, 2003). In contrast to model organisms, neither *dinB* homolog is included in the mycobacterial DNA damage response (Boshoff *et al.*, 2003). Instead, the mycobacterial SOS regulon is limited to the d*naE1* and *dnaE2*-encoded catalytic α subunits of PolIII (Boshoff *et al.*, 2003, Warner *et al.*, 2010). Loss of DnaE2 activity renders Mtb hypersensitive to DNA damage and eliminates induced mutagenesis. In addition, inactivation of *dnaE2* attenuates virulence and reduces the frequency of drug resistance in a mouse model (Boshoff *et al.*, 2003). Coupled with the induction of *dnaE2* during stationary infection, these observations implicate a DnaE2-mediated mutagenic mechanism in both pathogenesis and the adaptive evolution of drug resistance during persistence (Boshoff *et al.*, 2003).

Recent studies demonstrated that DnaE2 operates in a novel mutagenic pathway comprising two additional accessory proteins, ImuA' and ImuB. These proteins are often

encoded in bacterial genomes as a mutagenic gene "cassette", which is widely distributed across bacteria that do not encode the *umuD$_2$C* gene complex (Erill *et al.*, 2006, Galhardo *et al.*, 2005). ImuB is one of three putative Y-family polymerase homologs in the Mtb genome but it appears from sequence analysis to be catalytically inactive (Warner *et al.*, 2010). Although this predication requires formal demonstration, the current model for the mycobacterial "mutasome" holds that DnaE2 functions as the TLS polymerase with ImuB acting as hub protein that interacts with both ImuA' and DnaE2 via the C-terminal domain, and with the β_2 clamp via a β_2 clamp binding motif (Warner *et al.*, 2010). These observations suggest that the split *imuA'-imuB/dnaE2*-encoded system constitutes a non-orthologous replacement of the *E. coli* UmuD'$_2$C-RecA-ATP PolV mutasome (Jiang *et al.*, 2009).

The basis for the functional specialization of C-family replicative and TLS polymerases in Mtb and other Gram-positive bacteria remains unclear (Bruck *et al.*, 2003, Le Chatelier *et al.*, 2004). Unlike Y-family polymerases whose structures are adapted to specialist lesion bypass (Yang and Woodgate, 2007), sequence analysis reveals few clues as to DnaE2 function. Structural determinants such as active site architecture are likely to contribute significantly to inherent fidelity; however, it is also possible that differential interactions with other DNA metabolic proteins modulate polymerase function.

1.15 Structural and functional differences between the mycobacterial *dnaE*-encoded α subunits

Although DnaE1 and DnaE2 are similar in terms of amino acid identity, they perform different functions (Boshoff *et al.*, 2003). DnaE1 is an essential, high fidelity replicative polymerase in Mtb, whereas DnaE2 functions in non-essential, error-prone TLS. **Table 1.2** highlights some of the differences between DnaE1 and DnaE2. Previous studies identified DnaE2 as central player in the DNA damage response in Mtb and, importantly, demonstrated a critical role for the TLS polymerase in virulence and the emergence of drug resistance *in vivo* (Boshoff *et al.*, 2003), as mentioned above. However, the role for DnaE2 as an error-prone TLS polymerase is not general across bacterial phyla; for example, *Pseudomonas putida* DnaE2 has been shown to have antimutator properties (Koorits *et al.*, 2007) whereas the homologous protein in *P. aeruginosa* has been shown to be dispensable

for damage tolerance (Cirz *et al.*, 2006) but not for induced mutagenesis (Sanders *et al.*, 2006). In addition, in *Streptomyces coelicolor*, *dnaE2* was shown to be SOS-inducible, but dispensable for DNA replication, linear chromosome end patching, ultraviolet resistance and mutagenesis (Tsai *et al.*, 2012).

All major DNA PolIIIα structural features are readily identifiable in both DnaE1 and DnaE2, except for the C-terminal τ-interacting domain which is absent in DnaE2 (**Table 1.2**) (Timinskas *et al.*, 2014, Warner *et al.*, 2010). A strong α-τ interaction enables simultaneous leading and lagging strand synthesis by PolIII HE in *E. coli* (Johnson & O'Donnell 2005); the absence of this region might, therefore, account for the inability of DnaE2 and other non-essential polymerases to substitute essential replicative function (Boshoff *et al.*, 2003, Bruck *et al.*, 2003, Le Chatelier *et al.*, 2004). Evidence implicating a defective τ-PolIIIα interaction in a mutator phenotype (Gawel *et al.*, 2008) suggests that the ability to bind τ might contribute to relative polymerase activity. In addition, DnaE2 lacks a consensus β_2 clamp binding motif, QL[S/D]LF, that is required for accessing DNA during DNA synthesis (**Table 1.2)** (Dalrymple, 2001) as well as the complete set of PHP domain metal co-ordinating residues that are required for intrinsic proofreading (Baños *et al.*, 2008, Stano *et al.*, 2006).

Another critical interaction in the *E. coli* replisome is that which occurs between PolIIIα and the *dnaQ*-encoded ε subunit. In *E. coli*, disruptions to proofreading activity enable PolIII-mediated TLS in the absence of specialist DNA repair polymerases IV and V (Borden et al., 2002, Vandewiele et al., 1998). Moreover, in organisms such as *S. pyogenes*, the essential DnaE subunit that catalyses error-prone TLS does not bind DnaQ (Bruck et al., 2003). In combination, these observations suggest that differential interactions with the DnaQ-like Rv3711c and Rv2191 might determine α subunit function in Mtb. Additionally, a Zn^{2+}-dependent 3'-5' exonuclease activity was recently located to the PHP domain of DnaE-type polymerases (Barros *et al.*, 2013, Stano *et al.*, 2006). As detailed above, one of the key objectives of the current study was to evaluate the relative contributions of both PHP domain and DnaQ mediated exonuclease activity to PolIII function and replicative fidelity.

In conclusion, while comparative genomic surveys of DNA replication and repair proteins in Mtb have identified components of most major pathways, at the inception of this project, no studies had been conducted to investigate the functional contribution of key replisome components and interacting proteins to DNA replication fidelity. In particular, the structural and functional basis for the differences between DnaE1 and DnaE2 were poorly understood. This project was therefore formulated to address this gap.

Table 1.2: Structural and/or functional differences between DnaE1 and DnaE2 polymerases of Mtb

Structural and/or functional properties	DnaE1	DnaE2	References
Function	High fidelity chromosomal replicase	Error prone TLS polymerase	(Boshoff *et al.*, 2003)
β_2 clamp binding motif	Yes	No	(Boshoff *et al.*, 2003, Warner *et al.*, 2010)
In vitro essentiality [a]	Yes	No	(Boshoff *et al.*, 2003)
τ-binding domain	Yes	No	(Warner *et al.*, 2010)
Intrinsic proofreading domain	Yes	Unknown	
Complete set of metal co-ordinating residues [b]	Yes	No	(Warner *et al.*, 2010, Timinskas *et al.*, 2014)
Partners during replisome function	DnaN (β_2 clamp)	-	(Warner *et al.*, 2010)
Partners during damage response	-	ImuA' and ImuB	

[a] *In vitro* essentiality, as determined by TraSH (Griffin *et al.*, 2011, Sassetti *et al.*, 2003).

[b] These residues were determined by sequence alignment using protein sequences from *E. coli* (Barros *et al.*, 2013, Lamers *et al.*, 2006) and *T. aquaticus* (Bailey *et al.*, 2006) PolIIIα subunits as reference.

CHAPTER 2: AIMS AND OBJECTIVES

As noted above (**Table 1.2 and Figure 1.2**), previous work from the Warner and Mizrahi laboratory (Boshoff *et al.*, 2003, Warner *et al.*, 2010) and others (Timinskas *et al.*, 2014) has identified at least four major differences in the structural organizations of the DnaE1 *versus* DnaE2 α subunits, specifically:

1. The presence in DnaE1 of conserved residues that are predicted to be required for proofreading;
2. A pentapeptide C-terminal motif, SRDFR, that is present in DnaE2 and other non-essential DnaE-type polymerases;
3. An extended N-terminal region in DnaE2 that is conserved in DnaE2 homologs in *Actinobacteria* only
4. The absence in DnaE2 of the τ binding domain that is essential for co-ordinated leading and lagging strand DNA synthesis by the replisome

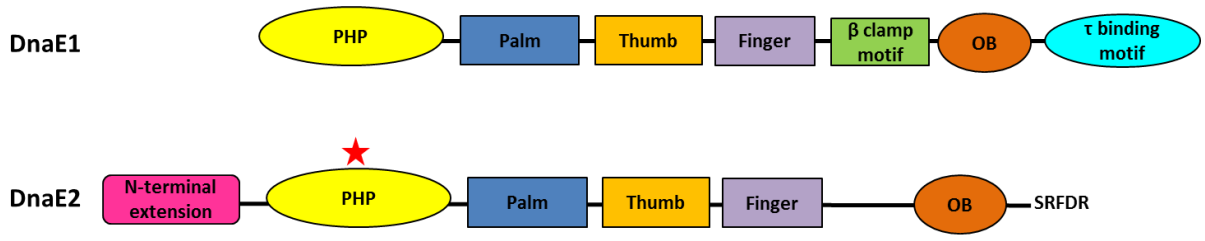


Figure 1.2: Schematic representation of key differences in the structural organizations of DnaE1 and DnaE2 PolIIIα subunits. The asterisk denotes mutations in conserved residues in the PHP domain of DnaE2. τ binding domain = tau-binding domain. See text for details.

2.1 Aim

The overall aim of this study was to investigate possible factors contributing to the differential functions of DnaE1 and DnaE2 under conditions of genotoxic stress.

2.2 Objectives

The specific objectives of this study were:

1. To determine the contribution of conserved PHP domain residues to the fidelity of the *dnaE1*-encoded α subunit;
2. To investigate the role of the mycobacterial proofreading DnaQ subunit homologs in maintaining the fidelity of DnaE1-dependent DNA replication; and
3. To determine the role of the C-terminal pentapeptide motif and N-terminal extension in the activity of the *dnaE2*-encoded error-prone TLS polymerase

CHAPTER 3: MATERIALS AND METHODS

All culturing and DNA manipulations were performed according to standard protocols (Parish and Stoker, 2000, Snapper *et al.*, 1990, Sambrook *et al.*, 1989, Sambrook and Russell, 2001). Details of growth media and general solutions used in this study are outlined in **Appendix A**. Basic microbiological and molecular genetics work was carried out using standard methods that are well established in the Molecular Mycobacteriology Research Unit (MMRU).

3.1 Mycobacterial strains, plasmids and growth conditions

All plasmids and bacterial strains used in this study are listed in **Tables 3.1** and **3.2**. Bacterial strains were stored in 33% (v/v) glycerol and stored at -80°C.

3.1.1 Bacterial culturing conditions

E. coli strains

E. coli strains used for cloning procedures were grown in Luria Bertani (LB) broth supplemented with an appropriate antibiotic, either standing at 30°C for 48 h in an IncoTherm incubator (Labotec) or overnight with vigorous shaking at 37°C in an IncoShake incubator (Labotec), depending on the size of the plasmid. For growth on solid medium, cells were streaked on Luria Bertani agar (LA) plates containing the relevant supplements, and incubated at 37°C or 30°C in the case of knockout constructs. Antibiotics were used at the following concentrations: 100 µg/ml ampicillin (Ap), 200 µg/ml hygromycin (Hyg), and 50 µg/ml kanamycin (Km). For blue/white selection, 40 µg/ml of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) was added to the growth medium.

Mycobacterial strains

Msm strains were grown in Middlebrook 7H9 medium (Difco, Becton Dickinson, USA) supplemented with 10% Middlebrook Oleate-Albumin-Dextrose-Catalase (OADC), 0.05% Tween 80 and 0.05% glycerol (7H9 OADC) or Middlebrook 7H10 solid medium supplemented with 10% OADC and 0.05% glycerol (7H10 OADC). Antibiotics were used at the following concentrations: 50 μg/ml Hyg and 20 μg/ml Km.

***Saccharomyces cerevisiae* (Yeast) strains**

Yeast strains were grown in rich YPD broth (0.1% Yeast extract, 0.2% Peptone and 0.2% Glucose) or on YPD agar (YPDA, 0.15% agar) supplemented with 0.003% adenine. Selection of strains carrying plasmids was on minimal synthetic drop-out (SD) medium without leucine and tryptophan. SD medium (0.67% yeast nitrogen base without amino acids, 2% glucose, 2% Bacto agar and 0.02% drop-out mix) was used for selection of plasmids based on the use of auxotrophic mutant strains which cannot grow in the absence of a specific medium component. Only yeast strains containing plasmid(s) with the gene required for growth in the absence of the specific component enables the transformants to grow on medium lacking that required component.

Table 3.1: Strains used in this study

Strain designation	Genotype	Description	Reference/ Source
Yeast			
AH109		*HIS3*, *ADE2*, *lacZ*, *MEL1* containing reporter strain for screening protein interactions, *MATa*, *trp*1-901, *leu*2-3, 112, *ura*3-52, *his*3-200, g*al*4Δ, *gal*80Δ, *LYS2*::*GAL1UAS-GAL1TATA-HIS3*, *MEL1 GAL2UAS-GAL2TATA-ADE2*, *URA3*::*MEL1UAS-MEL1TATA-lac*Z	Clontech
Msm			
WT	mc^2155	High-frequency transformation mutant of Msm mc^26; ATCC 706	(Snapper *et al.*, 1990)
ZD 1	*dnaE1*^{D49A}	mc^2155 mutant with the aspartate to alanine substitution at position 49 in DnaE1	This study
ZD 2	*dnaE1*^{E133A}	mc^2155 mutant with a glutamate to alanine substitution at position 133 in DnaE1	This study
ZD 3	*dnaE1* ::p2NIL*dnaE1*^{D228N}	Merodiploid mutant of mc^2155 containing both WT and the mutant allele with an aspartate to asparagine substitution at position 228 in DnaE1 at the native *dnaE1* chromosomal locus	This study
ZD 4	*dnaE1*^{E133A}::*dnaE1*	*dnaE1*^{E133A} mutant of mc^2155 complemented with wild-type (WT) *dnaE1*	This study
ZD 5	Δ*dnaQ* (MSMEG_6275)	*dnaQ* knockout mutant of mc^2155	This study
ZD 6	Δ*dnaQ dnaE1*^{E133A}	Δ*dnaQ dnaE1*^{E133A} double mutant of mc^2155	This study
Δ*dnaQ-uvrC*	ΔMSMEG_4259	mc^2155 deletion mutant of the second *dnaQ* homolog which comprises the N-terminal 3'-	(Ford, 2012)

		5’ exonuclease domain and the C-terminal UvrC-like endonuclease domain	
Δ*dnaQ* Δ*dnaQ-uvrC*	ΔMSMEG_6275 ΔMSMEG_4259	Double *dnaQ* and *dnaQ-uvrC* knockout mutant of mc^2155	(Ford, 2012)
Δ*dnaE2*	Δ*dnaE2*::*aph*	*dnaE2* knockout mutant of mc^2155; Km^R	(Boshoff *et al.*, 2003)
ZD 7	Δ*dnaE2*::*aph* *attB*::*dnaE2*	Δ*dnaE2*::*aph* complemented with *dnaE2* integrated at *attB* site; Km^R	This study
ZD 8	Δ*dnaE2*::*aph* *attB*::$dnaE2^{\Delta SRDFR}$	Δ*dnaE2* mutant complemented at *att*B site with a mutant form of *dnaE2* in which the C-terminal pentapeptide motif, SRDFR, is truncated; Km^R	This study
ZD 9	Δ*dnaE2*::*aph* *attB*::$dnaE2^{N38}$	Δ*dnaE2* mutant complemented at *att*B site with a mutant form of *dnaE2* in which the 38 N-terminal amino acids are truncated; Km^R	This study
ΔL	Δ*dinB1* Δ*dinB2* Δ*dinB2::hyg* Δ*dinB3* Δ*dnaE2*	Quintuple gene deletion mutant of mc^2155 in which both *dinB2* alleles, *dinB3*, *dinB1* and *dnaE2* are inactivated; Hyg^R	D. Warner, unpublished
ZD 10	ΔL $dnaE1^{D49A}$	$dnaE1^{D49A}$ mutant of mc^2155 in the ΔL background; Hyg^R	This study
ZD 11	ΔL $dnaE1^{E133A}$	$dnaE1^{E133A}$ mutant of mc^2155 in the ΔL background; Hyg^R	This study

Table 3.2 Plasmids used in this study

Plasmids	Description	Source
p2NIL	*E. coli* cloning vector, Km^R	(Parish and Stoker, 2000)
pMC1r	Derivative of pMC1s (Ehrt *et al.*, 2005) with *Not*I fragment containing *tet* repressor removed, Km^R	(Warner *et al.*, 2010)
pGOAL17	Plasmid carrying *lacZ* and *sacB* genes as a *Pac*I cassette; Ap^R	(Parish and Stoker, 2000)
pGOAL19	Plasmid carrying *hyg*, *lacZ* and *sacB* genes as a *Pac*I cassette; Ap^R	(Parish and Stoker, 2000)
pMC1rE1	pMC1r carrying full length *dnaE1* at the bacterial *att* locus; Km^R	This study
p2NILE1	p2NIL carrying full length *dnaE1* containing *Pac*I cassette from pGOAL19; Km^R, Hyg^R, Suc^S	This study
$p2NILE1^{D49A}$	p2NIL carrying full length *dnaE1* with the D49A substitution and *Pac*I cassette from pGOAL19; Km^R, Hyg^R, Suc^S	This study
$p2NILE1^{E133A}$	p2NIL carrying full length *dnaE1* with the E133A substitution and *Pac*I cassette from pGOAL19; Km^R, Hyg^R, Suc^S	This study
$p2NILE1^{D228N}$	p2NIL carrying full length *dnaE1* with the D228N substitution and *Pac*I cassette from pGOAL19; Km^R, Hyg^R, Suc^S	This study
p2NILΔ*dnaQ*	p2NIL carrying Δ*dnaQ* allele containing *Pac*I cassette from pGOAL19; Km^R, Hyg^R, Suc^S	This study
pOmsmE2	Multicopy *E. coli–Mycobacterium* integrating shuttle vector carrying full length *dnaE2*; Hyg^R	(Boshoff *et al.*, 2003)
pMC1r*dnaE2*	Msm *dnaE2* complementation vector - pMC1r carrying full length *dnaE2;* Km^R	This study
$pMC1rdnaE2^{\Delta SRDFR}$	pMC1r carrying *dnaE2* with the C-terminal pentapeptide sequence SRDFR truncated; Km^R	This study
$pMC1rdnaE2^{N38}$	pMC1r carrying *dnaE2* with the 38 N-terminal amino acids truncated; Km^R	This study
pGADT7	Yeast-two-hybrid (Y2H) vector to produce activation domain (AD) fusions, GAL4 $_{(768\text{-}881)}$ AD, LEU2, HA epitope tag; Ap^R	Clontech
pGBKT7	Y2H vector to produce binding domain (BD) fusions, GAL4 $_{(1\text{-}147)}$ DNA-BD, TRP1, c-MYC epitope tag; Km^R	Clontech
pGBDDnaE1	pGBKT7 containing Mtb DnaE1 ORF fused to GAL4 BD	(Kana *et al.*, 2010)
pGBDDnaE2	pGBKT7 containing Mtb DnaE2 ORF fused to GAL4 BD	(Kana *et al.*, 2010)

pGADImuB	pGADT7 containing Mtb ImuB ORF fused to GAL4 AD	(Kana *et al.*, 2010)
pGBDDnaE2	pGBKT7 containing Msm DnaE2 ORF fused to GAL4 BD	This study
pGBDDnaE2$^{\Delta SRDFR}$	pGBKT7 containing Msm DnaE2 ORF, with the C-terminal pentapeptide sequence SRDFR truncated, fused to GAL4 BD	This study
pGBDDnaE2^{N38}	pGBKT7 containing Msm DnaE2 ORF, with the 38 N-terminal amino acids truncated, fused to GAL4 BD	This study

Table 3.3 Primers used in this study

Name	Sequence (5'-3')	Restriction site	Application	Region targeted/ location in the gene
*dnaE1*WTF	tcagcgtgtgttagcg**gaattc**ggagggcgtcg tttgccacagcagatagtt**atcgat**gcaatagct atgccata	*Eco*RI and *Cla*I	Forward primer used to amplify full length *dnaE*1	4475 bp amplicon, primer located 810 bp upstream of the *dnaE1* gene
*dnaE1*WTR	tccgctgcgcgagaagct		Reverse primer used to amplify full length *dnaE*1	4475 bp amplicon, primer located 107 bp downstream of the *dnaE1* gene
*dnaQ*KOFP1	tgtcaccgcgcggccgtat		Forward primer used to amplify a region downstream the *dnaQ* gene	2019 bp amplicon downstream full length *dnaQ*
*dnaQ*KORP1	accgctggtgcgtgacgacg**agatct**tgcgtct gctcgaccat	*Bgl*II	Reverse primer used to amplify a region downstream *dnaQ*	2019 bp amplicon downstream full length *dnaQ*
*dnaQ*KOFP2	aggatggcacgc**agatct**gacaagccgggtcg gcacgcca	*Bgl*II	Forward primer used to amplify a region upstream the *dnaQ* gene	1965 bp amplicon upstream full length *dnaQ*
*dnaQ*KORP2	ttcgacctcggg**aagctt**gagcgggggcagat	*Hin*dIII	Reverse primer used to amplify a region upstream *dnaQ*	1965 bp amplicon upstream full length *dnaQ*
*att*BS1	acgtggcggtccctaccg		Reverse primer used to confirm site-specific recombination at the bacterial attachment (*attB)* site	Hybrid primer which amplifies both the vector and phage attachment sites.
*att*BS2	acaggatttgaacctgcggc		Forward primer used to confirm site-specific recombination at the *attB* site	Hybrid primer which amplifies both the vector and phage attachment sites.
*att*L2	cttggatcctcccgctgcgc		Forward primer used to confirm site-specific recombination at the *attB* site	Hybrid primer which amplifies both the vector and phage attachment sites.
*att*L4	aattcttgcagacccctgga		Reverse primer used to confirm site-specific recombination at the *attB* site	Hybrid primer which amplifies both the vector and phage attachment sites.

*dnaQ*F	gcttgtcggagccgcgtgccat		Forward primer used to amplify full length *dnaQ*	1140 bp amplicon, primer located 30 bp upstream *dnaQ*
*dnaQ*R	ccaggagatcgccgacgccaa		Reverse primer used to amplify full length *dnaQ*	1140 bp amplicon, primer located 97 bp downstream *dnaQ*
*dnaQ-uvrC*F	cgggcagcctctgctgcatt		Forward primer used to amplify full length *dnaQ-uvrC*	2130 bp amplicon, primer located 141 bp upstream *dnaQ*
*dnaQ-uvrC*R	ccgtgccgtgctcggcggtgccaa		Reverse primer used to amplify full length *dnaQ-uvrC*	2130 bp amplicon, primer located 85 bp downstream *dnaQ*

3.2 DNA manipulations

3.2.1 Plasmid and genomic DNA extraction

3.2.1.1 *E. coli* small scale plasmid isolation

A single colony was inoculated into 3 ml of LB containing the appropriate antibiotic, in a 5 ml tube, which was incubated either overnight at 37°C with shaking or standing at 30°C for 48 h, to avoid plasmid rearrangement. Cells were harvested by centrifugation at 15700 × *g* using the Eppendorf 5415 D centrifuge (Merck) for 1 min. The pellet was re-suspended in 100 µl of Solution I (50 mM glucose, 25 mM Tris-hydrochloride, 10 mM ethylenediaminetetraacetic acid (EDTA), pH 8.0) followed by 200 µl of Solution II (0.2 M sodium hydroxide, 1% sodium dodecyl sulphate (SDS)) and mixed by gentle inversion of the tube. This was followed by addition of 150 µl of Solution III (5M potassium acetate, 11.5% glacial acetic acid) into the cell suspension, vigorous shaking and centrifugation at 15700 × *g* for 10 min to remove cellular debris. The supernatant was collected into a fresh sterile 1.5 ml Eppendorf tube and 350 µl of isopropanol was added to precipitate the DNA. The DNA was harvested by centrifugation at 15700 × *g* for 10 min at room temperature. The DNA pellet was washed with 200 µl of 100% ethanol and vacuum-dried for 20 min in the vacuum-dryer (GeneVac, Thermo Scientific). The DNA pellet was then re-suspended in 20 µl of sterile RNase-free water (Qiagen).

3.2.1.2 *E. coli* large scale plasmid isolation

A single colony was inoculated into 100 ml of LB broth supplemented with the appropriate antibiotic in a 250 ml Erlenmeyer flask. Cells were grown with gentle shaking at 37°C overnight or standing at 30°C for 48 h. Cells were then transferred into two 50 ml Falcon tubes and harvested by centrifugation at 3901 × *g* (Beckman Coutler Allegra X-22R 4250 rotor) for 10 min. Pellets were re-suspended in 1 ml of Solution I, followed by 2 ml of Solution II and inverted gently for mixing. This was followed by addition of 1.5 ml of Solution III and vigorous shaking. The suspension was aliquoted into 1.5 ml Eppendorf tubes and

centrifuged for 10 min at 15700 × *g*. The supernatant was transferred into a fresh 1.5 ml Eppendorf tube and 3 µl RNaseA (10 mg/ml stock) was added to each tube, followed by incubation at 37°C for 1 h. Following incubation, 700 µl of isopropanol was added to each tube to precipitate the DNA and centrifuged for 10 min at 15700 × *g*. The DNA pellet was washed with 500 µl of cold 70% ethanol and dried in the vacuum-dryer to remove excess ethanol. The DNA pellet from each tube was re-suspended in 100 µl of sterile autoclaved water (dH_2O) and pooled to a volume of 500 µl. Following this, 50 µl of 3M sodium acetate (pH 5.2) and 700 µl of phenol:chloroform (1:1) was added to the DNA mixture. The mixture was vortexed and centrifuged at 15700 × *g* for 10 min. The aqueous layer was transferred into a clean 1.5 ml Eppendorf tube, 350 µl of chloroform:isoamyl alcohol (24:1) was added to the aqueous solution and the mixture was vortexed and centrifuged at 15700 × *g* for 10 min. Following this, the aqueous layer was transferred into a clean 1.5 ml Eppendorf tube and 1 ml of cold 100% ethanol was added to the mixture. To precipitate the DNA, the tube containing the mixture was gently inverted several times and stored at -20°C for 1 h before centrifugation at room temperature for 25 min. The DNA pellet was washed with 70% cold ethanol, vacuum-dried and re-suspended 50-100 µl of sterile RNase free water. The DNA was quantified using a NanoDrop 2000c Spectrophotometer (Thermo Scientific) according to the manufacturer's instructions.

3.2.1.3 Genomic DNA extraction from Msm

Mycobacterial genomic DNA was extracted using the modified cetyltrimethylammonium bromide (CTAB, ICN Biomedicals, Aurora, Ohio) method described by Larsen (Larsen, 2000). Briefly, mycobacterial cells were harvested and re-suspended in 500 µl of TE buffer (10 mM Tris-hydrochloride and 0.1 mM EDTA, pH 8.0). Cells were heat-killed at 65°C for 20 min, followed by centrifugation at 15700 × *g* for 5 min. Cells were re-suspended in 500 µl of TE buffer and 50 µl lysozyme (10 mg/ml) was added and this mixture was incubated at 37°C overnight. This was followed by addition of 70 µl 10% SDS and 6 µl of proteinase K (10 mg/ml) and the mixture incubated at 65°C, shaking at 400 rpm in a Eppendorf thermomixer (Merck) for 2 h. Whilst the mixture was still at 65°C, 100 µl of pre-warmed 5 M NaCl and CTAB/NaCl mix (10% CTAB in 0.5 M NaCl chloride) was added to the mixture and incubated

for 15 min. Following incubation, the mixture was stored at -80°C for 15 min, allowed to thaw first before re-incubation at 65°C for a further 15 min. The mixture was allowed to cool first before addition of 700 µl chloroform:isoamyl alcohol (24:1) and inverted several times until formation of a homogeneous white solution. Samples were centrifuged at 15700 × *g* for 10 min, followed by the removal of the aqueous phase containing the DNA into fresh microcentrifuge tubes. The DNA pellet was precipitated by addition of 700 µl of isopropanol and incubation at 4°C overnight. This was followed by centrifugation at 4°C for 10 min and the DNA pellet washed with 70% cold ethanol. The DNA pellet was vacuum-dried and re-suspended in 50 µl of sterile RNase free water.

3.2.2 Polymerase Chain Reaction (PCR)

For amplification of genes to be used for cloning purposes, a high-fidelity polymerase, Phusion® (Finnzymes, NEB) was used, whereas a low-fidelity polymerase, Faststart Taq (Roche Applied Science), was used for preliminary PCRs and screening purposes. Reactions were performed according to manufacturers' instructions. For reactions using the NEB Phusion kit, the standard 50 µl PCR reaction consisted of 1× HF buffer, 200 µM of each dNTP [dATP, dCTP, dGTP and dTTP (Roche Applied Science)], 1× GC rich buffer (Roche Applied Science), 0.5 µM of each primer (IDT, Whitehead Scientific (Pty) Ltd), 1.5 µl DMSO and 2.5 U of the DNA polymerase. DNA amplification conditions were as follows: initial denaturation at 98°C for 5 min followed by 30 cycles of denaturation at 98°C for 1 min, annealing at 60°C for 1 min, extension at 72°C for 30 s/kbp and final extension at 72°C for 7 min. For the Roche Faststart Taq kit, the standard 50 µl PCR reaction consisted of 1× FastStart PCR buffer with $MgCl_2$ (Roche Applied Science), 200 µM of each dNTP, 1× GC rich buffer, 0.5 µM of each primer, 1.5 µl DMSO, 2.5 U of the DNA polymerase and 2 µl of the template DNA. DNA amplification conditions were as follows: initial denaturation at 95°C for 3 min followed by 30 cycles of denaturation at 95°C for 30s min, annealing at 60°C for 1 min, extension at 72°C for 1 min/kbp and final extension at 72°C for 7 min. The PCR reaction mixture without the template DNA and dH_2O were used as negative controls to test for contaminating chromosomal DNA and non-specific fragments. All the PCRs were performed using MyCycler™ or C1000™ Thermal cyclers (Bio-Rad).

3.2.3 Agarose gel electrophoresis

Standard gel electrophoretic techniques were used for separation of DNA fragments (Sambrook *et al.*, 1989). High molecular weight (500 bp–12 kbp) DNA fragments were separated in 0.8-1% agarose gels and low molecular weight DNA fragments of ≤500 bp were separated in 2-4% agarose gels. Agarose gels were prepared in 1×TAE buffer (1 mM EDTA, 40 mM Tris-acetic acid, pH 8.0) and contained 0.5 µg/ml ethidium bromide for detection. DNA samples were loaded with a loading dye (0.025% bromophenol blue in 30% glycerol) for visual tracking of DNA migration during electrophoresis. DNA molecular weight markers (III, IV and V; Roche Applied Science) were used to determine DNA fragment sizes. Agarose gels were electrophoresed between 80–100V in a Mini-Sub® or Sub-Cell® GT Cell horizontal electrophoresis system (Bio-Rad). Fragments to be cloned were visualized using the blue-light Dark Reader DR88M transilluminator (Inqaba) to avoid UV–induced DNA damage; otherwise gels were visualized using the Wealtec Keta UV transilluminator (Inqaba).

3.2.4 Purification of DNA from agarose gels and PCR reactions

The DNA fragment of interest was excised from the gel and purified using the Nucleospin kit (Macherey-Nagel) as per manufacturer's instructions. Briefly, the excised agarose gel containing the fragment of interest was incubated at 60°C in a 2 ml Eppendorf tube containing the DNA-binding buffer (Nucleospin, Macherey-Nagel). After all the agarose gel had solubilized, the sample was loaded onto a Nucleospin column (Macherey-Nagel) to bind the DNA. For purification of PCR products, 1× volume of the PCR mixture was mixed with 2× volume of the DNA-binding buffer and the sample loaded onto a Nucleospin column to bind the DNA. The column containing the DNA was washed twice with the wash buffer (Nucleospin, Macherey-Nagel) to remove excess salts. The DNA was then eluted using pre-warmed sterile RNase-free water. The DNA was quantified either on agarose gels or using the NanoDrop.

3.2.5 DNA sequencing

DNA was sequenced at the Central Analytical Facilities (CAF) DNA Sequencing Unit at Stellenbosch University. Sequencing data were visualized and analysed using the CLC DNA Workbench 6 software (http://www.clcbio.com/products/clc-main-workbench).

3.2.6 DNA manipulation for cloning

3.2.6.1 Digestion of DNA with restriction enzyme(s)

Enzymes were obtained from Fermentas, New England Biolabs (NEB), Amersham or Roche Applied Science and used as per manufacturer's instructions. Up to 5 µg plasmid DNA and 1 µg PCR product was digested in a 20 µl reaction volume for 4-24 h. Digested samples were incubated at the recommended temperature for maximal enzyme activity for at least 4 h. For double digestions, an appropriate buffer in which both restriction enzymes had maximal activity was selected; otherwise, the digestions were performed sequentially. DNA fragments were then separated and analysed on agarose gels by electrophoresis (**section 3.2.3**).

3.2.6.2 DNA phosphorylation

Phosphorylation of blunt-ended PCR products to enable ligation with de-phosphorylated vector was performed for 30 min at 37°C using polynucleotide kinase (Roche Applied Science), as per the manufacturer's instruction. Reactions were stopped by inactivation at 65°C and the DNA was purified using the PCR clean-up protocol (**section 3.2.4**).

3.2.6.3 DNA dephosphorylation

Following digestion from plasmid DNA, the vector was treated with Antarctic Alkaline Phosphatase (NEB) to remove the 5'-phosphate group from the linearized vector DNA to prevent vector re-ligation. Dephosphorylation reactions were carried out at 37°C for 1h

followed by inactivation at 65°C for 20 min. The DNA was purified using the PCR clean-up protocol (**section 3.2.4**) as per manufacturer's instructions.

3.2.6.4 DNA ligation

DNA ligations were performed using either the Fast-Link™ ligation kit (Epicentre® Biotechnologies) or T4 DNA Ligase (Roche Applied Science), as per manufacturer's instructions. Blunt-end ligations were performed at 4°C and sticky-end ligations were performed at room temperature.

3.2.7 Transformation of bacterial cells

3.2.7.1 Chemical transformation of *E. coli* cells

Rubidium chloride-treated *E. coli* DH5α cells were used for transformation of plasmid DNA into *E. coli* cells. Cells were treated with rubidium chloride as follows: One millilitre from the stationary phase culture was used to inoculate 100 ml of LB broth and grown at 37°C to an OD_{600} = 0.48. Cells were kept on ice for 15 min and harvested by centrifugation at 3901 × *g* for 5 min at 4°C. Pellets were re-suspended in 20 ml of transformation buffer (Tfb) I solution (30 mM potassium acetate; 100 mM rubidium chloride; 10 mM calcium chloride; 50 mM manganese chloride; 15% v/v glycerol; pH 5.8) and stored on ice for a further 15 min. Cells were harvested by centrifugation at 3901 × *g* for 5 min at 4°C and re-suspended in 2 ml Tfb II solution (10 mM MOPS; 75 mM calcium chloride; 10 mM rubidium chloride; 15% glycerol; pH 6.5). Following this, 500 µl aliquots of cells were stored on ice for 15 min and either used immediately or quick-frozen in ethanol dry ice at -80°C, prior to storage at -80°C until further use. For transformation, chemically competent DH5α cells were thawed on ice and 100 µl was used per transformation. Up to 1 µg of plasmid DNA or 10 µl of the ligation mixture was incubated with the cells on ice for 30 min. Following incubation, cells were heat shocked at 42°C for 90s and incubated on ice for 1 min to facilitate the uptake of plasmid DNA. The transformation mixture was rescued in 400 µl of 2×TY (0.16% tryptone, 0.10%

yeast extract and 0.05% sodium chloride) broth and incubated at 37°C for 1 h. Following incubation, cells were plated on LA media containing the appropriate antibiotic and incubated overnight at 37°C or at 30°C for 48 h.

3.2.7.2 Electroporation into Msm

Msm electroporations were carried out as described by Larsen (Larsen, 2000). Briefly, a single colony of Msm was used to inoculate a pre-culture of 5 ml 7H9 OADC and grown overnight at 37°C. A volume of 1 ml of the pre-culture was inoculated into 100 ml 7H9 OADC and grown at 37°C until mid-log phase (OD_{600} of 0.6-1.0). Cells were harvested by centrifugation at 2600 × *g* for 10 min at 4°C and the pellet washed three times by gentle re-suspension in 40, 30 and 20 ml of ice-cold 10% glycerol, respectively. Cells were harvested by centrifugation at 2600 × *g* for 10 min at 4°C. The pellet was re-suspended in 1 ml of ice-cold 10% glycerol. Up to 4 µg of plasmid DNA was added to 400 µl of Msm competent cells. The mixture was transferred into a 0.2 cm pre-chilled electroporation cuvette and pulsed using the GenePulser Xcell™ (Bio-Rad) at the following settings: 2.5 kV, 1000 Ω resistance, 25 µF capacitance and 2 mm path length. Cells were rescued immediately with 800 µl of 2×TY for 4 h, shaking at 37°C before plating on 7H10 OADC media supplemented with appropriate antibiotics. Plates were incubated for 7 days at 37°C before scoring for growth of potential recombinants.

3.2.7.3 Lithium acetate-mediated yeast transformation

A single colony from the YPDA plate was inoculated into 5 ml YPD broth and vigorously vortexed to disperse the clumps, followed by incubation at 30°C with shaking for 48 h. A volume of 2 ml of the pre-culture was used to inoculate 100 ml YPD broth and incubated at 30°C for 16–18 h with shaking at 250 rpm until the culture reached stationary phase (OD_{600} >1.5). This culture was diluted 1:10 to an OD_{600} =0.2-0.3 and grown for two generations until an OD_{600} =0.6-0.8 was reached. Cells were harvested at 2600 × *g* for 5 min at room temperature. Pellets were re-suspended in 10 ml sterile RNase-free water and pooled into a 50 ml tube. Cells were then harvested at 2600 × *g* for 5 min at room temperature. The pellet

was re-suspended in 1.5 ml freshly made 1×TE/1×LiAc solution (10 mM Tris-hydrochloride, 1 mM EDTA, pH 8.0 plus 0.1 M lithium acetate, pH 7.5). The mixture was incubated for 1 h at 30°C with gentle agitation. Plasmid DNA (0.1 µg) and 0.1 mg of herring sperm carrier DNA (Invitrogen) were added to a fresh 1.5 ml tube and mixed thoroughly by passing them up and down through a needle. Following incubation, 100 µl of the yeast suspension was added to the tube containing the transforming and carrier DNA. This was incubated at 30°C for 30 min. This was followed by addition of 600 µl of sterile PEG/LiAc solution (40% polyethylene glycol, 10 mM Tris-hydrochloride, 1 mM EDTA, pH 8.0 plus 0.1 M lithium acetate, pH 7.5) and further incubation at 30°C for 30 min with gentle agitation. The yeast-DNA mixture was heat-shocked at 42°C for 15 min then chilled on ice for 1 min. Cells were harvested for 5 s at 15700 × *g* and re-suspended in 500 µl TE buffer. This step was repeated twice followed by final re-suspension in 200 µl TE buffer. The suspension was plated on SD agar to select for the desired transformants and the plates incubated at 30°C for 3-5 days.

3.3 Generation of *dnaE1* constructs for complement switching

The essentiality of *dnaE1* complicates genetic approaches to elucidate structure-function-relationships in the encoded protein. A set of *dnaE1* constructs that could be used for both allelic exchange and complement switching methods were generated. Complement switching involves the deletion of a chromosomal copy of an essential gene in the presence of a second copy of the gene (complementing allele) integrated at the mycobacterial *attB* locus. Integration of a cassette expressing the protein of interest and containing a particular antibiotic resistance gene into the *attB* site creates a merodiploid strain. The native copy of the gene is knocked out by allelic exchange resulting in one copy of the gene integrated at the *attB* locus. Mutant alleles are introduced on alternative vectors containing different antibiotic markers (Williams *et al.*, 2011).

Fragments containing WT *dnaE1* and targeted *dnaE1* point mutations were initially cloned into pMC1r, respectively to generate constructs for the complement switching method (details shown below). In parallel to this, the same fragments were excised from representative pMC1r constructs and subsequently cloned into the suicide vector, p2NIL, for the allelic exchange method **(Figure 3.1)** (Parish and Stoker, 2000), which enabled for

generation of the *dna*E1 point mutants quicker and was then used throughout to generate all the *dnaE1* mutants (**Figure 3.1**) and the *dnaQ* knock-out mutant (details shown below).

pMC1rE1: The sequence of the WT Msm *dnaE1* gene was obtained by BLAST homology search (Altschul *et al.*, 1990) of the preliminary Msm genome sequence database at The Institute for Genomic Research (TIGR; http://www.tigr.org) using its respective Mtb gene sequence (TubercuList; http://www.pasteur.fr/Bio/TubercuList) as a probe. The primers, *dnaE1*WTF and *dnaE1*WTR (**Table 3.3),** were designed to amplify the full length *dnaE1* from WT Msm mc^2155 genomic DNA with 800 bp upstream and 107 bp downstream of *dnaE1*. The forward primer contained an engineered *Eco*RI restriction site which enabled cloning into the integrative mycobacterial shuttle vector, pMC1r. Both vector and full length *dnaE1* PCR product were digested with *Eco*RI and ligated to generate pMC1rE1. This construct was used as the site-directed mutagenesis (SDM) template to generate the *dnaE1* mutants.

pMC1rE1^{D49A}: The synthesized suicide vector pIDT*dnaE1*^{D49A}, containing the aspartate to alanine mutation at position 49 in DnaE1, was digested with *Eco*RI and *Mfe*I to release the 700 bp fragment containing the targeted mutation. This fragment was swapped into pMC1rE1 using the same restriction sites. The D49A mutation introduced an additional *Nco*I site in *dnaE1* and this was used to distinguish WT *dnaE1* from *dnaE1*^{D49A}.

pMC1rE1^{E133A}: A similar strategy was used to generate the pMC1rE1^{E133A} construct. A 1500 bp fragment containing the glutamate to alanine mutation at position 133 in DnaE1, from the synthesized suicide vector pIDT*dnaE1*^{E133A}, was swapped into pMC1rE1 using the native *Eco*RV and *Mfe*I sites in *dnaE1.* The E133A mutation introduced an additional *Msc*I site in *dnaE1* which was used to distinguish WT *dnaE1* from *dnaE1*^{E133A}.

pMC1rE1^{D228N}: A 1500 bp fragment containing the aspartate to asparagine mutation at position 228 in DnaE1, from the synthesized suicide vector pIDT*dnaE1*^{D228N}, was swapped into pMC1rE1 using the native *Eco*RV and *Pvu*II sites in *dnaE1.* The D228N mutation introduced an additional *Hind*III site in *dnaE1* which was used to distinguish WT *dnaE1* from *dnaE1*^{D228N}.

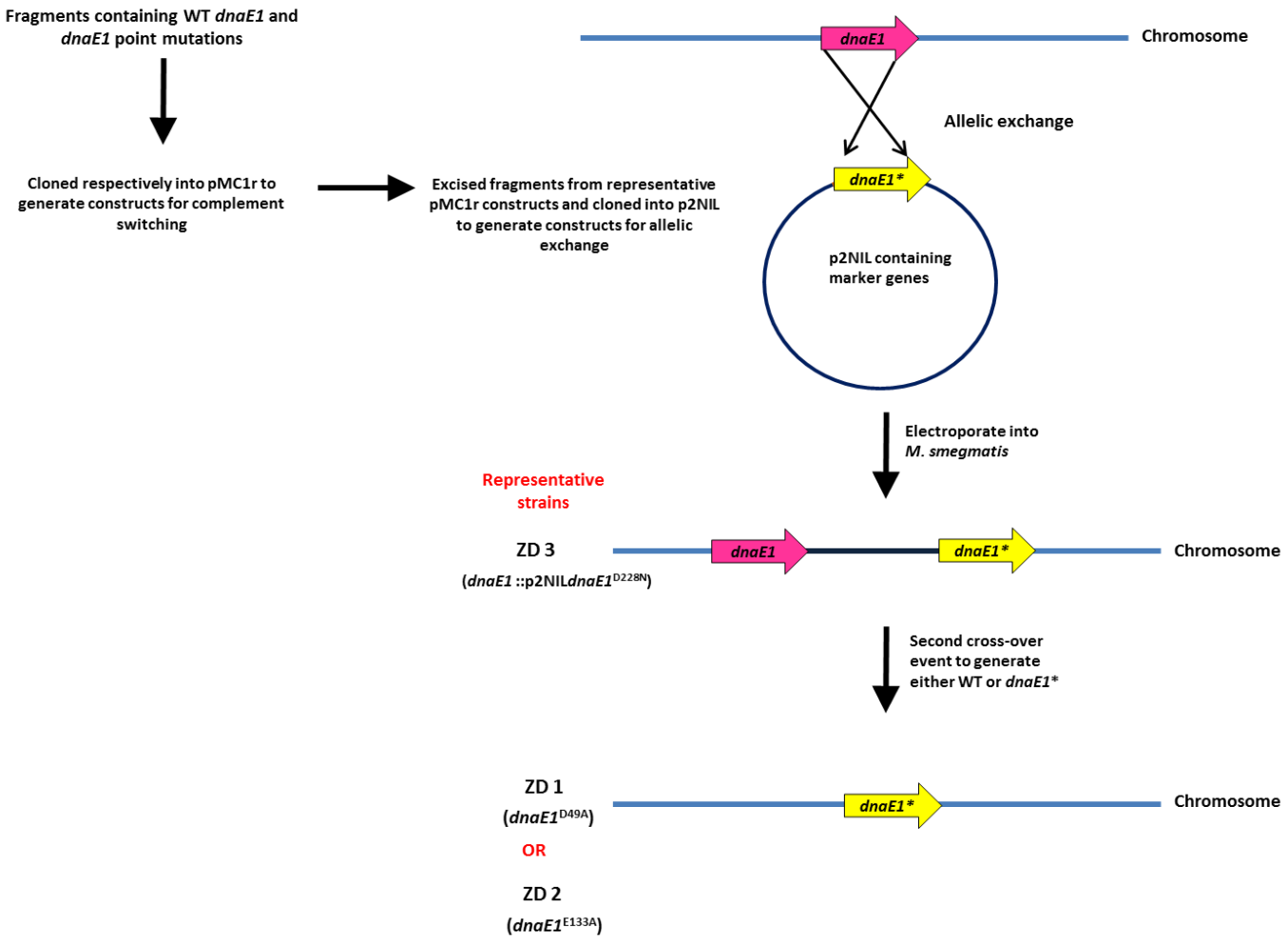


Figure 3.1: Generation of the *dnaE1* alleles. *dnaE1* constructs that could be used for both allelic exchange and complement switching methods were generated in pMC1r and then sequentially cloned into p2NIL for the allelic exchange method which was used throughout to generate all the *dnaE1* mutants**.**

3.4 Construction of unmarked suicide delivery plasmids for use in allelic exchange mutagenesis

p2NILE1, p2NILE1^{D49A}, p2NILE1^{E133A} and p2NILE1^{D228N}: The 4470 bp fragments containing WT *dnaE1, dnaE1*^{D49A}*, dnaE1*^{E133A} *and dnaE1*^{D228N} were excised as an *Eco*RI fragment from pMC1rE1, pMC1rE1^{D49A}, pMC1rE1^{E133A} and pMC1rE1^{D228N} and cloned separately into p2NIL using the same restriction site. The 7939 bp *Pac*I cassette from pGOAL19 - containing a Hyg resistance gene (*hyg*), as well as the *lacZ* and *sacB* genes to enable blue/white and sucrose counterselection, respectively (Parish and Stoker, 2000), was cloned into each construct generating unmarked p2NILE1, p2NILE1^{D49A}, p2NILE1^{E133A} and p2NILE1^{D228N}, respectively .

p2NILE1^{E133A}KI: A 1500 bp fragment containing WT *dnaE1* was swapped into pMC1rE1^{E133A} using the native *Eco*RV and *Mfe*I sites in *dnaE1.* This regenerated the WT *dnaE1* sequence in this background. The *Msc*I site was used for screening between p2NILE1^{E133A} and p2NILE1^{E133A}KI to verify loss of this site in the knock-in construct. The *Pac*I cassette from pGOAL 19 (*hyg, lacZ and sacB)* (Parish and Stoker, 2000) was cloned into this construct, generating unmarked p2NILE1^{E133A}KI.

All the p2NILE1 constructs containing mutations in *dnaE1*, including the knock-in construct, were verified by restriction enzyme analysis and sequencing at CAF, University of Stellenbosch.

p2NILΔ*dnaQ*: The sequence of the WT Msm *dnaQ* (MSMEG_6275) gene was obtained by BLAST homology search (Altschul *et al.*, 1990) of the preliminary Msm genome sequence database at The Institute for Genomic Research (TIGR; http://www.tigr.org) using its respective Mtb gene sequence (TubercuList; http://www.pasteur.fr/Bio/TubercuList) as a probe. Two primer sets d*naQ*KOFP1/*dnaQ*KORP1 and *dnaQ*KOFP2/*dnaQ*KORP2 were used to amplify sequences upstream and downstream of the Msm *dnaQ* gene. The two PCR products were ligated and this resulted in the elimination of 990 bp *dnaQ* gene. The ligated PCR product was then cloned into p2NIL. The *Pac*I cassette from pGOAL 19 (*hyg, lacZ* and *sacB*) (Parish and Stoker, 2000), was cloned into this construct, generating the unmarked p2NILΔ*dnaQ* construct, respectively.

3.5 Construction of unmarked *dnaE1* and *dnaQ* mutants using allelic exchange

The allelic exchange method was used throughout to generate the *dnaE1* and *dnaQ* mutants in Msm (Parish and Stoker, 2000) (**Figure 3.1).** Briefly, WT mc^2155 was electroporated with p2NILE1, p2NILE1^{D49A}, p2NILE1^{E133A}, p2NILE1^{D228N} and p2NILΔ*dnaQ* separately and plated on 7H10 OADC containing Km, Hyg, and X-gal. Five putative single cross-over (SCO) recombinants per construct that were Km^R, Hyg^R and blue in colour were selected and grown in liquid 7H9 OADC containing no antibiotic, to allow for the second cross-over event to occur. Serial dilutions of each culture were plated on 7H10 medium containing 2% sucrose and X-gal to enable phenotypic selection of putative double-cross over (DCO) recombinant mutants (Parish and Stoker, 2000). White, sucrose-resistant colonies were screened by PCR using *dnaE1* internal primers to confirm the genotype. Genomic DNA was extracted using the crude method of boiling the colony at 65°C to lyse the mycobacterial wall, followed by precipitation of DNA using chloroform. The genomic DNA obtained was used as a template for PCR. An internal fragment in *dnaE1* of ~2000 bp covering the region where the point mutations were introduced was amplified and screened by restriction enzyme analysis to verify the mutants relative to WT.

3.6 Construction of *dnaE2* constructs

pMC1r*dnaE2*: The multicopy *E. coli–Mycobacterium* shuttle vector carrying full length *dnaE2*, pOmsmE2 (Boshoff *et al.*, 2003) was digested with *Eco*RI to release the full length WT *dnaE2*. This fragment was cloned into pMC1r using the same restriction site, to generate pMC1r*dnaE2*. This construct was verified by restriction enzyme analysis and PCR amplification of the bacterial *att* locus to confirm integration (see **Appendix B**). This construct was also used as a template for generating the *dnaE2*$^{\Delta SRDFR}$ and *dnaE2*^{N38} constructs and for complementation.

pMC1r*dnaE2*$^{\Delta SRDFR}$: The synthesized suicide vector, pIDT*dnaE2*$^{\Delta SRDFR}$, containing Msm *dnaE2* with the C-terminal SRDFR motif truncated, was digested with *Srf*I and *Hind*III to release a fragment containing the targeted truncation and an additional *Bgl*II silent mutation, for screening purposes. This fragment was swapped into pMC1r*dnaE2* using the same restriction sites and confirmed by restriction enzyme analysis.

pMC1r*dnaE2*^{N38}: The synthesized suicide vector, pIDT*dnaE2*^{N38}, containing Msm *dnaE2* with 38 N-terminal amino acids truncated, was digested with *Acv*I and *Srf*I to release a fragment containing the targeted truncation and an additional *Spe*I silent mutation for screening purposes. This fragment was swapped into pMC1r*dnaE2* using the same restriction sites and confirmed by restriction enzyme analysis.

3.7 Construction of *dnaE2* mutants

Following verification, WT mc^{2}155 was separately electroporated with pMC1r*dnaE2*, pMC1r*dnaE2*$^{\Delta SRDFR}$ and pMC1r*dnaE2*^{N38} and plated into 7H10 OADC containing 20µg/ml Km. This resulted into integration at the *attB* site which was verified by the PCR strategy shown in **Appendix B**. The *att*B primers amplify the bacterial attachment site which is 276 bp long and the *att*L primers amplify the phage attachment site (vector sequence) which is 326 bp long. Site specific recombination of *att*B and *att*P creates the flanking *att*L and *att*R sites. These sites contain hybrid portions originating from *att*B and *att*P which are used as a marker for the recombination event. Following homologous recombination, screening with a combination of either primer set (attBS2 and attL4 or attL2 and attBS1) results in a 320 bp

or a 282 bp fragment as shown in **Appendix B**. If recombination is not successful neither combination will give a product.

3.8 Construction of vectors for Y2H assays

Full length WT *dnaE2*, *dnaE2*$^{\Delta SRDFR}$ and *dnaE2*^{N38} fragments were digested from pMC1r*dnaE2*, pMC1r*dnaE2*$^{\Delta SRDFR}$ and pMC1r*dnaE2*^{N38} using *EcoR*I and cloned in-fusion with the GAL4 BD of the GAL4 transcription regulator in pGBKT7, using the same restriction site, to generate pGBDDnaE2, pGBDDnaE2$^{\Delta SRDFR}$ and pGBDDnaE2^{N38} respectively. Correct clones were confirmed by restriction enzyme analysis. pGADImuB and pGBDDnaE1 were generated as described (Kana *et al.*, 2010, Warner *et al.*, 2010).

3.9 Microbiological characterization of the *dnaE1*, *dnaQ* and *dnaE2* strains

Assays used were common across the *dnaE1*, *dnaQ* and *dnaE2* strains. However, the Luria-Delbruck assay was not performed for *dnaE2* mutants relative to the WT strain. Details of the methods are shown below.

3.9.1 Mutation rate (μ) assessment using the Luria-Delbrück fluctuation assay

Rates of spontaneous mutations to rifampicin were assessed for *dnaE1* and *dnaQ* strains relative to WT mc^{2}155 using the Luria-Delbrück fluctuation assay and analysed according to Rosche and Foster (Rosche and Foster, 2000). Briefly, a single colony of each Msm strain was used to inoculate a pre-culture of 7H9 OADC, which was grown till mid-log phase (OD_{600}= 0.6-1.0). This culture was used to inoculate 7H9 OADC at a concentration of ~100 colony forming units per millilitre (CFU/ml). Aliquots were plated on standard 7H10 OADC plates for enumeration of CFU in the initial culture (N_0) as well as on 7H10 OADC plates containing 200 μg/ml Rif (Sigma, R3501) to detect any pre-existing Rif-resistant mutants in the starting culture. Approximately 3 ml of the starting culture was aliquoted into 20 parallel 15 ml culture tubes for each strain. The tubes were incubated for 3 days at 37°C shaking until mid-

log phase (OD_{600}= 0.6-1). Cultures were serially diluted in un-supplemented 7H9 media containing 0.5% Tween 80, vortexed and plated on standard 7H10 OADC to determine the final number of cells per culture (N_t). Undiluted cultures were plated on 7H10 OADC plates containing 200 µg/ml Rif to determine the number of Rif-resistant mutants. The mutation rate and 95% confidence intervals were calculated using the software developed in the MMRU by Z. Martin and J. Kent (unpublished), optimized from the MSS method as described by Rosche and Foster (Rosche and Foster, 2000).

3.9.2 DNA damage assays

To determine damage tolerance of the *dnaE1*, *dnaQ* and *dnaE2* mutants, mitomycin C (MMC) sensitivity and UV-induced mutagenesis assays were performed as described by Boshoff *et al.* (Boshoff *et al.*, 2003). Briefly, mycobacterial cultures were grown until mid-log phase (OD_{600} = 0.6-1.0). Log_{10}-fold dilutions (10^{-1} – 10^{-7}) of each culture were spotted on 7H10 OADC solid media without and with varying concentrations (0.02 and 0.04 µg/ml) of MMC. A volume of 1 ml culture for each strain was serially diluted and used for CFU enumeration to ensure that the total number of cells per culture was comparable across all strains. The remaining cultures were used to set up UV-induced mutagenesis assays as described below.

Briefly, 1 ml of each culture was serially diluted and used for CFU enumeration before UV treatment. The remaining cells were harvested and re-suspended in 5 ml aliquots. Cultures were then transferred into open petri dishes and irradiated at 25 mJ/cm^2 (Spectrolinker XL-1000, Spectroline). After irradiation, cells were rescued in fresh 45 ml 7H9 OADC media at 37°C for 4 h. After incubation, 1 ml of the irradiated culture was plated on 7H10 OADC solid media without antibiotic for CFU enumeration and 1 ml was plated on 7H10 OADC plates containing 200 µg/ml rifampicin $(Rif)^{200}$ for determination of Rif-resistant mutants. The frequency of Rif-resistant mutants was determined as described (Boshoff *et al.*, 2003).

3.10 Y2H assays

Y2H assays were used to determine the role for the N-terminal extension and C-terminal motif in DnaE2 in mediating protein-protein interactions during function of the mutasome. Protein-protein interactions were assessed using the Clontech Matchmaker® Y2H system, as per the manufacturer's instructions. Genes encoding proteins of interest were cloned in-frame with the GAL4 AD and BD of a GAL4 transcription factor, and then co-transformed into the yeast strain, AH109. Fusion of ImuB to the BD of a GAL4 transcription factor has been shown to result in auto-activation, therefore giving a false positive signal (Warner *et al.*, 2010). ImuB was therefore cloned as a fusion protein to the GAL4 AD of the transcription factor to determine the interactions between ImuB and the DnaE2 mutants. WT DnaE1 and DnaE2 as well as DnaE2 mutants were cloned in as BD fusion proteins. Yeast strains carrying plasmid to be assessed were grown in SD-LT broth to an OD_{600} of 0.20 and then spotted onto different media with increasing stringency to identify interacting partners (Warner *et al.*, 2010).

CHAPTER 4: RESULTS

4.1 Investigating the role of conserved PHP domain residues in replication fidelity in mycobacteria

4.1.1 The PHP domain of mycobacterial DnaE1 contains a complete set of metal co-ordinating residues

The role of the PHP domain in bacterial DNA PolIIIα subunits remains unclear. The PHP domains of *Thermus thermophilus* and *T. aquaticus* PolIIIα subunits, which have complete sets of metal co-ordinating residues [H H D/H H E H C/H D/N H], possess exonuclease proofreading activity (Stano *et al.*, 2006, Wing *et al.*, 2008). The presence of an intrinsic exonuclease in these *dnaE*-type polymerases resembles similar observations for the PHP domains of PolX from both *B. subtilis* (Banos *et al.*, 2008) and *T. thermophilus* (Nakane *et al.*, 2009). In contrast, no exonuclease activity could be detected for the PHP domain of *G. kaustophilus* PolC (Evans *et al.*, 2008) in which an asparagine residue substitutes for the metal-binding aspartate at position 8 (Barros *et al.*, 2013). This implies that a complete set of metal co-ordinating residues is required for exonuclease proofreading activity. In *E. coli*, the PHP domain of the sole DnaE-type PolIIIα subunit lacks most of the conserved metal co-ordinating residues, and does not exhibit exonuclease activity in enzyme assays *in vitro* (Barros *et al.*, 2013). It is notable, though, that introduction of the complete set of residues into *E. coli* DnaE does not enable exonuclease function *in vitro*, suggesting that an additional factor is required (Barros *et al.*, 2013). Instead, it appears that the E. coli PHP domain might be required to stabilize the structure of the PolIII core.

To determine whether the *dnaE*-encoded α subunit in Mtb contains an intact PHP domain, protein sequences from representative mycobacterial DnaE subunits were aligned against the sequences of PolIIIα subunits from *E. coli* (Barros *et al.*, 2013, Lamers *et al.*, 2006) and *T. aquaticus* (Bailey *et al.*, 2006) (**Figure 4.1**). As reported previously (Barros *et al.*, 2013), *E. coli* PolIII contains a variant PHP domain in which five of the metal co-ordinating residues are replaced by residues incompatible for this function (H14R, H48F, E73D, C158G and H228G). Mtb DnaE1 contains a complete set of metal co-ordinating residues [H14, H16, D23,

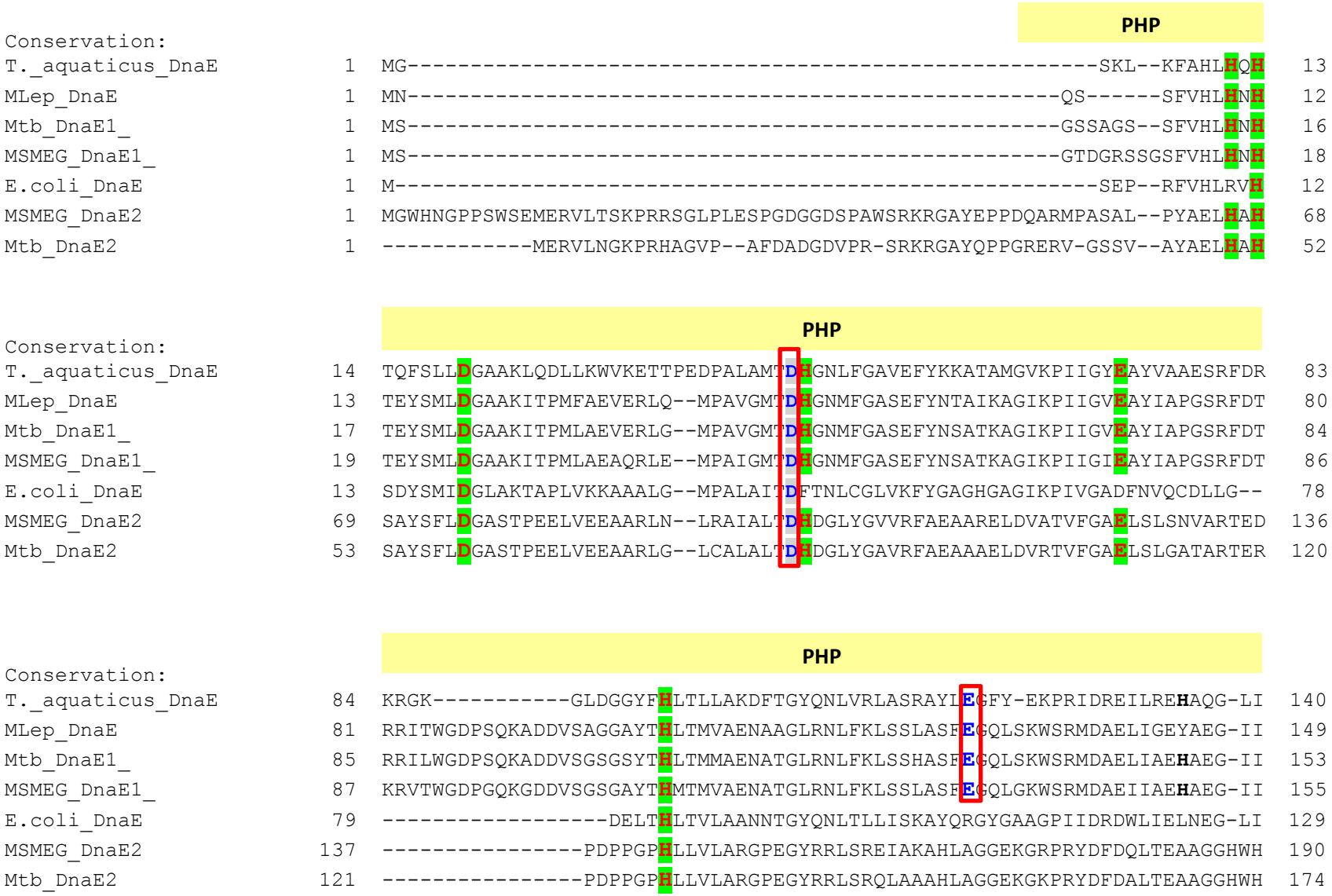
H48, E73, H107, C158, D226 and H228]. In contrast, an aspartate to glycine substitution (D228G) in DnaE2 (**Figures 4.1** and **S1**) suggests that this α subunit is unlikely to possess intrinsic exonuclease function. Analysis of sequenced mycobacterial DnaE2 polymerases suggests that conservation of eight of the nine canonical metal co-ordinating residues is a characteristic of all mycobacteria (**Figure S1**), as suggested (Timinskas *et al.*, 2014). Both *M. leprae* and Msm DnaE1 homologs also possess the full nine residues (**Figure S1**), which was important for two reasons: (i) it confirmed the applicability of Msm as mycobacterial model for these assays and (ii) it reinforced the inferred likelihood that strong selection has maintained the DnaE1 PHP domain given the massive reductive evolution that characterises the *M. leprae* genome (Monot *et al.*, 2009).

4.1.2 Identification of PHP domain residues which might contribute to the differential fidelities of mycobacterial DnaE1 and DnaE2

Based on the sequence alignments, it was decided that this study should focus on the role of residues which are conserved in the PHP domain of the essential *dnaE1*-encoded DnaE-type polymerase in Msm but are absent in the corresponding domain of the non-essential DnaE2 (**Figures 4.1** and **S1**), specifically:

i. The glutamate at position 133 (E133) in Msm DnaE1, which was predicted to be required for exonuclease function based on observations with other intrinsic exonucleases (Barros *et al.*, 2013, Nakane *et al.*, 2009, Wieczorek and McHenry, 2006)
ii. The aspartate at position 228 (D228), which was selected based on studies of exonuclease function (Barros *et al.*, 2013, Nakane *et al.*, 2009, Wieczorek and McHenry, 2006) and, subsequently, was implicated in metal binding (Barros *et al.*, 2013)

In addition, given the key role of DnaQ in replisome function in model systems such as *E. coli*, as well as evidence implicating the PHP domain aspartate at position 49 (D49) in the DnaQ-DnaE1 protein-protein interaction (Wieczorek and McHenry, 2006), this residue was also investigated despite its conservation in both DnaE1 and DnaE2.

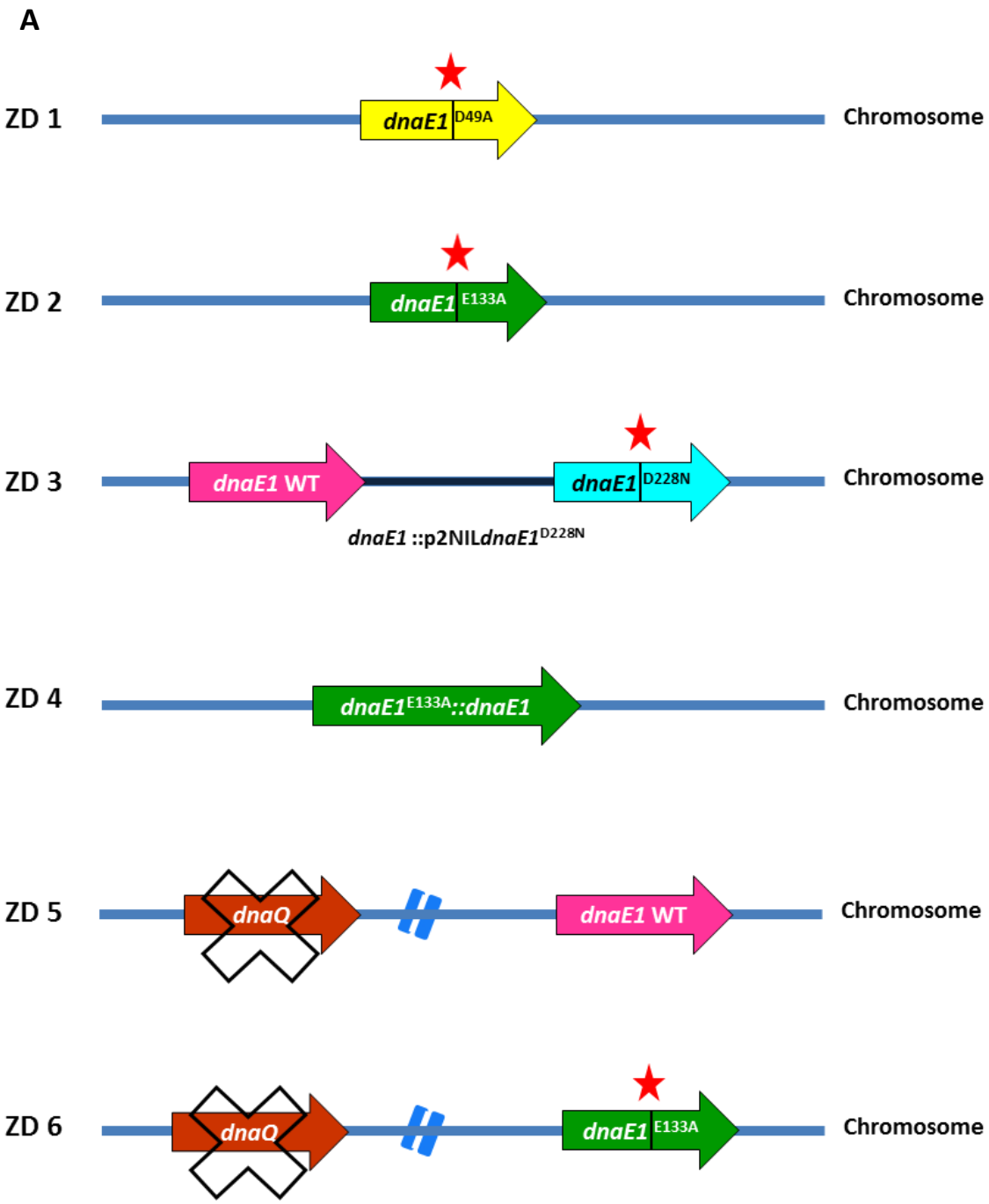

PHP
Conservation:
T._aquaticus_DnaE 1 MG----------SKL--KFAHLHQH 13
MLep_DnaE 1 MN----------QS------SFVHLHNH 12
Mtb_DnaE1_ 1 MS----------GSSAGS--SFVHLHNH 16
MSMEG_DnaE1_ 1 MS----------GTDGRSSGSFVHLHNH 18
E.coli_DnaE 1 M----------SEP--RFVHLRVH 12
MSMEG_DnaE2 1 MGWHNGPPSWSEMERVLTSKPRRSGLPLESPGDGGDSPAWSRKRGAYEPPDQARMPASAL--PYAELHAH 68
Mtb_DnaE2 1 ----------MERVLNGKPRHAGVP--AFDADGDVPR-SRKRGAYQPPGRERV-GSSV--AYAELHAH 52
PHP
Conservation:
T._aquaticus_DnaE 14 TQFSLLDGAAKLQDLLKWVKETTPEDPALAMTDHGNLFGAVEFYKKATAMGVKPIIGYEAYVAAESRFDR 83
MLep_DnaE 13 TEYSMLDGAAKITPMFAEVERLQ--MPAVGMTDHGNMFGASEFYNTAIKAGIKPIIGVEAYIAPGSRFDT 80
Mtb_DnaE1_ 17 TEYSMLDGAAKITPMLAEVERLG--MPAVGMTDHGNMFGASEFYNSATKAGIKPIIGVEAYIAPGSRFDT 84
MSMEG_DnaE1_ 19 TEYSMLDGAAKITPMLAEAQRLE--MPAIGMTDHGNMFGASEFYNSATKAGIKPIIGIEAYIAPGSRFDT 86
E.coli_DnaE 13 SDYSMIDGLAKTAPLVKKAAALG--MPALAITDFTNLCGLVKFYGAGHGAGIKPIVGADFNVQCDLLG-- 78
MSMEG_DnaE2 69 SAYSFLDGASTPEELVEEAARLN--LRAIALTDHDGLYGVVRFAEAARELDVATVFGAELSLSNVARTED 136
Mtb_DnaE2 53 SAYSFLDGASTPEELVEEAARLG--LCALALTDHDGLYGAVRFAEAAAELDVRTVFGAELSLGATARTER 120
PHP
Conservation:
T._aquaticus_DnaE 84 KRGK----------GLDGGYFHLTLLAKDFTGYQNLVRLASRAYLEGFY-EKPRIDREILREHAQG-LI 140
MLep_DnaE 81 RRITWGDPSQKADDVSAGGAYTHLTMVAENAAGLRNLFKLSSLASFEGQLSKWSRMDAELIGEYAEG-II 149
Mtb_DnaE1_ 85 RRILWGDPSQKADDVSGSGSYTHLTMMAENATGLRNLFKLSSHASFEGQLSKWSRMDAELIAEHAEG-II 153
MSMEG_DnaE1_ 87 KRVTWGDPGQKGDDVSGSGAYTHMTMVAENATGLRNLFKLSSLASFEGQLGKWSRMDAEIIAEHAEG-II 155
E.coli_DnaE 79 ----------DELTHLTVLAANNTGYQNLTLLISKAYQRGYGAAGPIIDRDWLIELNEG-LI 129
MSMEG_DnaE2 137 ----------PDPPGPHLLVLARGPEGYRRLSREIAKAHLAGGEKGRPRYDFDQLTEAAGGHWH 190
Mtb_DnaE2 121 ----------PDPPGPHLLVLARGPEGYRRLSRQLAAAHLAGGEKGKPRYDFDALTEAAGGHWH 174


PHP
Conservation:
T._aquaticus_DnaE 141 ALSGCLGAEIPQFILQDRLDLAEARLNEDLSIF-GDRFFIEIQNHGLPEQKKVNQVLKEFARKYGLGMVA 209
MLep_DnaE 150 VTTGCPSGEVQTRLRLGHDREALESAAKWREIVGPDNYFLELMDHGLSIEQRVREGLLNIGRKLNIPPLA 219
Mtb_DnaE1_ 154 ITTGCPSGEVQTRLRLGQDREALEAAAKWREIVGPDNYFLELMDHGLTIERRVRDGLLEIGRALNIPPLA 223
MSMEG_DnaE1_ 156 ATTGCPSGEVQTRLRLGHEREALEAAAKWREIFGPENFFLELMDHGLDIERRVREGLLEIGRKLGIPPLA 225
E.coli_DnaE 130 LLSGGRMGDVGRSLLRGNSALVDECVAFYEEHF-PDRYFLELIRTGRPDEESYLHAAVELAEARGLPVVA 198
MSMEG_DnaE2 191 ILTGCRKGHVRQALSDGGPDAAAAALADLVDRFGADRVSIELTHHGHPCDDERNAALAELAPRFGLGVVA 260
Mtb_DnaE2 175 ILTGCRKGHVRQALSQGGPAAAQRALADLVDRFTPSRVSIELTHHGHPLDDERNAALAGLAPRFGVGIVA 244
PHP
Conservation:
T._aquaticus_DnaE 210 TNDGHYVRKEDARAHEVLLAIQSKTTLDDPERWRFPCDEFYVKTPEEMRAMLPEAEWGDEPFDNTVEIAR 279
MLep_DnaE 220 TNDCHYVTRDAVHNHEALLCVQTGKTLSDPNRFKFHGDGYYLKSAAEMRQLWDDE--VPGACDSTLLIAE 287
Mtb_DnaE1_ 224 TNDCHYVTRDAAHNHEALLCVQTGKTLSDPNRFKFDGDGYYLKSAAEMRQIWDDE--VPGACDSTLLIAE 291
MSMEG_DnaE1_ 226 TNDCHYVTREAARNHEALLCVQTGKTLSDPTRFKFDGDGYFLKSAEEMRALWDSQ--VPGACDSTLLIGE 293
E.coli_DnaE 199 TNDVRFIDSSDFDAHEIRVAIHDGFTLDDPKRPRNYSPQQYMRSEEEMCELFAD---IPEALANTVEIAK 265
MSMEG_DnaE2 261 TTAAHFATPSRGRLAMAMAAIRARNSIDTAAGWLAPLGGVHLRSGEEMARLFD-----PEFVAAAADLGE 325
Mtb_DnaE2 245 TTGAHFADPSRGRLAMAMAAIRARRSLDSAAGWLAPLGGAHLRSGEEMARLFAWC---PEAVTAAAELGE 311
Palm
Conservation:
T._aquaticus_DnaE 280 MCDVDLPIGDKMVYRIPRFPLPEGRTEAQYLRELTFLGLLRRYPDRITEAFYREVLRLLGTMPPHGDERA 349
MLep_DnaE 288 RVQSYADVWEP-RNRMPVFPVPVGHDQASWLRHEVDAGLKRRFPDG---------- 332
Mtb_DnaE1_ 292 RVQSYADVWTP-RDRMPVFPVPDGHDQASWLRHEVDAGLRRRFPAG---------- 336
MSMEG_DnaE1_ 294 RVQSYADVWEP-RDRMPVFPVPEGHDQASWLTHEVKAGLERRFRGG---------- 338
E.coli_DnaE 266 RCNVTVRLGE---YFLPQFPTGD-MSTEDYLVKRAKEGLEERLAFLFPDEEERL---------- 315
MSMEG_DnaE2 326 QCAFGLALIA---PQLPPFDVPDGHTEDSWLRHLAMAGAARRYGPP---------- 368
Mtb_DnaE2 312 RCAFGLQLIA---PRLPPFDVPDGHTEDSWLRSLVMAGARERYGPP---------- 354

Figure 4.1: Multiple sequence alignment of the PHP domains of representative bacterial DnaE homologs. DnaE-type sequences from *T. aquaticus* (AAD44403), *M. leprae* (ML1207), Mtb [DnaE1 (Rv1547) and DnaE2 (Rv3370c)], Msm [DnaE1 (MSMEG_3718) and DnaE2 (MSMEG_1633)], and *E. coli* DnaE (AP_000844) were aligned using the PROMALS3D multiple sequence and structure alignment server (http://prodata.swmed.edu/promals3d/promals3d.php). (1) Bold red residues shaded in green are the conserved metal co-ordinating residues [H H D/H H E H C/H D/N H] (Barros *et al.*, 2013), (2) residues shaded in grey are predicted to bind DnaQ in *E. coli* (Vandewiele *et al.*, 1998, Borden *et al.*, 2002), (3) Bold blue residues inside boxes are conserved in essential DnaE-type polymerases and absent in non-essential DnaE-type polymerases, with the exception of D49. The annotation of the polymerase domains is derived from crystal structures of *E. coli* (Lamers *et al.*, 2006) and *T. aquaticus* (Bailey *et al.*, 2006) PolIIIα subunits.

4.1.3 Construction of site-directed *dnaE1* mutants

To investigate the contribution of the conserved PHP domain residues to the fidelity of the *dnaE1*-encoded α subunit, mutant Msm strains containing targeted substitutions (D49A, E133A and D228N) in DnaE1 were generated by allelic exchange (**Figure 3.1**) (Parish and Stoker, 2000). The genotype (**Figure 4.2A**) of each mutant was verified by restriction enzyme analysis of a 2075 bp PCR amplicon spanning the region containing the targeted *dnaE1* point mutations (**Figure 4.2B**). This result was confirmed by Sanger sequencing of the PCR product. Attempts to generate a *dnaE1*^{D228N} DCO were not successful: a total of 180 potential DCOs were screened by restriction enzyme analysis and PCR sequencing, however, all contained the WT genotype. Although formal proof is required, this result suggested that the D228N mutation is not compatible with optimal growth *in vitro*. Therefore, the merodiploid SCO mutant (ZD 3) containing both WT and *dnaE1*^{D228N} mutant alleles at the native *dnaE1* chromosomal locus was utilized in all subsequent experiments.


A
ZD 1
dnaE1 D49A
Chromosome
ZD 2
dnaE1 E133A
Chromosome
ZD 3
dnaE1 WT
dnaE1 D228N
Chromosome
dnaE1 ::p2NILdnaE1D228N
ZD 4
dnaE1E133A::dnaE1
Chromosome
ZD 5
dnaQ
dnaE1 WT
Chromosome
ZD 6
dnaQ
dnaE1 E133A
Chromosome

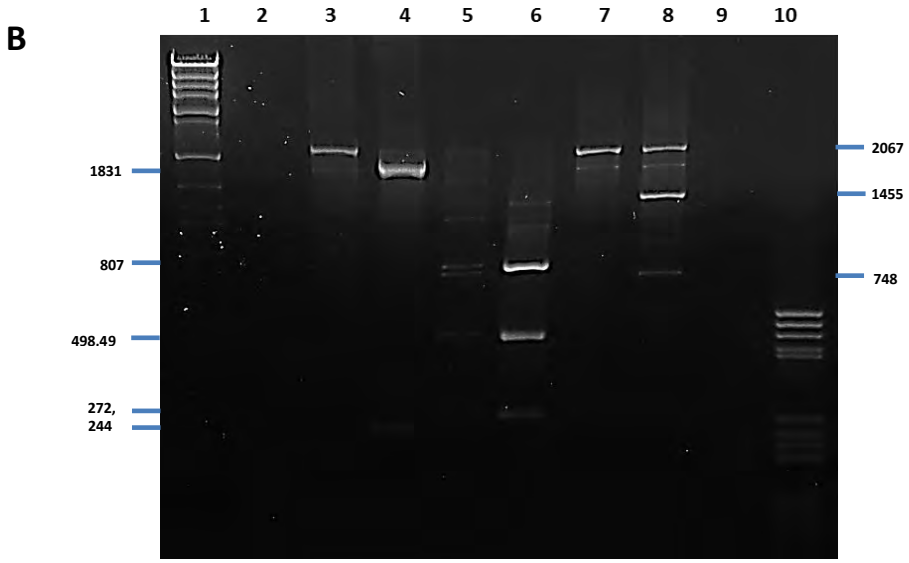


Figure 4.2: Genotypic confirmation of site-directed *dnaE1* mutants. (**A**) Schematic diagram detailing the allelic exchange method used to generate the *dnaE1* mutants containing targeted point mutations in selected PHP domain residues. The asterisks denote mutations in specific residues in the *dnaE1* locus. The black cross depicts deletion of the *dnaQ* gene (B) Chromosomal DNA was amplified by PCR from putative *dnaE1* mutants (ZD 1, ZD 2 and ZD 3) containing site-directed mutations in PHP domain residues, and PCR products digested with *Nco*I (ZD 1, lane 4), *Msc*I (ZD 2, lane 6) and *Hin*dIII (ZD 3, lane 8), and analysed on a 2% agarose gel (**B**). There is no natural *Nco*I site in WT *dnaE1*, therefore amplification of mc^{2}155 genomic DNA yields a 2075 bp PCR product that is not subject to *Nco*I restriction (lane 3); the engineered D49A mutation introduces an *Nco*I site in the *dnaE1*^{D49A} allele which results in 1831 bp and 244 bp fragments after digestion of the PCR product (lane 4). Digestion of the PCR product from WT mc^{2}155 with *Msc*I results in 807 bp, 766 bp and 494 bp fragments, whereas the engineered E133A mutation introduces an additional *Msc*I site in the *dnaE1*^{E133A} allele which results in 766 bp, 498 bp and 268 bp fragments after digestion of the PCR product, respectively (lanes 5 and 6). There is no natural *Hin*dIII site in WT *dnaE1*, therefore digestion of the PCR product from the ZD 3 containing both WT and mutant allele results in 2067 bp (from WT) and 1455 bp and 620 bp fragments from the engineered D228N allele which introduces an additional *Hin*dIII site (lanes 7 and 8). The PCR product amplified from WT mc^{2}155 was included as a control (lanes 3, 5 and 7). Lane 1, molecular weight marker IV (Roche); lane 3, WT mc^{2}155 digested with *Nco*I; lane 4, ZD 1 digested with *Nco*I; lane 5, WT mc^{2}155 digested with *Msc*I; lane 6, ZD 2 digested with *Msc*I; lane 7, WT mc^{2}155 digested with *Hin*dIII; lane 8, ZD 3 digested with *Hin*dIII; and lane 10, molecular weight marker V (Roche).

4.1.4 Conserved PHP domain residues are required for replication fidelity

To determine whether conserved residues in the PHP domain in DnaE1 are required for replication fidelity, spontaneous mutation rates (μ) were assessed for the ZD 1, ZD 2 and ZD 3 mutants by fluctuation assay (Rosche and Foster, 2000) using Rif-resistance as the readout. A total of 20 cultures was used to set up each fluctuation assay, and spontaneous Rif-resistant mutants enumerated on solid medium containing 200 μg/ml Rif. There was no difference in the mutation rates of the ZD 1 mutant versus WT ($p>0.05$) (**Figure 4.3A**). In contrast, mean 3-fold and 10-fold ($p<0.05$) increases, respectively, were observed in the spontaneous mutation rates of the ZD 2 and ZD 3 mutants. To confirm that the apparent loss in replication fidelity was attributable to the *dnaE1*^{E133A} mutation alone, the wild-type *dnaE1* allele was restored in the ZD 2 background by allelic exchange. The mutation rate of the resulting "knock-in" strain (ZD 4) was the same as WT (**Figure 4.3A)**, thereby establishing that the mutator phenotype of ZD 2 arose as a consequence of the E133A substitution in the DnaE1 PHP domain.

In performing the fluctuation assays, it became apparent that slightly different mutation rates were observed for the same strain each time the experiment was performed (*i.e.*, intra-strain variation). This variation in experimentally determined mutation rates is an intrinsic feature of fluctuation analyses (Rosche and Foster, 2000) and, while not significant in magnitude, it can complicate attempts to determine the reproducibility of observed mutation rate differences between strains in serial independent experiments. To address this concern, the fold-change relative to WT was calculated for each experiment according to the formula:

$$\boldsymbol{mutation\ rate}\ (\mu)\ \boldsymbol{of\ the\ mutant\ strain} \Big/ \boldsymbol{mutation\ rate}\ (\mu)\ \boldsymbol{of\ the\ WT\ strain}$$

Notably, by plotting fold-difference in the mutation rates of each allelic exchange mutant relative to the wild-type parental strain in independent experiments, the reproducibility of

inferred mutation rate changes in the different mutant strains was confirmed (**Figure 4.3B**). That is, replacement of the wild-type PolIIIα subunit in Msm with a DnaE1^{E133A} mutant resulted in a 3-fold increase in the spontaneous mutation rate. This was a significant result, since it provided the first evidence of a small mutator effect consequent on impaired DnaE1 function in mycobacteria. Similarly, the presence in the merodiploid ZD3 mutant of both wild-type and *dnaE1*^{D228N} alleles was associated with a 10-fold increase in the mutation rate. Given that the error-prone DnaE2 subunit contains a natural substitution (D228G) at this site, this result reinforced the possibility that the differential intrinsic fidelities of the mycobacterial PolIIIα subunits might be explained, in part, by the elimination of proofreading exonuclease function in DnaE2.

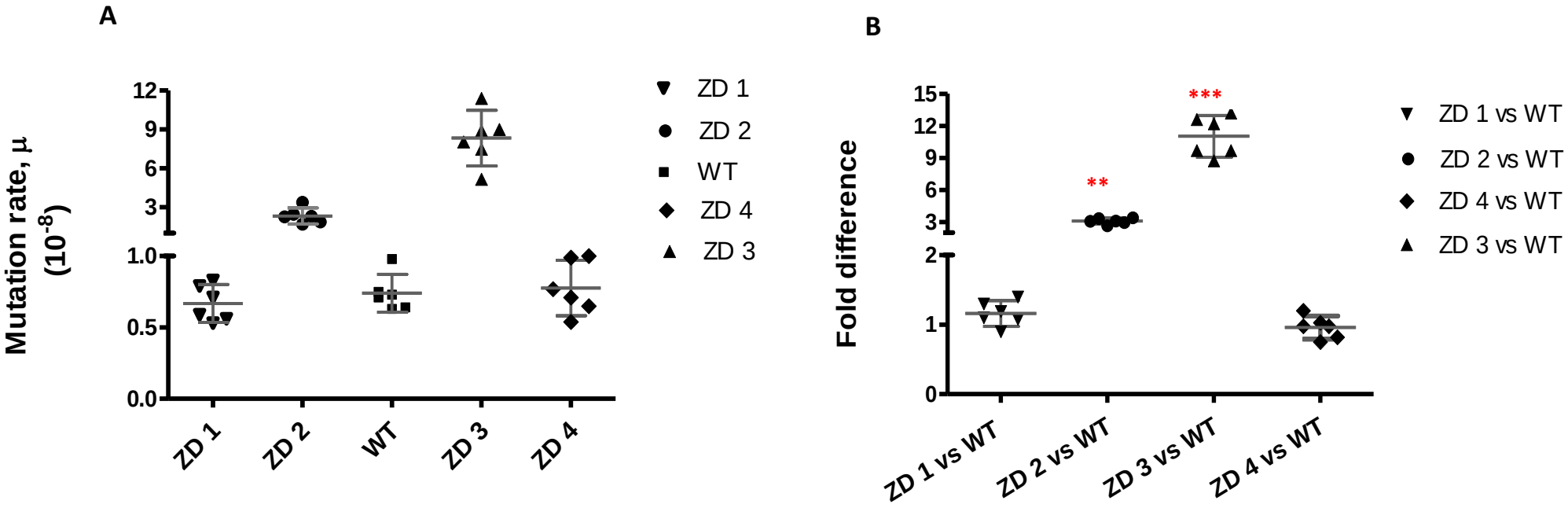


Figure 4.3: Mutation of conserved residues in the PHP domain in DnaE1 results in increased spontaneous mutation rates. Fluctuation assays were used to measure spontaneous rates of acquiring Rif-resistance in mutant strains carrying targeted amino acid substitutions in the putative DnaQ-binding residue, D49A (ZD 1), and residues that are conserved in essential DnaE-type polymerases, E133A (ZD 2) and D228N (ZD 3). Data represent mutation rates (μ) (**A**) and fold differences (**B**) for six independent experiments. Error bars represent the 95% confidence interval. Significance was determined by comparing strain pairs using the Mann-Whitney U test. **p<0.01; ***p<0.001

4.1.5 The mycobacterial proofreading DnaQ subunit does not contribute to DNA replication fidelity

The observation that mutation of the D49 residue in DnaE1 did not impact the mutation rate (**Figure 4.3)** could be interpreted as suggesting that this residue was not required to mediate the interaction with DnaQ. However, the role of DnaQ in mycobacterial replication fidelity had not been investigated at the time of this study. Therefore, while strong mutator phenotypes have been reported in *E. coli* strains deficient in DnaQ-mediated proofreading function (Fijalkowska and Schaaper, 1995, Horiuchi *et al.*, 1978, Degnen and Cox, 1974, Fijalkowska and Schaaper, 1996), the impact of impaired DnaQ-dependent proofreading function on replication fidelity in mycobacteria was unknown. To investigate the role of the mycobacterial proofreading DnaQ subunit homolog, allelic exchange mutagenesis was used to construct a targeted Msm Δ*dnaQ* mutant by in-frame elimination of the gene encoding the Msm DnaQ homolog, *MSMEG_6275*. In addition, evidence from homology modelling had suggested that the DnaE1 PHP domain residue was located in the DnaQ-binding region (Česlovas Venclovas, personal communication). Therefore, a second Msm mutant was constructed in parallel in which *dnaQ* was deleted in the ZD 2 (*dnaE1*^{E133A}) background, yielding the Δ*dnaQ dnaE1*^{E133A} double mutant (ZD 6). The genotypes of both mutants were confirmed by PCR and restriction enzyme analysis.

To determine the impact of abrogated DnaQ function on mycobacterial replication fidelity, spontaneous mutation rates were assessed in both Δ*dnaQ* and Δ*dnaQ dnaE1*^{E133A} (ZD 6) relative to WT. As before, a total of 20 cultures was used to set up each fluctuation assay, and Rif-resistant mutants were isolated on medium containing 200 µg/ml Rif. A very small (1.4-fold; $p<0.05$), but reproducible, increase in mutation rate was calculated for Δ*dnaQ* (**Figure 4.4**), indicating that the mycobacterial *dnaQ* homolog does not contribute significantly to replication fidelity *in vitro*. This observation was surprising, and contrasted sharply with the very well-characterised models of replisome organization and function in other bacteria. In addition, a mean 4.3-fold increase in mutation rate was observed for the Δ*dnaQ dnaE1*^{E133A} double mutant (ZD 6) compared with WT (**Figure 4.4**). This value appeared to be an additive effect of the mutation rates observed for the Δ*dnaQ* (1.4-fold) and *dnaE1*^{E133A} (3-fold) alleles, and suggested that the roles (albeit small) of the intrinsic PHP

domain and the extrinsic DnaQ subunit in maintaining replication fidelity were non-redundant.

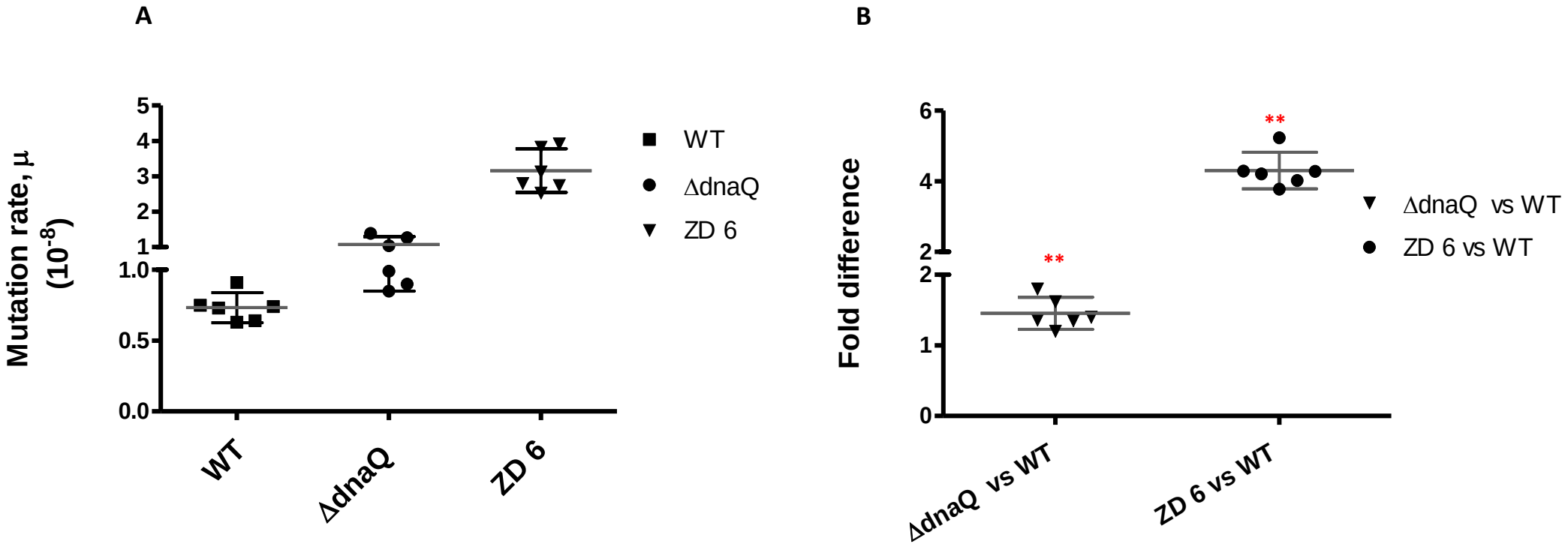


Figure 4.4: DnaQ is not required for proofreading in mycobacteria. Fluctuation assays were used to measure rates of acquiring Rif-resistance in Δ*dnaQ* and ZD 6 strains. Data represent mutation rates (μ) (**A**) and fold differences (**B**) for six replicates. A total of 20 cultures was used to set up each fluctuation assay. Rif was used at a concentration of 200 μg/ml. Error bars represent the 95% confidence interval. Significance was determined by comparing strain pairs using the Mann-Whitney U test. **$p<0.01$

4.1.6 The conserved DnaE1 PHP domain residues are not required for UV-induced mutagenesis

In *E. coli*, disruptions of DnaQ-dependent proofreading activity enable PolIII-mediated TLS in the absence of specialist Y-family polymerases (Borden *et al.*, 2002, Vandewiele *et al.*, 1998). Moreover, in *S. pyogenes*, the essential DnaE subunit that catalyses TLS does not bind DnaQ (Bruck *et al.*, 2003). To determine whether substitution of conserved DnaE1 residues impacts replication fidelity following exposure to genotoxic stress, the mutation frequencies to Rif-resistance were measured following UV treatment of ZD 1 and ZD 2 mutants. WT Msm, Δ*dnaE2* and the *dnaE2* complementation mutant were used as controls since the dominant role of DnaE2 in UV-induced mutagenesis has been established (Boshoff *et al.*, 2003). Treatment of WT Msm with UV has been previously shown to increase the mutation frequency to Rif-resistance, whereas the *dnaE2* knockout mutant is expected to eliminate UV-induced mutagenesis. This UV-induced phenotype is reversed by complementation with the WT copy of *dnaE2* at the *attB* locus (Boshoff *et al.*, 2003). Consistent with previous findings (Boshoff *et al.*, 2003), deletion of *dnaE2* eliminated UV-induced mutagenesis and this phenotype was restored by integration of the WT copy of *dnaE2* at the *attB* locus. In contrast, no phenotype was observed in the ZD 1 and ZD 2 PHP domain mutants, which phenocopied the WT strain (**Figure S2**). Since the dominant role of DnaE2 in damage-induced mutagenesis (Boshoff *et al.*, 2003) may have masked the impact of mutations in the PHP domain of DnaE1, the ZD 1 and ZD 2 mutations were therefore transferred into a genetic background in which all *dinB* homologs and *dnaE2* had been deleted from the Msm mc^2155 genome (ΔL mutant; D. Warner, unpublished), to generate strains ZD 10 and ZD 11 (**Table 3.1**). However, even in this genetic background, which is depleted of genes encoding specialist DNA polymerases, a DNA damage phenotype was not associated with either *dnaE1*^{D49A} or *dnaE1*^{E133A} alleles (**Figure S2**).

4.1.7 Contribution of conserved DnaE1 PHP domain residues to damage tolerance

Prior work (Warner *et al.*, 2010) has established that deletion of *dnaE2* renders mycobacteria hypersensitive to treatment with MMC, a genotoxic agent that forms crosslinks between the complementary strands of DNA preventing their separation and

inhibiting DNA replication (Tomasz, 1995). To determine the impact on DNA damage tolerance of the targeted DnaE1 PHP domain mutations, ZD 1 and ZD 2 mutants were plated in 7H10 OADC without and with varying concentrations of MMC. Again, the total number of CFU per culture was calculated to ensure comparable starting values across all the strains tested. Consistent with previous observations (Boshoff *et al.*, 2003, Warner *et al.*, 2010), the *dnaE2* mutant was hypersensitive to MMC treatment and this phenotype was reversed by complementation. In contrast, no DNA damage phenotype was observed for the ZD 1 and ZD 2 mutants, which phenocopied the WT strain (**Figure S3**). In addition, no DNA damage phenotype was observed for the ZD 10 and ZD 11 mutants following exposure to MMC (**Figure S3**), eliminating the possibility that any contribution, albeit small, of the PHP residues was obscured by the presence in Msm of multiple specialist TLS polymerases.

4.1.8 Investigating the role of DnaQ: compiling an expanded panel of *dnaQ* mutants

During the course of this study, a report emerged (Ford, 2012) describing the results of an independent analysis of the contributions to Msm replication fidelity of the canonical Msm *dnaQ* homolog (*MSMEG_6275*) and a second, *dnaQ*-like homolog comprising DnaQ and UvrC domains on a single peptide, *MSMEG_4259* (*dnaQ-uvrC*). The panel of Msm *dnaQ* mutants (**Table 3.1**) comprising Δ*dnaQ*, Δ*dnaQ-uvrC* and Δ*dnaQ* Δ*dnaQ-uvrC* (Ford, 2012) was obtained from the author of that study for inclusion in DNA damage sensitivity and UV-induced mutagenesis assays. Prior to commencing the work, the genotypes of all mutants were confirmed by PCR utilizing primers designed to amplify full length *dnaQ* (*MSMEG_6275*) and *dnaQ-uvrC* (*MSMEG_4259*) (**Table 3.3**). Amplification of the WT mc^{2}155 and Δ*dnaQ-uvrC* mutants using the *dnaQ*-specific primer set resulted in a PCR product of 1140 bp for both strains, whereas the Δ*dnaQ* and Δ*dnaQ* Δ*dnaQ-uvrC* strains yield a PCR product of 152 bp (**Figure 4.5**). Amplification of the WT mc^{2}155 and Δ*dnaQ* with *dnaQ-uvrC* specific primers resulted in a PCR product of 2130 bp respectively, whereas the Δ*dnaQ-uvrC* and Δ*dnaQ* Δ*dnaQ-uvrC* strains yield a PCR product of 196 bp. In combination, these analyses confirmed that the strain designations reflected the underlying genotypes.

Since a separate *dnaQ* mutant had been generated independently in the current study, it was considered useful to compare both Δ*dnaQ* mutant strains in parallel in fluctuation analyses (to determine spontaneous mutation rates) and DNA damage assays (to elucidate a role, if any, in DNA damage tolerance and induced mutagenesis). The *dnaQ* mutants phenocopied each other in the fluctuation experiments (the *dnaQ* mutants were associated with a 1.4-fold loss in fidelity) as well as in preliminary DNA damage assays (details below, section 4.1.9). Therefore, for all remaining assays, the Δ*dnaQ* mutant generated in this study was used as experimental strain.

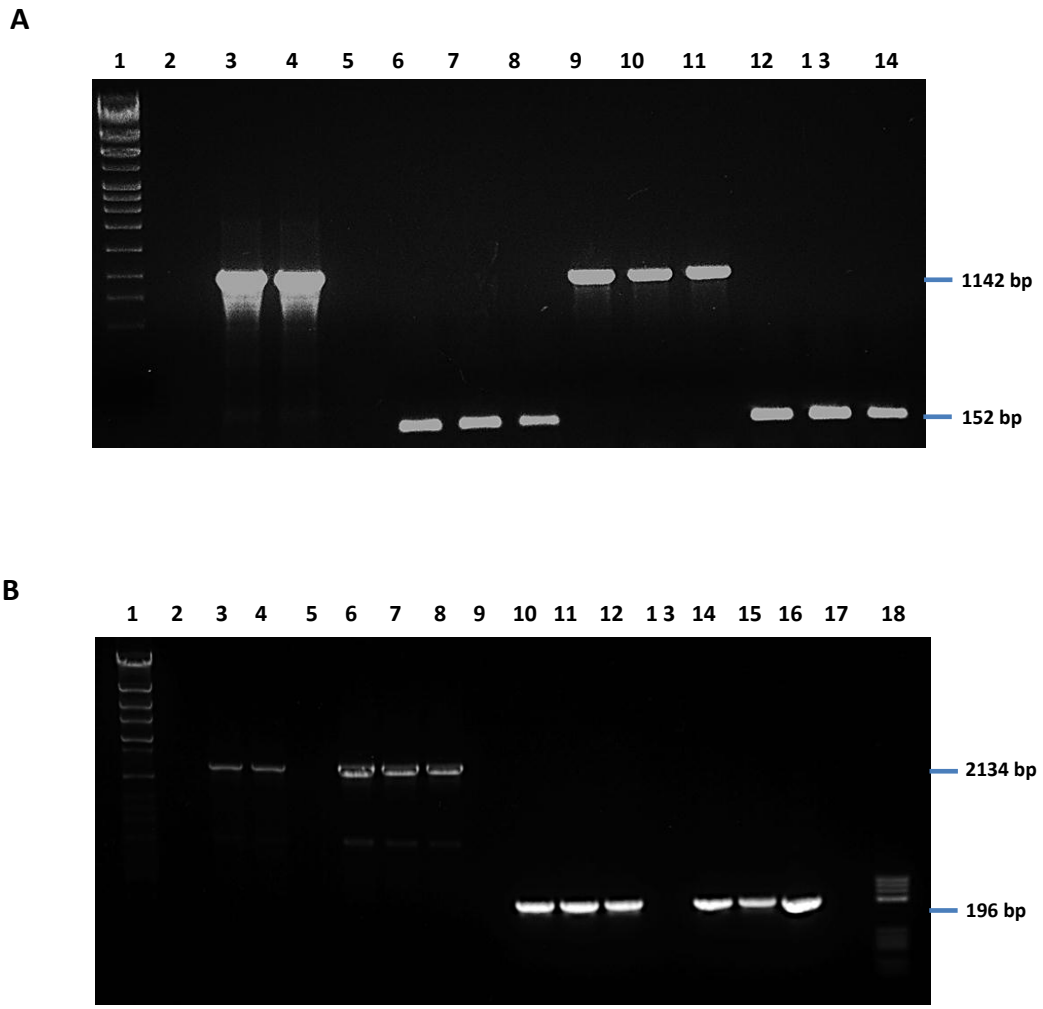


Figure 4.5: Genotypic confirmation of *dnaQ*, *dnaQ-uvrC* and *dnaQ dnaQ-uvrC* deletion mutants by PCR. WT (lanes 3 and 4), Δ*dnaQ* (lanes 6-8), Δ*dnaQ-uvrC* (lanes 10-12) and Δ*dnaQ* Δ*dnaQ-uvrC* (lanes 13- 16) strains were amplified using primer sets targeting the full-length *dnaQ* (**A**) and *dnaQ-uvrC* (**B**) alleles to verify deletion of these genes. Three colonies per strain were screened for the *dnaQ* mutants and two colonies were screened for the WT mc^2155 strain which was used as a positive control. Lane 1, molecular weight marker IV (Roche); lane 18, molecular weight marker V (Roche).

4.1.9 The role of the mycobacterial *dnaQ* homologs in UV-induced mutagenesis

To determine whether inactivation of *dnaQ* in Msm enhances UV-induced mutagenesis, as has been reported in *E. coli* (Borden *et al.*, 2002, Vandewiele *et al.*, 1998), the Δ*dnaQ,* Δ*dnaQ-uvrC* and Δ*dnaQ* Δ*dnaQ-uvrC* mutants were applied in UV-induced mutagenesis assays. Again, Δ*dnaE2* and the complemented *dnaE2* derivative were included as controls, and the total number of cells per culture was calculated to ensure comparable starting CFU across all strains tested. Consistent with previous observations (Boshoff *et al.*, 2003), a 50-100-fold decrease in mutation frequency was observed for the Δ*dnaE2* mutant relative to WT. The mutation frequencies to Rif-resistance for the Δ*dnaQ* and Δ*dnaQ-uvrC* mutants, as well as the Δ*dnaQ* Δ*dnaQ-uvrC* double knockout, were similar to that of WT Msm ($p>0.05$) suggesting that neither gene is required for UV-induced mutagenesis (**Figure 4.6**).

4.1.10 DnaQ-UvrC is required for DNA damage sensitivity

Finally, to assess the potential contribution of the DnaQ homologs to DNA damage tolerance, the panel of mutants was applied in MMC assays, again using the *dnaE2* deletion mutant as a negative control. Deletion of *dnaQ* did not increase damage sensitivity relative to WT. In contrast, the *dnaQ-uvrC* deletion mutant was hypersensitive to MMC treatment, as indicated by the 2-$\log_{10}$ reduction in CFU compared with WT (**Figure 4.7**). Surprisingly, the MMC-hypersensitivity phenotype was even more profound than that which was observed for the Δ*dnaE2* control. Moreover, damage hypersensitivity was not evident in the *dnaQ dnaQ-uvrC* double deletion mutant (**Figure 4.7**). In order to confirm that the single *dnaQ-uvrC* and double *dnaQ dnaQ-uvrC* mutants had not been inadvertently switched, the gentoypes of both strains were re-confirmed by PCR analysis. This result established that loss of DnaQ-UvrC renders Msm hypersusceptible to MMC-mediated DNA damage, and that this effect is reversed by deletion of *dnaQ* in the same background.

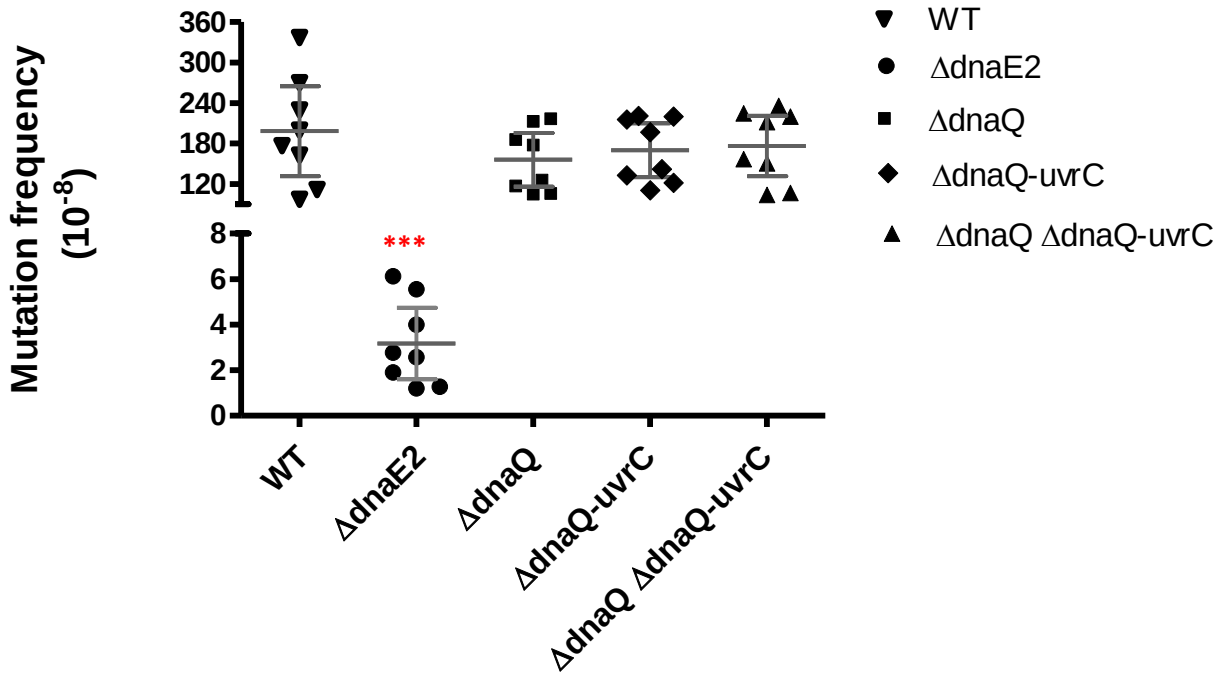


Figure 4.6: Msm *dnaQ* homologs are not required for UV-induced mutagenesis. UV-induced mutation frequencies of WT, Δ*dnaE2*, Δ*dnaQ*, Δ*dnaQ-uvrC* and Δ*dnaQ* Δ*dnaQ-uvrC* following UV-irradiation. Mycobacterial cultures of WT mc^2155 and mutant strains were grown until log phase harvested, re-suspended and UV irradiated at 25 mJ/cm^2. After irradiation, cells were rescued at 37°C for 4 h and plated on 7H10 OADC containing 200 µg/ml Rif. Error bars represent the 95% confidence interval. Significance was determined by comparing strain pairs using the Mann-Whitney U test. ***$p<0.001$

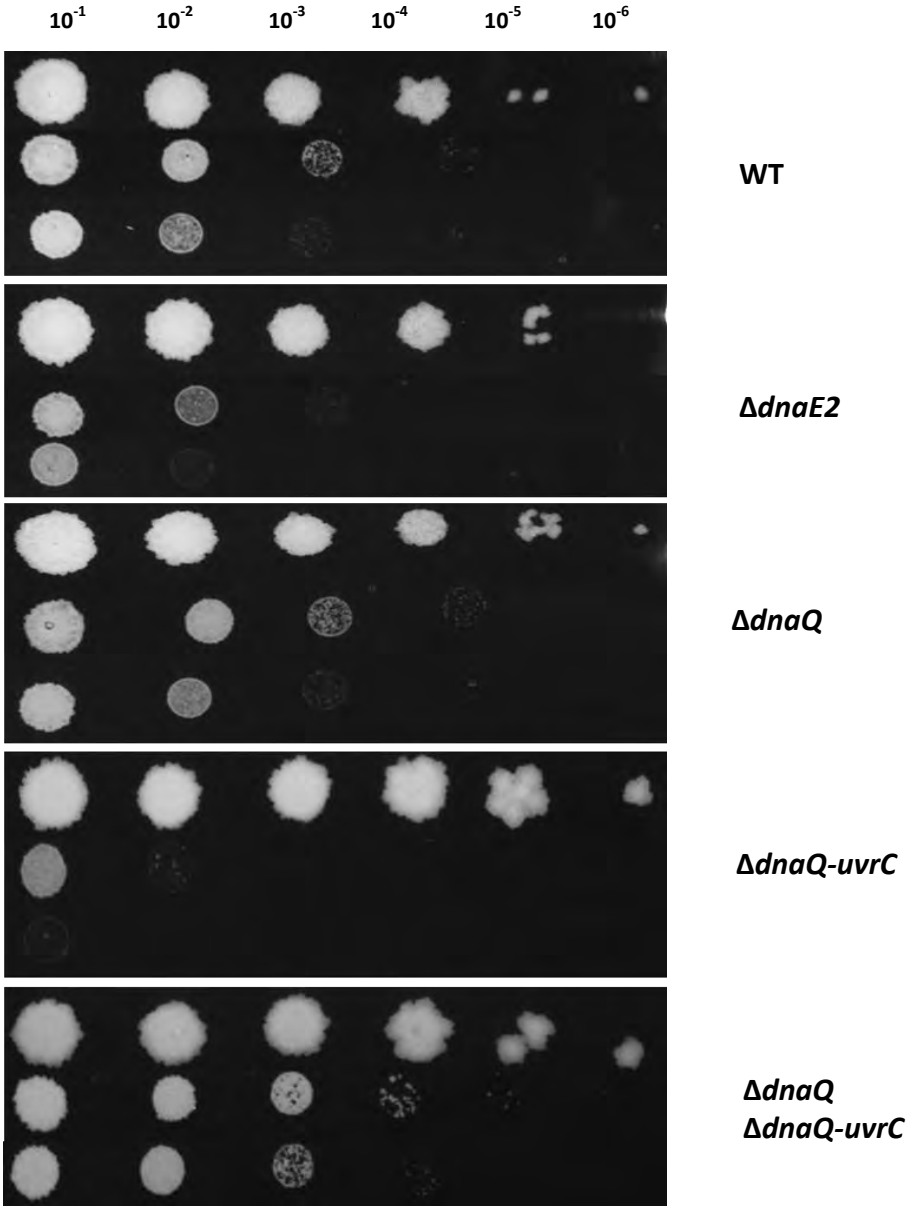


Figure 4.7: *dnaQ-uvrC* is required for DNA damage tolerance. Log_{10}-fold dilutions (10^{-1}–10^{-6}) of WT, Δ*dnaE2*, Δ*dnaQ*, Δ*dnaQ-uvrC* and Δ*dnaQ* Δ*dnaQ-uvrC* were spotted on standard 7H10 media without (first row for each strain) and with 0.02 µg/ml MMC (middle row for each strain) or 0.04 µg/ml MMC (last row for each strain). CFU determinations confirmed that all strains were spotted at equivalent cell numbers (not shown). Data are from a representative experiment performed in triplicate.

4.2 Determination of functional interactions that are essential for DnaE2-dependent induced mutagenesis in Mtb

The central role of DnaE2 in DNA damage survival and induced mutagenesis (Boshoff *et al.*, 2003) identified the mycobacterial PolIIIα subunit as the founder member of a novel family of DnaE-type family C polymerases from Gram-positive bacteria that catalyse TLS (Tippin *et al.*, 2004). Subsequent work then elaborated a model in which DnaE2 functions together with at least two accessory proteins, ImuA' and ImuB, both of which are upregulated together with DnaE2 in the mycobacterial SOS response (Warner *et al.*, 2010). The C-terminal domain of ImuB has been shown to be critical for the separate ImuB'-DnaE2 and ImuB-ImuA' protein-protein interactions which are thought to enable DnaE2 to access the damage lesion during TLS (Warner *et al.*, 2010). The specific domains in DnaE2 that are critical for this function remain unknown, however. As noted above **(Table 1.2**), DnaE2 lacks the C-terminal τ domain but instead contains a highly conserved C-terminal pentapeptide motif, SRDFR and an N-terminal extension of ~40 amino acid residues that is only found in *Actinobacteria* (**Figure S1**) (Timinskas *et al.*, 2014, Warner *et al.*, 2010). Owing to their high conservation and lack of a defined structure, it was hypothesized that these DnaE2 domains are required for mediating protein–protein interactions during mutasome function.

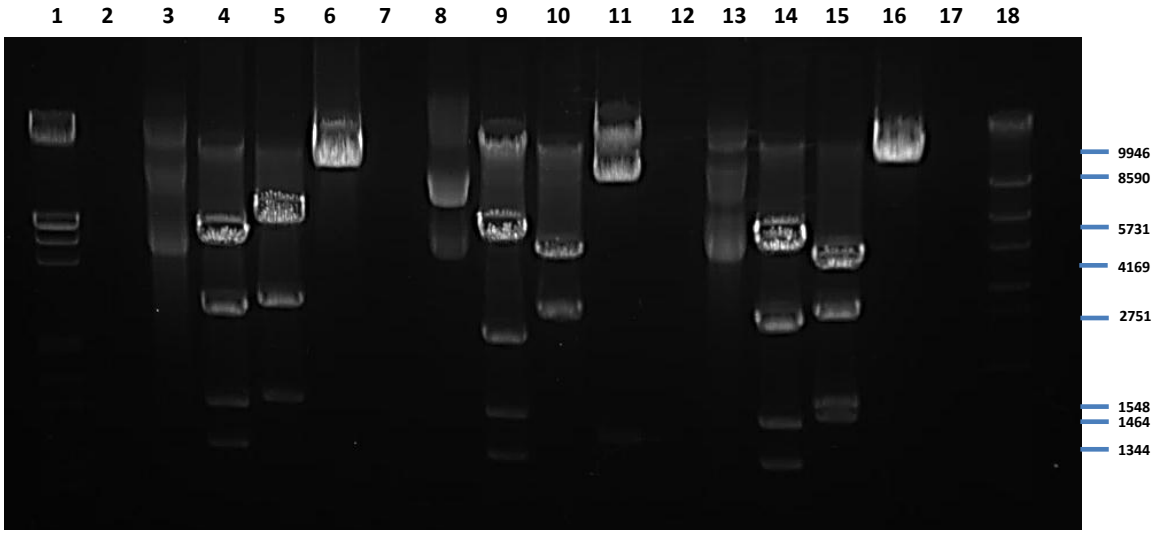
4.2.1 Construction of N- and C-terminally truncated *dnaE2* mutants

To test this hypothesis, mutant Msm strains were generated containing defined truncations of the N-terminal extension and, separately, the C-terminal pentapeptide motif in DnaE2. To this end, two alleles were engineered: (i) a *dnaE2* mutant allele in which the C-terminal pentapeptide-encoding motif was truncated (*dnaE2*$^{\Delta SRDFR}$) and (ii) a second *dnaE2* allele in which 38 N-terminal residues were eliminated, resulting in an N-terminally truncated DnaE2 (*dnaE2*^{N38}). The alleles were subcloned into the integrating vector, pMC1s (Warner *et al.*, 2010), to generate mutant complementation vectors that were subsequently electroporated into the Δ*dnaE2* mutant background. As such, this strategy investigated the ability of the N- and C-terminally truncated DnaE2 proteins to complement a *dnaE2* null mutant, analogous to the approach that was used previously to elucidate the structure-function relationships of the other mutasome components, ImuA' and ImuB (Warner *et al.*, 2010).

The *dnaE2*$^{\Delta SRDFR}$ and *dnaE2*^{N38} fragments contained engineered *Bgl*II and *Spe*I restriction sites, respectively, to enable differentiation of WT and mutant alleles. Digestion of pMC1r*dnaE2* with *Spe*I results in a linear fragment of 9946 bp (**Figure 4.8A**, lane 6). The additional *Spe*I site in pMC1r*dnaE2*^{N38} cleaves the 9946 bp fragment resulting in 1344 bp and 8590 bp, respectively (**Figure 4.8A**, lane 11). Digestion of pMC1r*dnaE2* with *Bgl*II results in 2751 bp, 18 bp, 1464 bp and 5713 bp fragments (**Figure 4.8A, lane 5**). The additional *Bgl*II site in pMC1r*dnaE2*$^{\Delta SRDFR}$ cleaves the 5713 bp in WT *dnaE2* resulting in 2751 bp, 18 bp, 1464 bp, 1548 bp and 4169 bp fragments, respectively (**Figure 4.8A, lane 15**).

The Msm Δ*dnaE2* mutant was then transformed with vectors carrying the WT *dnaE2* allele (yielding strain ZD 7) and the *dnaE2*$^{\Delta SRDFR}$ (ZD 8) and *dnaE2*^{N38} (ZD 9) alleles. In each case, the complementation vector integrated into the *att* locus in the Δ*dnaE2* chromosome; this was confirmed using the PCR strategy shown in **Appendix B**. Amplification of the WT *dnaE2* complementation mutant and *dnaE2* mutant strains using attL2 and attBS1 primers resulted in a PCR product of the expected size (282 bp) for each strain (**Figure 4.8**). Mutant strains of Msm with a validated *dnaE2* genotype were assessed for DNA damage tolerance and UV-induced mutagenesis (**section 4.2.2**). In addition, the same alleles were used as templates to generate a panel of "bait" and "prey" vectors for Y2H analyses designed to assess the impact of deletion of these terminal sequences on protein-protein interactions involving DnaE2 (**section 4.2.4**).

A


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
9946
8590
5731
4169
2751
1548
1464
1344

B

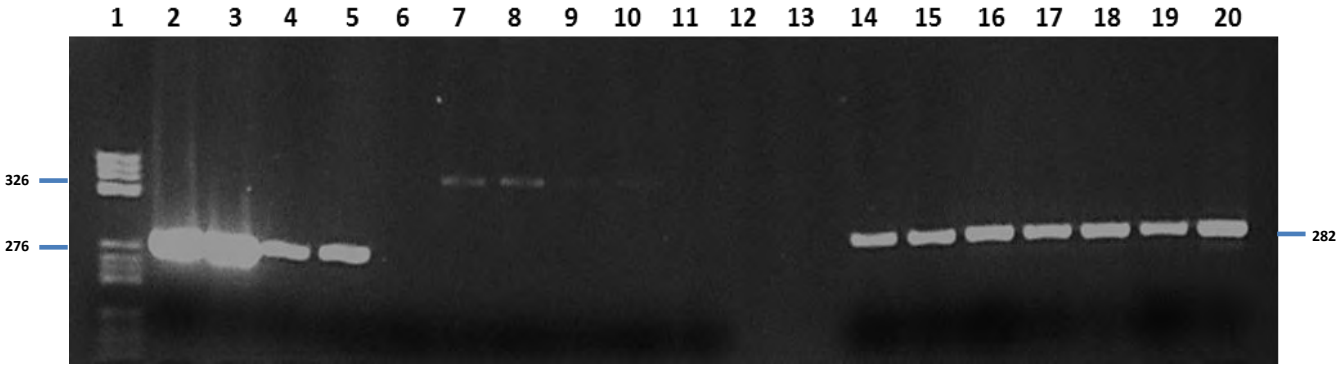

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
326
276
282

Figure 4.8: Genotypic confirmation of *dnaE2* constructs and mutants. (**A**) pMC1r*dnaE2* (lanes 4-6), pMC1r*dnaE2*^{N38} (lanes 9-7) and pMC1r*dnaE2*$^{\Delta SRDFR}$ (lanes 14-16) were digested with *Bam*HI, *Bgl*II, and *Spe*I, and analysed on a 1% agarose gel. Undigested controls were included for each construct (lanes 3, 8 and 13). Lane 1, molecular weight marker III (Roche); lane 3, undigested pMC1r*dnaE2*; lane 4, pMC1r*dnaE2* digested with *Bam*HI; lane 5, pMC1r*dnaE2* digested with *Bgl*II; lane 6, pMC1r*dnaE2* digested with *Spe*I; lane 8, undigested pMC1r*dnaE2*^{N38}; lane 9, pMC1r*dnaE2*^{N38} digested with *Bam*HI; lane 10, pMC1r*dnaE2*^{N38} digested with *Bgl*II; lane 11, pMC1r*dnaE2*^{N38} digested with *Spe*I; lane 13, undigested pMC1r*dnaE2*$^{\Delta SRDFR}$; lane 14, pMC1r*dnaE2*$^{\Delta SRDFR}$ digested with *Bam*HI; lane 15, pMC1r*dnaE2*$^{\Delta SRDFR}$ digested with *Bgl*II; lane 16, pMC1r*dnaE2*$^{\Delta SRDFR}$ digested with *Spe*I and lane 18, molecular weight marker IV (Roche) (**B**) Integration of the WT and *dnaE2* mutant alleles at the bacterial *att* locus was verified using primers that amplify hybrid portions originating from *att*B and *att*P. The *att*B primer set amplified the *attB* site which is 276 bp long (lanes 2-5) and the *att*L primers amplify a portion of the vector sequence which is 326 bp long (lanes 7 and 8). Genomic DNA was screened using *att*B primers and the vector (pMC1r) was screened using *att*L primers. Three colonies from ZD 9 and two from ZD 7 and ZD 8 were screened using attL2 and attBS1 primers which yielded PCR products of 282 bp.

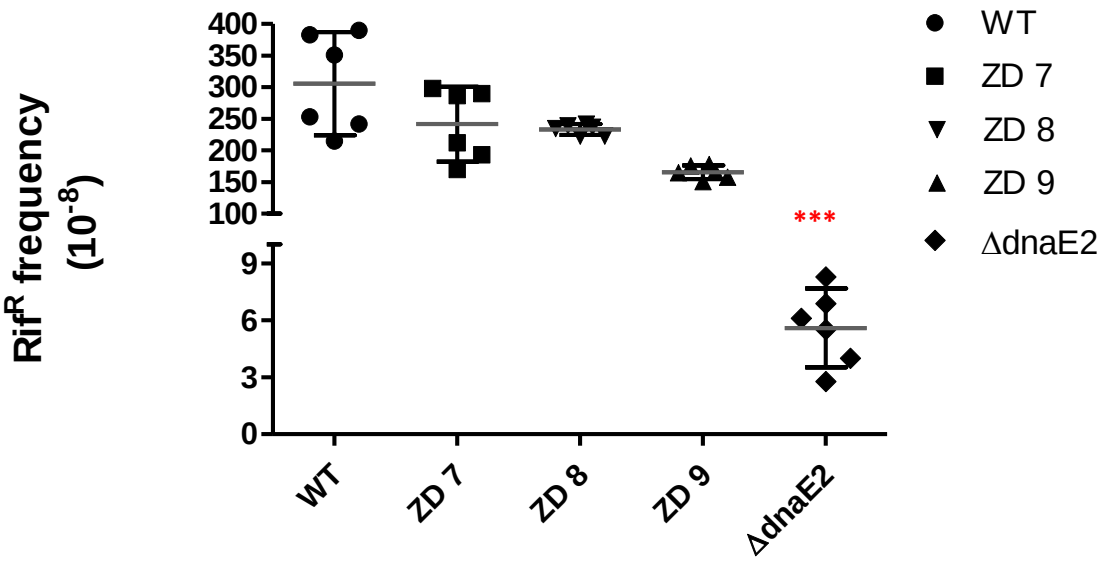
4.2.2 The N-terminal extension and C-terminal pentapetide motif are not required for DnaE2-dependent damage tolerance and UV-induced mutagenesis

To determine whether deletion of either the N-terminal extension or the C-terminal motif in DnaE2 impacted polymerase function, the ZD 8 (*dnaE2*$^{\Delta SRDFR}$) and ZD 9 (*dnaE2*^{N38}) mutants were assessed in DNA damage-induced mutagenesis. It was expected that, if either terminal sequence was essential for DnaE2 polymerase function, a Δ*dnaE2* mutant carrying the respective *dnaE2*$^{\Delta SRDFR}$ and *dnaE2*^{N38} allele should phenocopy the Δ*dnaE2* deletion mutant. That is, the mutant alleles should not reverse the deficiency in UV-induced mutagenesis to Rif that is characteristic of the Δ*dnaE2* strain (Boshoff *et al.*, 2003, Warner *et al.*, 2010).

As observed previously (Boshoff *et al.*, 2003, Warner *et al.*, 2010), the Δ*dnaE2* mutant was associated with a low frequency of UV-induced Rif-resistance; moreover, complementation with the WT *dnaE2* allele (ZD 7) reversed this phenotype, although it did not fully restore the Rif-resistance mutation frequencies to WT (**Figure 4.9A**). In contrast, no DNA-damage phenotype was observed for both ZD 8 and ZD 9 *dnaE2* mutants following UV damage, as evidenced by the fact that no differences in mutation frequencies to Rif were observed for both mutants relative to WT ($p>0.05$) (**Figure 4.9A**).

In addition to DNA damage-induced mutagenesis, the *dnaE2* deletion mutant was previously shown to be hypersensitive to MMC treatment; again, this effect could be reversed by integration of the WT copy of *dnaE2* at the *attB* locus (Boshoff *et al.*, 2003, Warner *et al.*, 2010). Here, $\log_{10}$-fold dilutions of each strain were plated on MMC at two concentrations known to elicit a damage tolerance phenotype (Warner *et al.*, 2010). As observed previously, the Δ*dnaE2* was hypersusceptible to MMC at both concentrations, and this phenotype was reversed by complementation with the WT *dnaE2* allele in the ZD 7 strain. In contrast, neither *dnaE2*$^{\Delta SRDFR}$ nor *dnaE2*^{N38} mutant allele resulted in a damage sensitivity phenotype following exposure to MMC, with both ZD 8 and ZD 9 mutants phenocopying the WT strain (**Figure 4.9B**). In summary, these observations indicated that elimination of either N-terminal extension or C-terminal pentapeptide motif did not disrupt DnaE2 function in DNA damage tolerance and mutagenesis assays *in vitro*.

A


Rif^R frequency (10^-8)
400
350
300
250
200
150
100
9
6
3
0
***
WT
ZD 7
ZD 8
ZD 9
ΔdnaE2
WT
ZD 7
ZD 8
ZD 9
ΔdnaE2

B

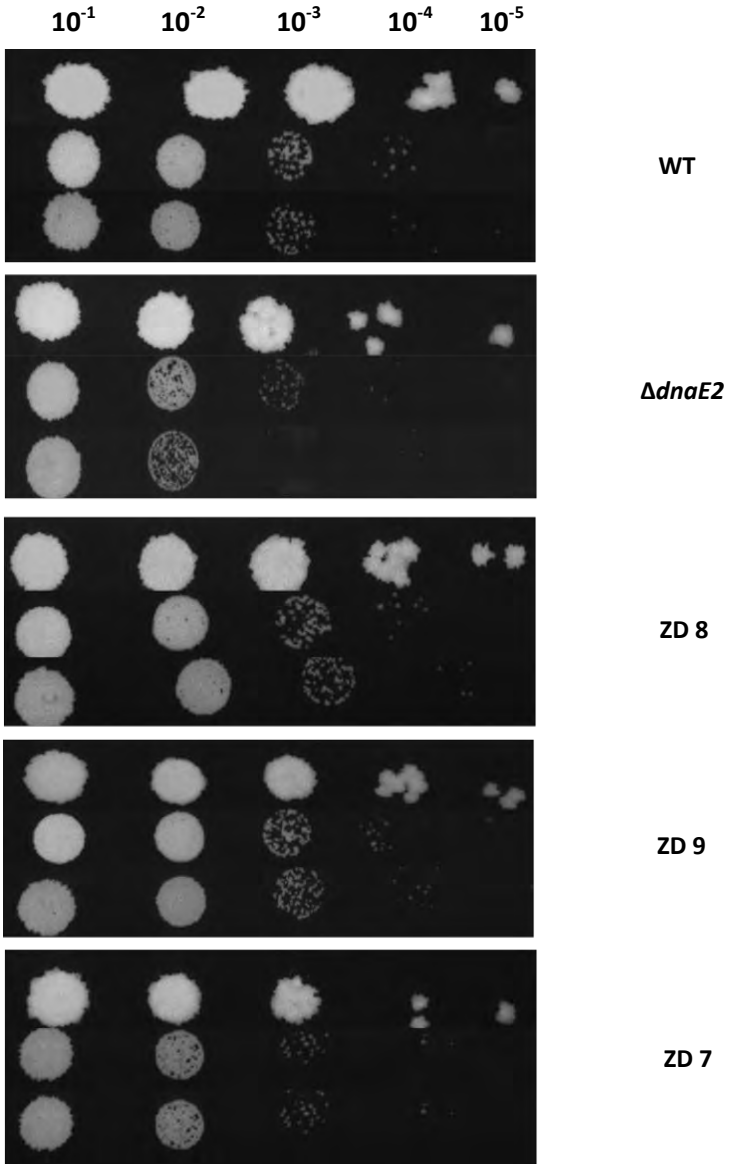

10^-1
10^-2
10^-3
10^-4
10^-5
WT
ΔdnaE2
ZD 8
ZD 9
ZD 7

Figure 4.9: The N-terminal extension and C-terminal motif in DnaE2 are not required for induced mutagenesis and DNA damage tolerance. (**A**) UV mutation frequencies of WT, *dnaE2* complementation strain (ZD 7), ZD 8, ZD 9 and Δ*dnaE2* following UV-irradiation. (**B**) Log_{10}-fold dilutions of the mycobacterial cultures (10^{-1}-10^{-5}) were spotted on 7H10 OADC media without MMC (top row for each strain), with 0.02 µg/ml MMC (second row for each strain); and 0.04 µg/ml MMC (third row for each strain. ***$p<0.001$

4.2.3 Identification of protein-protein interactions between putative mutasome components.

To determine whether the C-terminal pentapeptide motif and/or the N-terminal extension in DnaE2 was required for mediating protein-protein interactions during mutasome function, "bait" and "prey" constructs containing the targeted DnaE2 mutations ($DnaE2^{\Delta SRDFR}$ and $DnaE2^{N38}$) were generated and assessed in Y2H assays. Protein-protein interactions were previously detected between Mtb ImuB and DnaE1, and both proteins were also shown to interact separately with the *dnaN*-encoded β_2 clamp processivity factor, consistent with the presence in DnaE1 and ImuB of β_2 clamp-binding motifs (Kana *et al.*, 2010, Warner *et al.*, 2010). In addition, DnaE2 was shown to interact with ImuB (Warner *et al.*, 2010). Therefore, for the current analyses, Msm WT DnaE2, WT DnaE1 and ImuB genes were cloned into pGADT7 (ImuB) or pGBKDT7 (DnaE1 and DnaE2) and these constructs were used as positive controls (**Table 3.2**). For negative controls, pGADT7 was co-transformed with the empty plasmid pGBKT7, pGBDDnaE2 or $pGBDDnaE2^{N38}$ (**Table 3.2**) constructs, respectively.

Consistent with previous findings (Warner *et al.*, 2010), an interaction was detected between DnaE2 and ImuB, although cellular growth was poor even on low stringency (LTH) medium. This contrasted with DnaE1 and ImuB, for which a strong interaction was observed even on high stringency medium (LTHA-3AT) (**Figure 4.10)**. As expected, no interaction was observed for the pGADT7-pGBKT7, pGADT7-pGBDDnaE2 and pGADT7-$pGBDDnaE2^{N38}$ negative controls. Both DnaE2 mutants co-transformed with ImuB failed to grow on high stringency medium (LTH-3AT) compared with WT DnaE2-ImuB (**Figure 4.10**). In combination, these results suggested that the N-terminal extension and C-terminal motif in DnaE2 are necessary for mediating the *imuA'-imuB/dnaE2* interaction during TLS. However, the observation that the corresponding alleles did not impact DnaE2 function in Msm (section 4.2.2) indicates that inferences about protein-protein interactions must be made with caution.

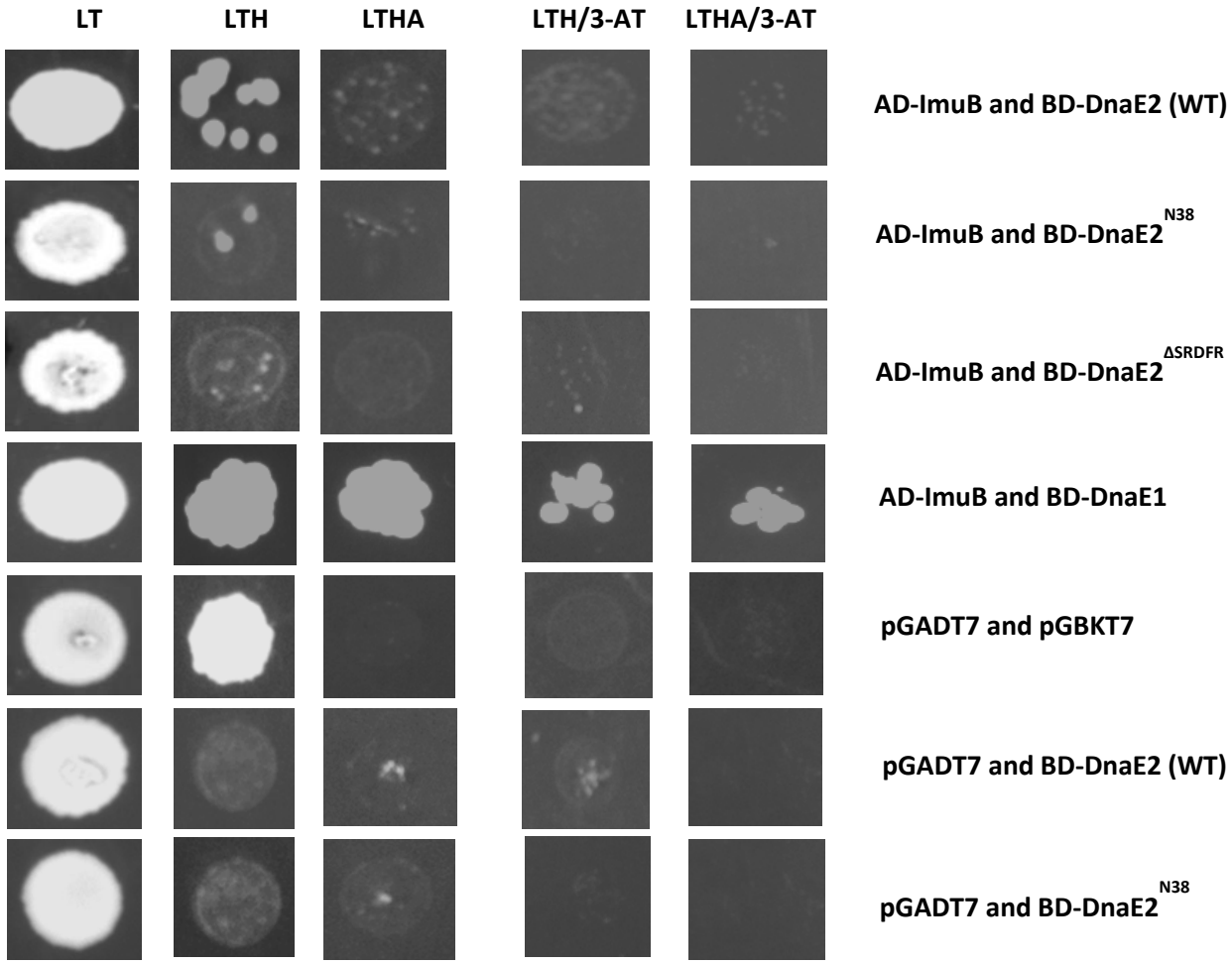


Figure 4.10: The N- and C-terminal domains in DnaE2 are not required for interaction with ImuA' and ImuB. Protein interactions were detected on synthetic drop-out (SD) medium lacking leucine and tryptophan (LT); leucine, tryptophan and histidine (LTH); and leucine, tryptophan, histidine and adenine (LTHA). 3-amino triazole (3-AT) was added to medium and high stringency media in order to reduce HIS3 expression required for growth on histidine-deficient media. pGADT7-pGBKT7, pGADT7-pGBDDnaE2 and pGADT7-pGBDDnaE2^{N38} were used as negative controls, whereas pGADImuB and pGBDDnaE1 were used as a positive controls. Data are from a single representative experiment performed in triplicate.

CHAPTER 5: DISCUSSION

5.1 Replication fidelity in mycobacteria

In most prokaryotes, replication fidelity is determined by a combination of three separate but inter-related processes: 1) base selection by the replicative polymerase 2) exonuclease proofreading and 3) post-replicative MMR (Fijalkowska *et al.*, 2012). Mycobacteria lack the MMR system (Mizrahi and Andersen, 1998), suggesting that base selection and proofreading are sufficient to ensure overall replication fidelity, or that alternative mechanisms exist for the correction of replication errors. To elucidate the molecular mechanisms governing replication fidelity in mycobacteria, the contributions of two subunits of the mycobacterial PolIII enzyme were investigated: DnaE1, the essential α subunit which contains a PHP domain with putative proofreading activity, and the *dnaQ*-encoded ε subunit which, in *E. coli,* encodes the major proofreading function. Given that the Mtb genome encodes a second, error-prone α subunit, DnaE2, the differential fidelity of the two α subunits (DnaE1 and DnaE2) was also investigated to determine whether 1) it was a direct function of their respective intrinsic biochemical properties (PHP-mediated exonuclease proofreading) or 2) whether this property was instead determined by differential interactions with other cellular components via specific structural domains. Therefore, the second part of this study adopted a genetic approach to investigate the requirement for specific N- and C-terminal domains of DnaE2 in damage-induced mutagenesis.

5.1.1 The PHP domain in DnaE1 contains a complete set of metal co-ordinating residues

The PHP domains of PolIIIs from some Gram-positive bacteria have been shown to contain a complete set of metal-binding residues that are important for 3'-5' exonuclease activity (Banos *et al.*, 2008, Nakane *et al.*, 2009, Stano *et al.*, 2006, Wing *et al.*, 2008). In contrast, no exonuclease activity could be detected for the PHP domain of *G. kaustophilus* PolC (Evans *et al.*, 2008) which contains an almost intact active site – suggesting that a complete set of the catalytic metal co-ordinating residues is required to constitute an "active" PHP domain; that is, a PHP domain with exonuclease activity. In this study, the essential PolIIIα subunit in

mycobacteria, DnaE1, was demonstrated to contain a complete set of the conserved metal co-ordinating residues in its PHP domain, suggesting that it might possess 3'-5' exonuclease activity. Interestingly, DnaE2 was also found to contain an almost intact 3'-5' exonuclease active site in its PHP domain, with eight out of nine residues conserved, the only difference being the D228A substitution in Msm or D228G substitution in Mtb instead. This was reported elsewhere (Timinskas *et al.*, 2014) and raises an important question as to whether the PHP domain in DnaE2 is active or not, given the observation that conservation of the same eight residues is not sufficient for exonuclease function in *G. kaustophilus* PolC. In this regard, the conservation in a subgroup of Actinobacterial DnaE2s of an almost intact PHP active site (**Figure S1**) (Timinskas *et al.*, 2014) might indicate a role in the maintenance of structural fidelity rather than proofreading since the results of this study (discussed below) indicate that proofreading can be eliminated by a single deviation (*dnaE1*^{D228N}) from the highly conserved nine-residue PHP exonuclease active site.

Even though the third DnaE-type C-family polymerase, DnaE3, strongly resembles DnaE1 in terms of domain architecture, the PHP domain in DnaE3-type polymerases is smaller, as it lacks several structural elements (Timinskas *et al.*, 2014). In contrast to DnaE2 which contains eight of the nine conserved metal co-ordinating residues, the PHP domain in DnaE3-type polymerases are very diverse and exhibit limited identity with the corresponding DnaE1 and DnaE2 PHP domains, suggesting that DnaE3 proteins are completely deficient in exonuclease proofreading activity (Timinskas *et al.*, 2014). Similarly, the PHP domain in *E. coli* has been shown to lack five of the nine metal co-ordinating results and therefore possesses no proofreading activity. However, re-introduction of the catalytic residues did not restore the proofreading function but instead stabilized the PolIII core, suggesting that the PHP domain in *E. coli* is critical for maintenance of the structural integrity of the PolIII HE, instead of genome maintenance (Barros *et al.*, 2013).

5.1.2 The conserved PHP domain residues in DnaE1 are required for genome maintenance

In addition to the metal co-ordinating residues, the PHP domain in DnaE1 was also shown to contain conserved residues that are predicted to be required for maintaining replication fidelity in mycobacteria (E133). It is worth noting that these residues are absent in the PHP

domain of DnaE2 and other non-essential DnaE-type polymerases (**Figure S1**) (Warner *et al.*, 2010), possibly correlating with the inability of the alternative DnaEs to carry-out high fidelity DNA synthesis. Mutation of the Glu-133 residue in DnaE1 resulted in loss of intrinsic replication fidelity, suggesting that additional residues not directly involved in metal co-ordination are also necessary for robust exonuclease activity (Barros *et al.*, 2013).

In addition to metal co-ordination, the Asp-228 residue was also found to be conserved in DnaE1 and absent in DnaE2. Efforts to generate an Msm DCO mutant containing only the *dnaE1*^{D228N} allele proved unsuccessful despite numerous attempts. It was striking, therefore, that a merodiploid strain (ZD 3) carrying both WT *dnaE1* and the *dnaE1*^{D228N} allele was associated with a 10-fold increase in spontaneous mutation rate. This study did not investigate gene expression from the different alleles; however, if it is assumed that the alleles are expressed equally, the mutator effect observed in ZD 3 implies that any disruption to intrinsic proofreading function in the active replisome can impact replicative fidelity significantly. It is tempting to speculate, therefore, that upregulation in the mycobacterial SOS response of *dnaE2*, a "natural" D228N mutant, might impact fidelity – a notion that is consistent with the elevated mutagenesis observed following DNA damage.

Failure to generate the isogenic *dnaE1*^{D228N} mutant suggests that there is a high selection pressure against this mutation. Mutation of this residue could possibly result in "error catastrophe" which will have an impact on genome maintenance as the fidelity of the replication machinery decreases. In support of this notion, a mutator phenotype was observed for the merodiploid strain carrying both WT and *dnaE1*^{D228N} alleles at the native *dnaE1* chromosomal locus, suggesting the importance of this specific residue in genome maintenance. Previous studies measured mutation rates for the Δ*dnaE2* mutant and demonstrated no differences between this mutant relative to WT (Boshoff *et al.*, 2003). Given the strong mutator phenotype observed in this study associated with the DnaE1^{D228N} mutation and observations that this residue is required to complete the set of metal co-ordinating residues in DnaE2, further studies are required to test whether introduction of this PHP domain residue in DnaE2 would enhance replication fidelity (but decrease damage tolerance) following exposure to genotoxic stress.

5.1.3 DnaQ is not required for maintaining replication fidelity in mycobacteria

The presence of a *dnaQ* homolog in mycobacteria suggests that the mycobacterial PolIII HE might use DnaQ for proofreading, as in the case of *E. coli* (Fijalkowska and Schaaper, 1995, Horiuchi *et al.*, 1978, Degnen and Cox, 1974, Fijalkowska and Schaaper, 1996). In the *E. coli* α subunit, the counterpart of the Asp-49 residue has been shown by structural analysis to bind the DnaQ proofreading subunit (Wieczorek and McHenry, 2006). In contrast to the model organism, mutation of the Asp-49 residue in DnaE1 did not impact the mutation rate of Msm, suggesting that this residue is not required for binding DnaQ. Moreover, a 4.3-fold increase in mutation rate was observed for the ZD 6 (Δ*dnaQ* *dnaE1*^{E133}) mutant, which appeared to be an additive effect of the Δ*dnaQ* (1.4-fold) and *dnaE1*^{E133} (3-fold) mutations. This suggests that DnaQ and Glu-133 residue in DnaE1 perform non redundant functions and that the Glu-133 residue in DnaE1 does not bind DnaQ during DNA replication.

Barros and colleagues demonstrated a correlation between a variant PHP domain – a PHP domain lacking one or more conserved catalytic metal co-ordinating residues – with the presence of a separate proofreading subunit, DnaQ in *E. coli* (Barros *et al.*, 2013). This raises questions in the context of mycobacteria which contain both the intact PHP domain and two *dnaQ* homologs. That is, is the inferred correlation between the presence of an intact PHP domain and the possession of a separate DnaQ subunit of any relevance to this organism, given that both mycobacterial *dnaQ* homologs have been demonstrated to be dispensable for replication fidelity?

5.1.4 Mycobacteria use non canonical mechanisms of proofreading to maintain genome integrity.

Deletion of *dnaQ* in mycobacteria has been shown to have a very modest effect on mutation rate (1.4-fold increase). During the course of this work, Ford reported the results of an independent analysis of the function of the canonical *dnaQ* gene and the second *dnaQ*-type homolog (*MSMEG_4259*) on replication fidelity in Msm and Mtb*,* whereby deletion of *MSMEG_4259* was found to have no impact on the mutation rate. Moreover, consistent with findings presented here, deletion of the canonical *dnaQ* gene was found to result in a

1.4-fold loss in fidelity (Ford, 2012). That study also showed that deletion of both *dnaQ* homologs did not further increase the mutation rate (Ford, 2012). The findings were intriguing since *E. coli* strains with inactivating mutations in *dnaQ* were reported to have 100-1000 fold increase in mutation rate (Fijalkowska and Schaaper, 1995, Horiuchi *et al.*, 1978, Degnen and Cox, 1974, Fijalkowska and Schaaper, 1996), suggesting that mycobacteria use alternative mechanisms of proofreading to maintain genome integrity.

Furthermore, in contrast to *E. coli* (Gerdes *et al.*, 2003) and other organisms (Akerley *et al.*, 2002, van Opijnen *et al.*, 2009), *dnaQ* homologs in mycobacteria were found to be dispensable for viability (Griffin *et al.*, 2011, Sassetti *et al.*, 2001, Sassetti and Rubin, 2003). Again, this observation raises the question: why does Mtb contain two homologs of *dnaQ*? *M. leprae*, which lacks both *dnaQ* homologs, may have provided some insight about the dispensability of the *dnaQ* homologs in genome maintenance in mycobacteria. It is worth noting that this organism, which is undergoing genomic decay (Cole *et al.*, 2001), contains a complete set of the metal co-ordinating residues in the PHP domain of its replicative polymerase, suggesting the crucial role for the PHP domain in genome maintenance in mycobacteria. In addition, in a comprehensive genome-wide screen for genes under selection in clinical drug-resistant Mtb isolates performed by Farhat and colleagues, *dnaQ* (Rv3711) was among those genes which were found to be associated with drug resistance (Farhat *et al.*, 2013). Does this mean that DnaQ is serving in another pathway that may not involve replicative fidelity directly? The identification of genes that are conditionally essential in the absence of the *dnaQ* homologs may provide some insight into proteins that serve in the same pathway as the *dnaQ* homologs, and hence their possible function. Most importantly, based on findings presented here, an important question arises: is the fidelity in mycobacteria solely determined by the intrinsic 3'-5' exonuclease activity in the PHP domain in DnaE1?

Springer and colleagues demonstrated that the key NER helicase, UvrD1, fulfils a critical role in limiting recombination-associated mismatches, therefore providing evidence that other repair components have enabled Mtb to mitigate the lack of the MMR system (Guthlein *et al.*, 2009). It is also possible that other replication-associated factors contribute to

replication fidelity in mycobacteria such as the interplay between replisome components and dNTP levels at the replication fork. There is no direct evidence for this mechanism in mycobacteria, however, in *E. coli*, increased dNTP levels were shown to modify the balance of PolIII activity by increasing the polymerase activity and reducing the proofreading function (Gon *et al.*, 2011). The same study demonstrated that spontaneous mutagenesis increased proportionally to dNTP pool levels, therefore defining a unique "dNTP spontaneous mutator" phenotype (Gon *et al.*, 2011). Another factor that contributes to replication fidelity is the co-ordination between the replication machinery and the pathways that function in chromosomal segregation and cell division (Aldridge *et al.*, 2012, Arjes *et al.*, 2014, Joyce *et al.*, 2012, Kester and Fortune, 2014, Kieser and Rubin, 2014, Kysela *et al.*, 2013). Further work focused on the relationship between dNTP pools and the replication rate of Mtb at different stages of infection is required to enable accurate estimates of the intracellular dNTP concentration during different phases of the mycobacterial cell cycle. In addition to this, even though there are insights into proteins that are involved in DNA replication, further studies focusing on identification of mycobacterial replisome components and their organization may assist in understanding how genome integrity is maintained in Mtb.

5.1.5 A role for mutator strains in Mtb

Considering that Mtb does not exhibit an elevated mutation rate relative to other bacteria (Ford *et al.*, 2011), how do mutator strains arise and is there any evidence to support a role for mutators in this organism? For many bacterial pathogens, the absence of key DNA repair functions might facilitate adaptation during specific stages of the life cycle; however, the loss of repair machinery can impact long-term colonization and transmission. To some extent, the benefits of a mutator phenotype for the development of drug resistance are likely to be outweighed by negative effects on virulence and the susceptibility of mutators to extinction as a result of evolutionary bottlenecks (Warner *et al.*, 2013). However, recent studies using WGS demonstrated strain-specific differences in mutation rates, whereby Mtb strains from lineage 2 which include the Beijing family, were shown to acquire drug resistance *in vitro* more rapidly compared with strains from lineage 4 (Euro-American lineage). This higher rate was attributed to a higher basal mutation rate in lineage 2 strains,

although the exact mechanism is unknown (Ford *et al.*, 2013). Therefore, even though small changes in mutation rate for the ZD 2 and ZD 3 PHP mutants (3-fold and 10-fold, respectively) were observed, small changes can profoundly affect the rate at which strains from different lineages acquire drug resistance, as demonstrated by Ford *et al.* (Ford *et al.*, 2013). The findings reported in this study suggest that polymorphisms in the PHP domain of DnaE1 may confer strain-specific alterations in replicative fidelity. Moreover, since none of PHP domain polymorphisms have been found among clinical strains of Mtb that have been sequenced so far, this implies that there is a high selection pressure in this domain.

This study has limitations, including the fact that the choice of amino acid residues selected for mutation in the PHP domain of Msm DnaE1 was guided by insights gained from homology modelling as the structure of DnaE1 has not been determined. Secondly, a genetic approach was used to understand the contribution of the conserved PHP domain residues in DnaE1 function. Biochemical studies were not performed to validate the exonuclease activity of the PHP domain in DnaE1 and to understand and contextualise the effects (or lack thereof) of the PHP domain mutations in maintaining genome integrity which is one of the limitations of this study. These studies would have also provided insights into the potential catalytic activity of the PHP domain in DnaE2. Moreover, the impact of the PHP domain point mutations on the overall structural integrity of DnaE1 was not investigated. This could be addressed using structural studies which were also not performed in this study. Based on recent observation in *E. coli* which demonstrated that mutations in PHP domain of PolIII affected the overall stability of the protein (Barros *et al.*, 2013), this warrants further investigation.

In conclusion, in contrast to model organisms where high fidelity DNA replication is dependent on the DnaQ proofreading subunit, this study demonstrated that the corresponding subunit is dispensable in Msm. Instead, conserved residues in the PHP domain in DnaE1 are required for genome maintenance, therefore updating the current view of the mycobacterial replisome (**Figure 5.1**) and demonstrating that mycobacteria use non-canonical mechanisms of proofreading to maintain genome integrity. This conclusion is supported by very recent evidence from the laboratory of Sarah Fortune (Harvard

University), which was published during the examination of this thesis. In their very elegant study, a combination of genetic, biochemical, and comparative genomic analyses established the essential role of the PHP-dependent 5'-3'exonuclease activity in maintaining the mycobacterial mutation rate (Rock *et al.*, 2015).

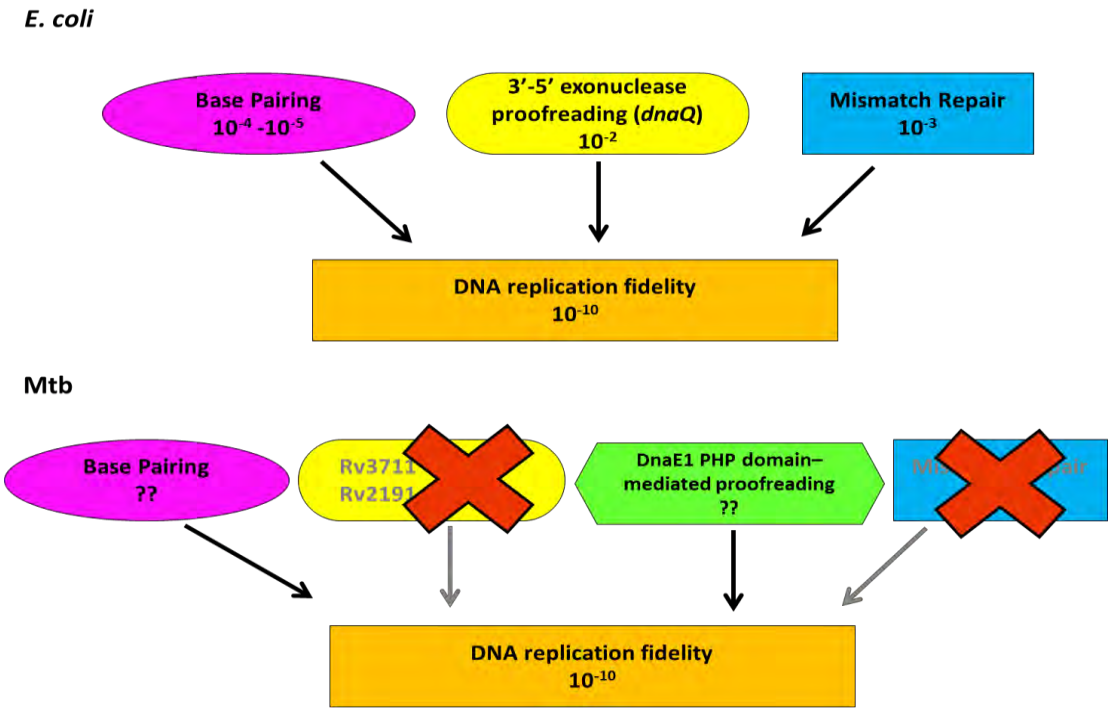


Figure 5.1: A model for DNA replication fidelity in *E. coli* versus Mtb. In *E. coli,* DNA replication fidelity is accomplished using three primary mechanisms - base selection by the replicative polymerase, 3'-5'exonuclease proofreading by the separate DnaQ subunit, and postreplicative MMR - which contribute to the overall error rate of 10^{-10} mutations per base pair per generation. The error rate for each process is shown in the figure. Mtb lacks the MMR system and this work, as well as a recent study (Rock et al., 2015), has established that the *dnaQ* homologs are dispensable for proofreading in mycobacteria. Instead, conserved residues in the PHP domain in DnaE1 are required to maintain overall mycobacterial replication fidelity. In Mtb, the contribution of each mechanism to the overall error rate of 10-10 mutations per base pair per generation is still unknown.

5.2 The mycobacterial SOS and damage-induced response

5.2.1 The role of the N- and C-terminal domains in DnaE2 in mediating protein-protein interactions during mutasome function

The dominant role for DnaE2 in DNA damage-induced mutagenesis in Mtb is well established (Boshoff *et al.*, 2003). Moreover, DnaE2 has been shown to function in association with other DNA damage response proteins, ImuA' and ImuB during TLS, thereby implicating differential protein interactions in specialist polymerase function (Warner *et al.*, 2010). The C-terminal region of ImuB has been shown to mediate interactions with other mutasome components suggesting that ImuB may act as a hub/adaptor protein that interacts with both DnaE2 and ImuA' in mediating TLS (Warner *et al.*, 2010). Y2H data demonstrated that the DnaE2^{N38} and DnaE2$^{\Delta SRDFR}$ mutants failed to interact with ImuB, thus potentially implicating interactions between ImuB and both the N-terminal extension and C-terminal motif in DnaE2 mutasome function. Consistent with previous studies (Warner *et al.*, 2010), an interaction was detected between WT DnaE2 and ImuB, however it was associated with very poor colony growth. This complicated comparison of very small changes between ImuB and the DnaE2 mutants relative to WT DnaE2. Based on this notion, it might be concluded that the deletion of the N-terminal extension and C-terminal motif in DnaE2 mutations did not impact the interaction with ImuB significantly.

It should be noted, however, that even though Y2H assays are still one of the most widely used methods for determining protein-protein/DNA interactions *in vivo*, there are important limitations to the technique: for example, the interacting proteins must be localized to the nucleus, since proteins which are less likely to be present in the nucleus are excluded due their inability to activate reporter genes. In addition, many true interactions may not be traced using Y2H assays, therefore leading to false negative results. Proteins which need posttranslational modifications to carry out their function are unlikely to behave or interact normally in a Y2H experiment. Most importantly, the proteins are not in their natural physiological environment and, therefore, may not fold properly to interact, also leading to false negative results (Rao *et al.*, 2014). Moreover, proteins form part of complexes or networks *in vivo*; in the case of the Y2H system, the proteins are not in the host cell, where other proteins which may form part of the functional interactome are present. If other

proteins compensate for the loss of function of another protein in the host cell, this effect will not be observed in the Y2H system which only determines interaction between protein pairs, therefore resulting in false negative results (Rao *et al.*, 2014). In addition, since the inferred interactions occur in a heterologous system: yeast *versus* Msm*,* under non-DNA-damaging conditions, it is not clear whether these interactions are relevant, in any way, during TLS.

Research has advanced such that techniques based on live cell imaging are now being used to address some of the limitations of Y2H assays and these can be included in future work involving determination of protein-protein interactions in mycobacteria. One such example is fluorescence resonance energy transfer (FRET) microscopy, which uses time-correlated single-photon counting to predict protein interactions (Lleres *et al.*, 2007). FRET is more advantageous in that it allows for retrieving information on molecular proximity, spatial organization, orientation and conformation of molecules on the nanometre scale and interactions between molecules can be quantified from living cells (Pietraszewska-Bogiel and Gadella, 2011), therefore increasing the sensitivity of this technique.

5.2.2 The N-terminal extension and C-terminal motif in DnaE2 are not required for DnaE2 function

Prior work in the Warner and Mizrahi laboratory characterised mutants of Msm and Mtb that are impaired in mutasome function (e.g. *dnaE2*, *imuA'* and *imuB* deletion mutants), showing specifically that these strains exhibit: an increase in sensitivity to certain DNA damaging agents (e.g. UV and MMC) and an inability to display UV-induced mutation to Rif-resistance (Boshoff et al., 2003; Warner et al., 2010). The same assays were therefore used to assess the impact of the DnaE2 mutations investigated in this study. The *dnaE2*$^{\Delta SRDFR}$ and *dnaE2*^{N38} mutants were indistinguishable from the parental WT strain in both assays, thus ruling out a significant role for the N-terminal extension and C-terminal pentapeptide motif in DnaE2 function during TLS.

One possible reason for the contrasting conclusions from these phenotypic assays *versus* the Y2H analysis, is that the phenotypic assays were conducted under damage-inducing conditions, where there is elevated expression of DnaE2, which may mask any effect that deletion of one domain may potentially have on polymerase function. In support of this, DnaE2 levels were shown to be 50-times greater than physiological conditions following UV-damage (Warner *et al.*, 2010). However, based on the sensitivity the phenotypic assays employed in this study, it is reasonable to conclude that the N-terminal extension and C-terminal motif are not critical for DnaE2 function.

5.2.3 The conserved PHP domain residues in DnaE1 are dispensable for damage tolerance and induced mutagenesis.

In addition to *dnaE2*, the mycobacterial damage response includes DnaE1 (Boshoff *et al.*, 2003, Warner *et al.*, 2010) instead of the Y-family polymerases, in contrast to most model organisms (Brooks *et al.*, 2001). No damage-induced phenotype was observed for the DnaE1 PHP mutants with defects in proofreading function, suggesting that the conserved PHP domain residues in DnaE1 are dispensable for damage tolerance and/or induced mutagenesis.

5.2.4 Mycobacterial *dnaQ-uvrC* is not required for UV-induced mutagenesis

Lack of a DNA damage phenotype for the Δ*dnaQ* and Δ*dnaQ-uvrC* following UV treatment suggests that neither mycobacterial *dnaQ* homolog is required for UV-induced mutagenesis. This finding contrasts observations from *E. coli*, whereby inactivation of the proofreading subunit, DnaQ, enabled PolIII-mediated TLS in the absence of specialist DNA repair polymerases IV and V (Borden *et al.*, 2002). Even though the present study demonstrated that both mycobacterial *dnaQ* homologs are dispensable for proofreading in mycobacteria, which may explain the absence of a phenotype following UV treatment, it is quite intriguing that this phenotype was still not observed for the DnaE1 PHP domain mutants with defects in proofreading function, even in a genetic background that is severely depleted of genes encoding specialist DNA polymerases (all *dinB* homologs and *dnaE2).* This implies that

molecular mechanisms used by mycobacteria to adapt to genotoxic stresses are different compared with model organisms.

UvrC is part of the UvrABC endonuclease, multi-enzymatic complex which carries out DNA repair in a sequential manner. Genes encoding the UvrABC complex have been identified in Mtb (Dos Vultos *et al.*, 2009, Mizrahi and Andersen, 1998). The *uvrB* deletion mutant of Msm (Kurthkoti *et al.*, 2008) and Mtb (Darwin and Nathan, 2005, Guthlein *et al.*, 2009) was shown to be hypersensitive to UV treatment and other DNA damaging agents (Kurthkoti *et al.*, 2008). The *uvrA* and *uvrB* genes were also shown to be induced following treatment by MMC (Rand *et al.*, 2003), as well as upon infecting human macrophages (Graham and Clark-Curtiss, 1999). The absence of a UV-induced damage phenotype in the *dnaQ-uvrC* mutant is quite intriguing since previous studies have demonstrated a role for the canonical Mtb UvrC protein in DNA repair following UV damage (Prammananan *et al.*, 2012). The *dnaQ-uvrC* gene is SOS-inducible and the encoded protein contains an additional GIY-YIG catalytic domain that is found in the N-terminus of UvrC-like nucleases – suggesting a potential role in NER. It is possible, therefore, that this GIY-YIG domain is not required for DNA repair following UV-induced damage, which may explain observations of the lack of phenotype for the Δ*dnaQ-uvrC* mutant. Further studies focused at individual contribution of different domains in DnaQ-UvrC to DNA damage following treatment with different DNA-damaging agents are required to further address these observations. In support of this notion, Davis and colleagues demonstrated differential requirements for different domains in UvrD2, which also functions in NER (Williams *et al.*, 2011). In this study they demonstrated that the C-terminal domain in UvrD2 is required for the helicase activity, whereas the N-terminal domain was demonstrated to be required for the ATPase activity (Williams *et al.*, 2011).

5.2.5 DnaQ-UvrC is required for MMC tolerance

Lack of a DNA damage sensitivity phenotype in the Δ*dnaQ* mutant following MMC treatment suggests that DnaQ is not required for tolerance of DNA damaging agents. In contrast the Δ*dnaQ-uvrC* mutant was profoundly hypersensitive to MMC treatment, even more than the Δ*dnaE2* mutant (Boshoff *et al.*, 2003). However, this phenotype was reversed in the Δ*dnaQ* Δ*dnaQ-uvrC* double knock-out mutant. The presence of a damage sensitivity

phenotype for the Δ*dnaQ-uvrC* mutant following MMC treatment, but absence of a phenotype following UV treatment, suggests that the regulation of mycobacterial damage response pathways might be lesion-specific (Boshoff *et al.*, 2003). Kesavan and colleagues demonstrated that the *dnaQ-uvrC* homolog in Mtb (Rv2191) is critical for persistence in the granuloma during chronic infection, where bacillary DNA damage requiring repair may be more likely (Kesavan *et al.*, 2009). These studies however, looked at the differential expression of this gene *in vivo*, which does not necessarily translate into functional importance. To the author's knowledge, the present work and that of Ford's (Ford, 2012) are the only studies, to date, which have attempted to determine the functional role of this protein.

Future work should then investigate whether deletion of *dnaE2* in the Δ*dnaQ-uvrC* background will further increase hypersensitivity to MMC, in order to assess whether these genes act in the same pathway or not. In addition, in order to assess whether the observed phenotype is due to the UvrC-like domain in DnaQ-UvrC, a strain with the entire UvrC-like domain truncated should be generated and assessed in similar assays. Deletion of Δ*dnaQ* in the Δ*dnaQ-uvrC* background resulted in the loss of the MMC-hypersensitive phenotype. Implication (s) of these observations also warrants further investigation.

APPENDICES

APPENDIX A: GROWTH MEDIA

All media was made up to a final volume of 1 litre with deionised water, and sterilised by autoclaving at 121°C for 20 minutes, unless otherwise stated.

2TY broth

16 g tryptone powder; 10 g yeast extract; 5 g sodium chloride.

Luria-Bertani broth (LB)

10 g tryptone powder; 5 g yeast extract; 10 g sodium chloride.

Luria-Bertani agar (LA)

10 g tryptone powder; 5 g yeast extract; 10 g sodium chloride; 15 g DIFCO™ agar powder.

Middlebrook-OADC (7H9-OADC)

4.7 g Middlebrook 7H9 broth base; 2 ml glycerol.

100ml ADC supplement added after autoclaving.

Middlebrook-OADC plates (7H10-OADC)

19 g Middlebrook 7H10 agar powder; 5 ml glycerol.

100 ml OADC supplement added after autoclaving.

YPD broth

1.0 g yeast extract, 2.0 g peptone and 2.0 g glucose

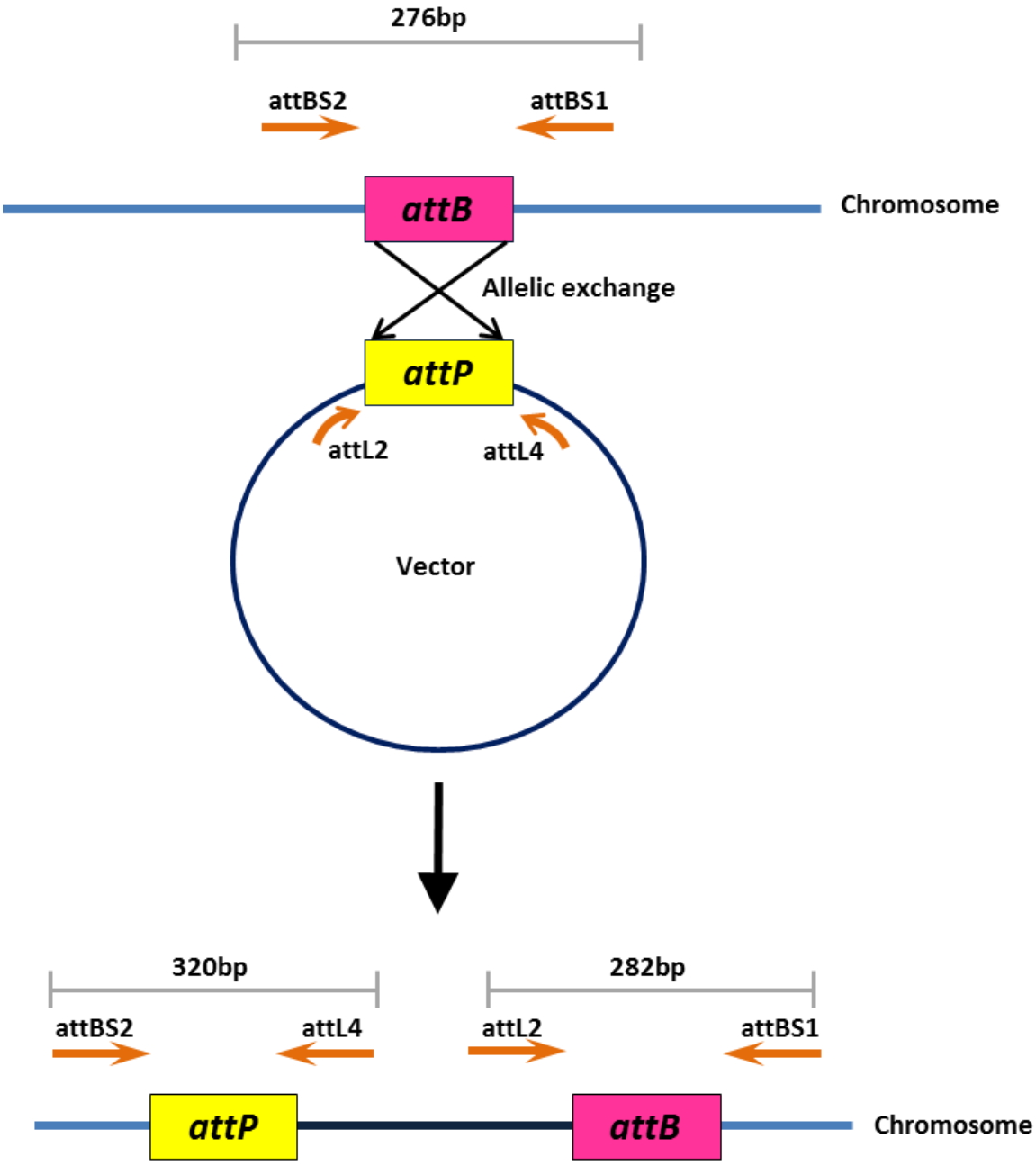
YPDA

YPD plus 0.15% agar both supplemented with 0.003% adenine.

Synthetic Drop out medium (SD)

6.7g yeast nitrogen base without amino acids, 0.2 g glucose, 20 g Bacto agar and 0.2 g dropout mix

APPENDIX B: PCR strategy for *attB* screening



Schematic representation of the PCR strategy that was used to confirm successful integration at the bacterial *att*B locus. The *att*B primers amplify the bacterial attachment site and the *att*L primers amplify the phage attachment site. Site specific recombination of *att*B and *att*P creates the flanking *att*L and *att*BS sites.

APPENDIX C: LIST OF ABBREVIATIONS

3-AT	3-Amino-1,2,4-triazole
A	Adenine
AD	Activation domain
ADE2	Gene encoding phosphoribosylaminoimidazole carboxylase
Ap	Ampicillin
aph	Gene encoding aminoglycoside phosphotransferase
attB	bacterial tRNAGly attachment site
BD	Binding domain
bp	base pair(s)
CFU	Colony forming unit
DCO	Double cross over
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
g	gravitational force
H	Histidine
h	hour
HIS3	Gene encoding imidazoleglycerol-phosphate-dehydratase
HIV	Human immunodeficiency virus
hyg	Gene conferring resistance to hygromycinB
Hyg	Hygromycin B
kbp	kilo base pair(s)
Km	Kanamycin
L	Leucine
LA	Luria-Bertani agar
lacZ	Gene encoding β-galactosidase
LB	Luria-Bertani broth

Ml	millilitre
MMR	Mismatch repair
Mtb	*Mycobacterium tuberculosis*
OADC	Oleic acid- albumin-dextrose catalase supplement for Middlebrook 7H9 and 7H10
OB	Oligonucleotide/oligosaccharide binding
OD_{600}	Optical density at 600 nanometre wavelength
PCR	Polymerase chain reaction
PHP	Polymerase and histidinol phosphatase
R	Resistant/resistance
Rif	Rifampicin
RNA	Ribonucleic acid
rpm	Revolutions per minute
sacB	Gene encoding levansucrase
SCO	Single cross over
SD	synthetic drop-out
SDS	Sodium dodecylsulphate
SSB	Single-stranded DNA-binding
Suc	Sucrose
T	Tryptophan
TB	Tuberculosis
Tris	Tris(hydroxymethyl)aminomethane
Tween	Polyoxyethylene sorbitan monooleate
U	Units
v/v	Volume per volume
WHO	World Health Organization
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside-
Y2H	Yeast-two-hybrid
µg	microgram

SUPPLEMENTARY INFORMATION

```
                                                                                             PHP
Conservation:
Mavium_MaviaA2_DnaE2        1  -------MGWFNGPPSWAEMERVLDSKPRRAGESPAPEP------DGP-LSRGRATYRP-PDEGRAAR--   53
Maviumparatub_MAP3476c      1  -------MGWFNGPPSWAEMERVLDSKPRRAGESAAPEP------DGP-LSRGRATYRP-PDEGRAAR--   53
Mavium_104_MAV4335          1  -------MGWFNGPPSWAEMERVLDSKPRRAGESPAPEP------DGP-LSRGRATYRP-PDEGRAAR--   53
CDC1551_MT3480              1  MFDILWNVGWSNGPPSWAEMERVLNGKPRHAGV---PAFDA--DGDVP-RSRKRGAYQP-PGRERVG---   60
MTB_DnaE2_Rv3370c           1  M-------------------ERVLNGKPRHAGV---PAFDA--DGDVP-RSRKRGAYQP-PGRERVG---   41
Mbovis_Mb3405c              1  M-------GWSNGPPSWAEMERVLNGKPRHAGV---PAFDA--DGDVP-RSRKRGAYQP-PGRERVG---   53
BCG_Pasteur_BCG3442c        1  M-------------------ERVLNGKPRHAGV---PAFDA--DGDVP-RSRKRGAYQP-PGRERVG---   41
Mcanettii_DnaE2             1  M-------------------ERVLDGKPRHAGV---PAFDA--DGDVP-RSRKRGAYQL-PGRERVG---   41
MintA_DnaE2                 1  -------MGWFNGPPSWAEMERVLDSKPRRAGEPAGLPE------DAP-LSRKRATYRP-PGDGRAPRPS   55
Mkansasii_MkanA1_DnaE2      1  -------MGWGNGPPSWAEMERVLDGKPRHAGAPVAAEPADDAGLEGP-WSRQRGTYQP-PENARVR---   58
Mmarinum_MMAR1158           1  -------MGWGNGPPSWAEMERVLDGKPRHAGVPDAAGPTAEAGWDGP-LSRKRETYAPKPDANRVD---   59
Mulcerans_MUL0923           1  -------------------MERVLDGKPRHAGVPDAAGPTAEAGWDGP-LSRKRETYAPKPDANRVD---   47
Myco_sp_Mmcs_1183           1  -------MGFSNG-PTWPEIERVLNGKPRRAGESLREPPGD--GGDSPAWSRKRGAYQA-PQLPRSG---   56
Myco_sp_Mjls_1210           1  -------MGFSNG-PTWPEIERVLNGKPRRAGESLREPPGD--GGDSPAWSRKRGAYQA-PQLPRSG---   56
Mvanbaalenii_Mvan1530       1  -------MGWHTGPPSWTEMERVLSGKPRRAGWPIDQQIGD--GGDSPAWSRKRGEYHA-PEGPGAQE--   58
Mgilvum_Mflv4893            1  -------------------MRRVLEGKPRRAGWPIDAQVGD--GGDSPAWSSKRGQYRA-PESRGPAT--   46
MSMEG_DnaE2_MSMEG_1633      1  -------------------MERVLTSKPRRSGLPLES-PGD--GGDSPAWSRKRGAYEP-PDQARMPA--   45
Mabscessus_MAB3703c         1  ----------------------MLDGR-----LNPHAPPGD--GGDGPAWSRKRQPYEP-PPRERGR---   37
Maviumparatub_MAP1257       1  ----------------------------------------------------------------MNHS--    4
Mavium_104_MAV3224          1  ----------------------------------------------------------------MNHS--    4
MintA_DnaE1                 1  ----------------------------------------------------------------MNHS--    4
Mabscessus_MAB2696c         1  ----------------------------------------------------------------MSSS--    4
Myco_sp_Mmcs_3087           1  ----------------------------------------------------------------MSGSD-    5
Myco_sp_Mjls_3107           1  ----------------------------------------------------------------MSGSD-    5
Mgilvum_Mflv3636            1  ----------------------------------------------------------------MSG---    3
Mvanbaalenii_Mvan2777       1  ----------------------------------------------------------------MSG---    3
MSM_MSMEG3178               1  ----------------------------------------------------------------MSGTD-    5
Mmarinum_MMAR2369           1  ----------------------------------------------------------------MSGS--    4
Mulcerans_MUL1545           1  ----------------------------------------------------------------MSGS--    4
ML_ML1207                   1  ----------------------------------------------------------------MNQS--    4
Mkansasii_MkanA1_DnaE1      1  ----------------------------------------------------------------MSGS--    4
MTB_DnaE1_Rv1547            1  ----------------------------------------------------------------MSGS--    4
Mbovis_Mb1574               1  ----------------------------------------------------------------MSGS--    4
BCG_Pasteur_BCG1600         1  ----------------------------------------------------------------MSGS--    4
Mcanettii_DnaE1             1  ----------------------------------------------------------------MSGS--    4
CDC1551_MT1598              1  ----------------------------------------------------------------MSGS--    4
```


PHP
Conservation:
Mavium_MaviaA2_DnaE2 54 --SS-VPYAELHAHSAFSFLDGVSTPEEMVEEAARLDLRALALTDHDGLYGAVRFAEAAAELDVRTVFGA 120
Maviumparatub_MAP3476c 54 --SS-VPYAELHAHSAFSFLDGASTPEEMVEEAARLDLRALALTDHDGLYGAVRFAEAAAELDVRTVFGA 120
Mavium_104_MAV4335 54 --SS-VPYAELHAHSAFSFLDGASTPEEMVEEAARLDLRALALTDHDGLYGAVRFAEAAAELDVRTVFGA 120
CDC1551_MT3480 61 --SS-VAYAELHAHSAYSFLDGASTPEELVEEAARLGLCALALTDHDGLYGAVRFAEAAAELDVRTVFGA 127
MTB_DnaE2_Rv3370c 42 --SS-VAYAELHAHSAYSFLDGASTPEELVEEAARLGLCALALTDHDGLYGAVRFAEAAAELDVRTVFGA 108
Mbovis_Mb3405c 54 --SS-VAYAELHAHSAYSFLDGASTPEELVEEAARLGLCALALTDHDGLYGAVRFAEAAAELDVRTVFGA 120
BCG_Pasteur_BCG3442c 42 --SS-VAYAELHAHSAYSFLDGASTPEELVEEAARLGLCALALTDHDGLYGAVRFAEAAAELDVRTVFGA 108
Mcanettii_DnaE2 42 --SS-VAYAELHAHSAYSFLDGASTPEELVTEAARLGLRALALTDHDGLYGAVRFAEAAAEVDVRTVFGA 108
MintA_DnaE2 56 AARS-VPYAELHAHSAFSFLDGASTPEELVEEAARLGLRALALTDHDGLYGAVRFAEAAAELDMRTVFGA 124
Mkansasii_MkanA1_DnaE2 59 --SS-AAYAELHAHTAYSFLDGASTPEELVEEAARLDLCALALTDHNGLYGAVRFAEAAAELDMRTVFGA 125
Mmarinum_MMAR1158 60 --SS-IAYAELHAHSAYSFLDGASTPEELVEEAARLDLRALALTDHNGLYGAVRFAEAASGLGVRTVFGA 126
Mulcerans_MUL0923 48 --SS-IAYAELHAHSAYSFLDGASTPEELVEEAARLDLRALALTDHNGLYGAVRFAEAASGLGVRTVFGA 114
Myco_sp_Mmcs_1183 57 --GA-VPYAELHAHSAYSFLDGAGTPEELVEEAARLDLRAIALTDHDGLYGVVRFAEAARELDMRTVFGA 123
Myco_sp_Mjls_1210 57 --GA-VPYAELHAHSAYSFLDGAGTPEELVEEAARLDLRAIALTDHDGLYGVVRFAEAARELDMRTVFGA 123
Mvanbaalenii_Mvan1530 59 --SS-TPYAELHAHSAYSFLDGASTPEELVEEAARLNLRAIALTDHDGLYGVVRFAEAARELDVATVFGA 125
Mgilvum_Mflv4893 47 -GRS-MPYAELHAHSAYSFLDGASTPEELVEEASRLGLRALALTDHDGLYGVVRFAEAAKELDVDTVFGA 114
MSMEG_DnaE2_MSMEG_1633 46 --SA-LPYAELHAHSAYSFLDGASTPEELVEEAARLNLRAIALTDHDGLYGVVRFAEAARELDVATVFGA 112
Mabscessus_MAB3703c 38 --SV-VPYAELHAHSAFSFLDGASLPEEMVQEAARLDLKALAITDHDGFYGVVRFAEAAKELGLPTVFGA 104
Maviumparatub_MAP1257 5 ------SFVHLHNHTEYSMLDGAAKITPMLAEVERLGMPAVGMTDHGNMFGASEFYNAATKAGIKPIIGV 68
Mavium_104_MAV3224 5 ------SFVHLHNHTEYSMLDGAAKITPMLAEVERLGMPAVGMTDHGNMFGASEFYNAATKAGIKPIIGV 68
MintA_DnaE1 5 ------SFVHLHNHTEYSMLDGAAKIAPMLAEVDRLQMPAVGMTDHGNMFGASEFYNAATKAGIKPIIGV 68
Mabscessus_MAB2696c 5 --NS--SFVHLHNHTEYSMLDGAAKIKPMLAEVERLGMSAVGMTDHGNMYGASEFYNVAAATGVKPIIGI 70
Myco_sp_Mmcs_3087 6 GRSS-GSFVHLHNHTEYSMLDGAAKVKPMLAEAQRLEMPAIGMTDHGNMFGASEFYNAATDAGIKPIIGI 74
Myco_sp_Mjls_3107 6 GRSS-GSFVHLHNHTEYSMLDGAAKVKPMLAEAQRLEMPAIGMTDHGNMFGASEFYNAATDAGIKPIIGI 74
Mgilvum_Mflv3636 4 ------SFVHLHNHTEYSMLDGAAKVKPMLAEAQRLEMPAIGMTDHGNMFGASEFYNAATDAGIKPIIGI 67
Mvanbaalenii_Mvan2777 4 ------SFVHLHNHTEYSMLDGAAKVKPMLAEAQRLEMPAIGMTDHGNMFGASEFYNAATDAGIKPIIGI 67
MSM_MSMEG3178 6 GRSS-GSFVHLHNHTEYSMLDGAAKITPMLAEAQRLEMPAIGMTDHGNMFGASEFYNSATKAGIKPIIGI 74
Mmarinum_MMAR2369 5 --SS-RSFVHLHNHTEYSMLDGAAKITPMLAEVERLDMPAIGMTDHGNMFGASEFYNSATKAGVKPIIGV 71
Mulcerans_MUL1545 5 --SS-RSFVHLHNHTEYSMLDGAAKITPMLAEVERLDMPAIGMTDHGNMFGASEFYNSATKAGVKPIIGV 71
ML_ML1207 5 ------SFVHLHNHTEYSMLDGAAKITPMFAEVERLQMPAVGMTDHGNMFGASEFYNTAIKAGIKPIIGV 68
Mkansasii_MkanA1_DnaE1 5 --SS-RSFVHLHNHTEYSMLDGAAKINPMLAEVERLQMPAIGMTDHGNMFGASEFYNSATQAGIKPIIGV 71
MTB_DnaE1_Rv1547 5 --SAGSSFVHLHNHTEYSMLDGAAKITPMLAEVERLGMPAVGMTDHGNMFGASEFYNSATKAGIKPIIGV 72
Mbovis_Mb1574 5 --SAGSSFVHLHNHTEYSMLDGAAKITPMLAEVERLGMPAVGMTDHGNMFGASEFYNSATKAGIKPIIGV 72
BCG_Pasteur_BCG1600 5 --SAGSSFVHLHNHTEYSMLDGAAKITPMLAEVERLGMPAVGMTDHGNMFGASEFYNSATKAGIKPIIGV 72
Mcanettii_DnaE1 5 --SAGSSFVHLHNHTEYSMLDGAAKITPMLAEVERLGMPAVGMTDHGNMFGASEFYNSATKAGIKPIIGV 72
CDC1551_MT1598 5 --SAGSSFVHLHNHTEYSMLDGAAKITPMLAEVERLGMPAVGMTDHGNMFGASEFYNSATKAGIKPIIGV 72


PHP
Conservation:
Mavium_MaviaA2_DnaE2 121 ELSLGPS---ART----------EAPDPPGPHLLVLARGAEGYRRLSRQLAVAHLAGG----EK 167
Maviumparatub_MAP3476c 121 ELSLGPS---ART----------EAPDPPGPHLLVLARGPEGYRRLSRQLAAAHLAGG----EK 167
Mavium_104_MAV4335 121 ELSLGPS---ART----------EAPDPPGPHLLVLARGPEGYRRLSRQLAAAHLAGG----EK 167
CDC1551_MT3480 128 ELSLGAT---ART----------ERPDPPGPHLLVLARGPEGYRRLSRQLAAAHLAGG----EK 174
MTB_DnaE2_Rv3370c 109 ELSLGAT---ART----------ERPDPPGPHLLVLARGPEGYRRLSRQLAAAHLAGG----EK 155
Mbovis_Mb3405c 121 ELSLGAT---ART----------ERPDPPGPHLLVLARGPEGYRRLSRQLAAAHLAGG----EK 167
BCG_Pasteur_BCG3442c 109 ELSLGAT---ART----------ERPDPPGPHLLVLARGPEGYRRLSRQLAAAHLAGG----EK 155
Mcanettii_DnaE2 109 ELSLGASVSAART----------ERPDPPGPHLLVLARGPEGYRRLSRQLAAAHLAGG----EK 158
MintA_DnaE2 125 ELSLGGE---ART----------ERPDPAGPHLLVLARGPEGYRRLSRQLAAAHLAGG----EK 171
Mkansasii_MkanA1_DnaE2 126 ELSLGSE---ART----------EQPDPPGPHLLVLARGPEGYRRLSRQLAAAHLAGG----EK 172
Mmarinum_MMAR1158 127 ELSLGPE---ART----------EQPDPPGPHLLVLARGPEGYRRLSRQLAAAHLAGG----EK 173
Mulcerans_MUL0923 115 ELSLGPE---ART----------EQPDPPGPHLLVLARGPEGYRRLSRQLAAAHLAGG----EK 161
Myco_sp_Mmcs_1183 124 ELSLGMG---ARSARN---------EVPDPPGPHLLVLARGPEGYRRLSREIAKAHLAGG----EK 173
Myco_sp_Mjls_1210 124 ELSLGMG---ARSARN---------EVPDPPGPHLLVLARGPEGYRRLSREIAKAHLAGG----EK 173
Mvanbaalenii_Mvan1530 126 ELSLGGG---TRT----------DVPDPPGPHLLVLARGPEGYRRLSRQLAAAHLAGG----EK 172
Mgilvum_Mflv4893 115 ELSLGGG---TRT----------DVPDPPGPHLLVLARGPEGYRRLSRQLAAAHLAGG----EK 161
MSMEG_DnaE2_MSMEG_1633 113 ELSLSNV---ART----------EDPDPPGPHLLVLARGPEGYRRLSREIAKAHLAGG----EK 159
Mabscessus_MAB3703c 105 ELSLGGQ---GNT----------EDSVHLLVLARGQEGYRRLSRQMAGAHLSGGTPKDRK 151
Maviumparatub_MAP1257 69 EAYIAPG---SRFDTRRILWGDPSQKADDVSGSGSYTHLTMVAENAAGLRNLFKLSSLASFEGQ----LG 131
Mavium_104_MAV3224 69 EAYIAPG---SRFDTRRILWGDPSQKADDVSGSGSYTHLTMVAENAAGLRNLFKLSSLASFEGQ----LG 131
MintA_DnaE1 69 EAYIAPA---SRFDTRRILWGDPGQKSDDVSGSGSYTHMTMVAETPTGLRNLFKLSSLASFEGQ----LG 131
Mabscessus_MAB2696c 71 EAYIAPA---SRFDTKRVFWGDPGQKSDDVSGSGSYTHMTMVAENATGLRNLFKLSSLASFEGQ----LG 133
Myco_sp_Mmcs_3087 75 EAYIAPA---SRFETKRVLWGDPSQKSDDVSGSGSYTHMTMVAENATGLRNLFKLSSLASFEGQ----LG 137
Myco_sp_Mjls_3107 75 EAYIAPA---SRFETKRVLWGDPSQKSDDVSGSGSYTHMTMVAENATGLRNLFKLSSLASFEGQ----LG 137
Mgilvum_Mflv3636 68 EAYIAPA---SRFDTKRVLWGDPGQKSDDVSGSGSYTHMTMVAENATGLRNLFKLSSLASFEGQ----LG 130
Mvanbaalenii_Mvan2777 68 EAYIAPA---SRFETKRVLWGDPSQKSDDVSGGGSYTHMTMVAENATGLRNLFKLSSLASFEGQ----LG 130
MSM_MSMEG3178 75 EAYIAPG---SRFDTKRVTWGDPGQKGDDVSGSGAYTHMTMVAENATGLRNLFKLSSLASFEGQ----LG 137
Mmarinum_MMAR2369 72 EAYVAPG---SRFDTRRITWGDPSQKGDDVSGSGSYTHLTMVAENATGLRNLFKLSSLASFEGQ----LG 134
Mulcerans_MUL1545 72 EAYVAPG---SRFDTRRITWGDPSQKGDDVSGSGSYTHLTMVAENATGLRNLFKLSSLASFEGQ----LG 134
ML_ML1207 69 EAYIAPG---SRFDTRRITWGDPSQKADDVSAGGAYTHLTMVAENAAGLRNLFKLSSLASFEGQ----LS 131
Mkansasii_MkanA1_DnaE1 72 EAYIAPA---SRFDTRRIFWGDPSQKADDISGSGSYTHLTMMAENATGLRNLFKLSSLASFEGQ----LS 134
MTB_DnaE1_Rv1547 73 EAYIAPG---SRFDTRRILWGDPSQKADDVSGSGSYTHLTMMAENATGLRNLFKLSSHASFEGQ----LS 135
Mbovis_Mb1574 73 EAYIAPG---SRFDTRRILWGDPSQKADDVSGSGSYTHLTMMAENATGLRNLFKLSSHASFEGQ----LS 135
BCG_Pasteur_BCG1600 73 EAYIAPG---SRFDTRRILWGDPSQKADDVSGSGSYTHLTMMAENATGLRNLFKLSSHASFEGQ----LS 135
Mcanettii_DnaE1 73 EAYIAPG---SRFDTRRILWGDPSQKADDVSGSGSYTHLTMMAENATGLRNLFKLSSHASFEGQ----LS 135
CDC1551_MT1598 73 EAYIAPG---SRFDTRRILWGDPSQKADDVSGSGSYTHLTMMAENATGLRNLFKLSSHASFEGQ----LS 135


PHP
Conservation:
Mavium_MaviaA2_DnaE2 168 GKPRYDLDALTEAAGGHWHILTGCRKGHVRQALSDGGPDAAARALADLVDRFGAARVSIELTRHGQPLDD 237
Maviumparatub_MAP3476c 168 GKPRYDLDALTEAAGGHWHILTGCRKGHVRQALSDGGPDAAARALADLVDRFGAARVSIELTRHGQPLDD 237
Mavium_104_MAV4335 168 GKPRYDLDALTEAAGGHWHILTGCRKGHVRQALSEGGPDEAARALADLVDRFGAARVSIELTRHGQPLDD 237
CDC1551_MT3480 175 GKPRYDFDALTEAAGGHWHILTGCRKGHVRQALSQGGPAAAQRALADLVDRFTPSRVSIELTHHGHPLDD 244
MTB_DnaE2_Rv3370c 156 GKPRYDFDALTEAAGGHWHILTGCRKGHVRQALSQGGPAAAQRALADLVDRFTPSRVSIELTHHGHPLDD 225
Mbovis_Mb3405c 168 GKPRYDFDALTEAAGGHWHILTGCRKGHVRQALSQGGPAAAQRALADLVDRFTPSRVSIELTHHGHPLDD 237
BCG_Pasteur_BCG3442c 156 GKPRYDFDALTEAAGGHWHILTGCRKGHVRQALSQGGPAAAQRALADLVDRFTPSRVSIELTHHGHPLDD 225
Mcanettii_DnaE2 159 GKPRYDFDALTEAAGGHWHILTGCRKGHVRQALSQGGPAAAQRALADLVDRFTPSRVSIELTHHGHPLDD 228
MintA_DnaE2 172 GKPRYDFDALTEAAGGHWHILTGCRKGHVRQALSDGGPDAAERALADLVDRFGAQSVSVELTHHGQPLDD 241
Mkansasii_MkanA1_DnaE2 173 GKPRYDFDALTEAAGGHWHILTGCRKGHVRQALSEGGPDAAQRALADLVDRFGAYRVSIELTHHGHPLDD 242
Mmarinum_MMAR1158 174 GKPRYDFDSLTEAAGGHWHILTGCRKGHVRQALRQGGPDAAERALADLVDRFTAGRVSVELTHHGHPLDD 243
Mulcerans_MUL0923 162 GKPRYDFDSLTEAAGGHWHILTGCRKGHVRQALRQGGPDAAERVLADLVDRFTAGRVSVELTHHGHPLDD 231
Myco_sp_Mmcs_1183 174 GKPRYDYDTLTEAAGGHWHILTGCRKGHVRQALSEGGPAAAEKALADLVDRFGRDRVSVELTHHGHPLDD 243
Myco_sp_Mjls_1210 174 GKPRYDYDALTEAAGGHWHILTGCRKGHVRQALSEGGPAAAEKALADLVDRFGRDRVSVELTHHGHPLDD 243
Mvanbaalenii_Mvan1530 173 GVLRYDFDALTEAAGGHWQILTGCRKGHVRQALSTGGPEAAEAALADLVDRFGPDRVTVELTHHGHPLDD 242
Mgilvum_Mflv4893 162 GVLRYDFDALTEAAGGHWQILTGCRKGHVRQALLRGGDSAAEAALADLVDRFGRERVTVELTHHGHPLDD 231
MSMEG_DnaE2_MSMEG_1633 160 GRPRYDFDQLTEAAGGHWHILTGCRKGHVRQALSDGGPDAAAAALADLVDRFGADRVSIELTHHGHPCDD 229
Mabscessus_MAB3703c 152 GKPRYDLDALTEAAGGHWHILTGCRKGHVRRALAGGGPVAAERALANLVDRFGADRVSIELNRHGHPGED 221
Maviumparatub_MAP1257 132 KWSRMDAELIAEHADG-IIATTGCPSGEVQTRLRLGQDREALESAAKWREIFGADNFFLELMDHGLSIEQ 200
Mavium_104_MAV3224 132 KWSRMDAELIAEHADG-IIATTGCPSGEVQTRLRLGQDREALESAAKWREIFGADNFFLELMDHGLSIEQ 200
MintA_DnaE1 132 KWSRMDAEIIAEHAEG-IIATTGCPSGEVQTRLRLGHDREALESAAKWREIFGADNYFLELMDHGLSIEQ 200
Mabscessus_MAB2696c 134 KWARMDADLIAEHASG-IIATTGCPSGEVQTRLRLGQESEALAAAAKWQDIFGKDNFFLELMDHGLSIER 202
Myco_sp_Mmcs_3087 138 KWSRMDAEIIAEHAEG-IIATTGCPSGEVQTRLRLGHQREALEAAAKWREIFGPQNFFLELMDHGLDIER 206
Myco_sp_Mjls_3107 138 KWSRMDAEIIAEHAEG-IIATTGCPSGEVQTRLRLGHQREALEAAAKWREIFGPQNFFLELMDHGLDIER 206
Mgilvum_Mflv3636 131 KWSRMDAEIIAEHAEG-IIATTGCPSGEVQTRLRLGHDREALEAAARWREIFGPENFFLELMDHGLDIER 199
Mvanbaalenii_Mvan2777 131 KWARMDAEIIAEHAEG-IIATTGCPSGEVQTRLRLGHDREALEAAARWREIFGPENFFLELMDHGLDIER 199
MSM_MSMEG3178 138 KWSRMDAEIIAEHAEG-IIATTGCPSGEVQTRLRLGHEREALEAAAKWREIFGPENFFLELMDHGLDIER 206
Mmarinum_MMAR2369 135 KWSRMDAELVAEHAEG-IIATTGCPSGEVQTRLRLGHDREALESAAKWREIFGADNYFLELMDHGLSIER 203
Mulcerans_MUL1545 135 KWSRMDAELVAEHAEG-IIATTGCPSGEVQTRLRLGHDREALESAAKWREIFGADNYFLELMDHGLSIER 203
ML_ML1207 132 KWSRMDAELIGEYAEG-IIVTTGCPSGEVQTRLRLGHDREALESAAKWREIVGPDNYFLELMDHGLSIEQ 200
Mkansasii_MkanA1_DnaE1 135 KWSRMDAELIAEHAEG-IIITTGCPSGEVQTRLRLGQEREALEAAARWREIVGPENYFLELMDHGLTIER 203
MTB_DnaE1_Rv1547 136 KWSRMDAELIAEHAEG-IIITTGCPSGEVQTRLRLGQDREALEAAAKWREIVGPDNYFLELMDHGLTIER 204
Mbovis_Mb1574 136 KWSRMDAELIAEHAEG-IIITTGCPSGEVQTRLRLGQDREALEAAAKWREIVGPDNYFLELMDHGLTIER 204
BCG_Pasteur_BCG1600 136 KWSRMDAELIAEHAEG-IIITTGCPSGEVQTRLRLGQDREALEAAAKWREIVGPDNYFLELMDHGLTIER 204
Mcanettii_DnaE1 136 KWSRMDAELIAEHAEG-IIITTGCPSGEVQTRLRLGQDREALEAAAKWREIVGPDNYFLELMDHGLTIER 204
CDC1551_MT1598 136 KWSRMDAELIAEHAEG-IIITTGCPSGEVQTRLRLGQDREALEAAAKWREIVGPDNYFLELMDHGLTIER 204


PHP
Conservation:
Mavium_MaviaA2_DnaE2 238 ERNAALAALAPRFGVGVVATTGAHFAGPSRRRLAMAVGAIRARESLDSAAGWLAPLGGSHLRSGEEMARL 307
Maviumparatub_MAP3476c 238 ERNAALAALAPRFGVGVVATTGAHFAGPSRRRLAMAMGAIRARESLDSAAGWLAPLGGSHLRSGAEMARL 307
Mavium_104_MAV4335 238 ERNAALAALAPRFGVGVVATTGAHFAGPSRRRLAMAVGAIRARESLDSAAGWLAPLGGSHLRSGEEMARL 307
CDC1551_MT3480 245 ERNAALAGLAPRFGVGIVATTGAHFADPSRGRLAMAMAAIRARRSLDSAAGWLAPLGGAHLRSGEEMARL 314
MTB_DnaE2_Rv3370c 226 ERNAALAGLAPRFGVGIVATTGAHFADPSRGRLAMAMAAIRARRSLDSAAGWLAPLGGAHLRSGEEMARL 295
Mbovis_Mb3405c 238 ERNAALAGLAPRFGVGIVATTGAHFADPSRGRLAMAMAAIRARRSLDSAAGWLAPLGGAHLRSGEEMARL 307
BCG_Pasteur_BCG3442c 226 ERNAALAGLAPRFGVGIVATTGAHFADPSRGRLAMAMAAIRARRSLDSAAGWLAPLGGAHLRSGEEMARL 295
Mcanettii_DnaE2 229 ERNAALAGLAPRFGVGIVATTGAHFADPSRGRLAMAMAAIRARRSLDSAAGWLAPLGGAHLRSGEEMARL 298
MintA_DnaE2 242 ERNAVLAALAPRFGVGVVATTGAHFAGPSRRRLAMAMGAIRARESLDSAAGWLAPLGGSHLRSGAEMARL 311
Mkansasii_MkanA1_DnaE2 243 ERNATLAGLAPRFGVGVVATTGAHFAHPSRSRLAMAMGAIRARQSLDSAAGWLAPLGGSHLRSGDEMARM 312
Mmarinum_MMAR1158 244 ENNAMLAGLAPRFGVGVVATTGAHFAHPARSRLAMAMGAIRARQSLDSAAGWLAPLGGSHLRSGAEMARL 313
Mulcerans_MUL0923 232 ENNAMLAGLAPRFGVGVVATTGAHFAHPARSRLAMAMGAIRARQSLDSAAGWLAPLGGSHLRSGAEMARL 301
Myco_sp_Mmcs_1183 244 ERNAALAELAPRFGLTVVATTAAHFAEPSRGRLAMAMGAIRARHSIDEAAGYLAPLGGSHLRSGEEMARL 313
Myco_sp_Mjls_1210 244 ERNAALAELAPRFGLTVVATTAAHFAEPSRGRLAMAMGAIRARHSIDEAAGYLAPLGGSHLRSGEEMARL 313
Mvanbaalenii_Mvan1530 243 ERNAALAALAPRFGLGVVATTAAHFAEPARGRLAMAMGAIRARNSIDEAAGYLAPLGGSHLRSGDEMARM 312
Mgilvum_Mflv4893 232 ERNAALAALAPRFGLTVIATTAAHFAEPSRGRLAMAMGAIRARNSIDEAAGYLAPLGGSHLRSGEEMARV 301
MSMEG_DnaE2_MSMEG_1633 230 ERNAALAELAPRFGLGVVATTAAHFATPSRGRLAMAMAAIRARNSIDTAAGWLAPLGGVHLRSGEEMARL 299
Mabscessus_MAB3703c 222 ERNAELAALARGFGVGVIATTAAHFATPESGRLAMAMAAVRARKSLDDAAGWLDPVGGAHLRSGDEMARL 291
Maviumparatub_MAP1257 201 RVRDGLLEIGRKLNIPPLATNDCHYITRDAAHNHEALLCVQTGKTLSDPNRFKFDGDGYYLKSAAEMRQI 270
Mavium_104_MAV3224 201 RVRGGLLEIGRKLNIPPLATNDCHYVTRDAAHNHEALLCVQTGKTLSDPNRFKFDGDGYYLKSAAEMRQI 270
MintA_DnaE1 201 RVRDGLLEIGRKLGIPPLATNDCHYVTRDAAHNHEALLCVQTGKTLSDPNRFKFDGDGYYLKSAAEMRRI 270
Mabscessus_MAB2696c 203 RVREGLLEVGKKLGIPPLATNDCHYVTKDHAHNHEALLCVQTGKTLSDPTRFKFDGDGYYLKPASEMRDL 272
Myco_sp_Mmcs_3087 207 RVREGLLEIGQKLGIPPLATNDCHYVTRDASQNHEALLCIQTGKTLSDPTRFKFDGDGYYLKSAAEMRAL 276
Myco_sp_Mjls_3107 207 RVREGLLEIGQKLGIPPLATNDCHYVTRDASQNHEALLCIQTGKTLSDPTRFKFDGDGYYLKSAAEMRAL 276
Mgilvum_Mflv3636 200 RVREGLLEIGQKLGIPPLATNDCHYVTRDASQNHEALLCIQTGKTLSDPNRFKFDGDGYYLKSAAEMRAL 269
Mvanbaalenii_Mvan2777 200 RVREGLLDIGQKLGIPPLATNDCHYVTRDAAQNHEALLCIQTGKTLSDPNRFKFDGDGYFLKSAAEMRAL 269
MSM_MSMEG3178 207 RVREGLLEIGRKLGIPPLATNDCHYVTREAARNHEALLCVQTGKTLSDPTRFKFDGDGYFLKSAEEMRAL 276
Mmarinum_MMAR2369 204 RVREGLIEIGRKLDIPPLATNDCHYVTRDAAHNHEALLCVQTGKTLSDPNRFKFDGDGYYLKSAAEMRQL 273
Mulcerans_MUL1545 204 RVREGLIEIGRKLDIPPLATNDCHYVTRDAAHNHEALLCVQTGKTLSDPNRFKFDGDGYYLKSAAEMRQL 273
ML_ML1207 201 RVREGLLNIGRKLNIPPLATNDCHYVTRDAVHNHEALLCVQTGKTLSDPNRFKFHGDGYYLKSAAEMRQL 270
Mkansasii_MkanA1_DnaE1 204 RVREGLLHIGRTLNIPPLATNDCHYVTRDAAHNHEALLCVQTGKTLSDPNRFKFDGDGYYLKSAAEMRQI 273
MTB_DnaE1_Rv1547 205 RVRDGLLEIGRALNIPPLATNDCHYVTRDAAHNHEALLCVQTGKTLSDPNRFKFDGDGYYLKSAAEMRQI 274
Mbovis_Mb1574 205 RVRDGLLEIGRALNIPPLATNDCHYVTRDAAHNHEALLCVQTGKTLSDPNRFKFDGDGYYLKSAAEMRQI 274
BCG_Pasteur_BCG1600 205 RVRDGLLEIGRALNIPPLATNDCHYVTRDAAHNHEALLCVQTGKTLSDPNRFKFDGDGYYLKSAAEMRQI 274
Mcanettii_DnaE1 205 RVRDGLLEIGRALNIPPLATNDCHYVTRDAAHNHEALLCVQTGKTLSDPNRFKFDGDGYYLKSAAEMRQI 274
CDC1551_MT1598 205 RVRDGLLEIGRALNIPPLATNDCHYVTRDAAHNHEALLCVQTGKTLSDPNRFKFDGDGYYLKSAAEMRQI 274


```
Conservation:
Mavium_MaviaA2_DnaE2          308  FA-WCPEAVTAAAELGEQCAFGLALI--APRLPPFDVPDGHTEDSWLRQLTMTGA-RDRYG---SPERAP  370
Maviumparatub_MAP3476c        308  FA-WRPQAVTAAAELGEQCAFGLALI--APRLPPFDVPDGHTEDSWLRQLTMTGA-RDRYG---SPEHAP  370
Mavium_104_MAV4335            308  FA-WCPEAVTAAAELGEQCAFGLALI--APRLPPFDVPDGHTEDSWLRQLTMTGA-RDRYG---SPEHAP  370
CDC1551_MT3480                315  FA-WCPEAVTAAAELGERCAFGLQLI--APRLPPFDVPDGHTEDSWLRSLVMAGA-RERYG---PPKSAP  377
MTB_DnaE2_Rv3370c             296  FA-WCPEAVTAAAELGERCAFGLQLI--APRLPPFDVPDGHTEDSWLRSLVMAGA-RERYG---PPKSAP  358
Mbovis_Mb3405c                308  FA-WCPEAVTAAAELGERCAFGLQLI--APRLPPFDVPDGHTEDSWLRSLVMAGA-RERYG---PPKSAP  370
BCG_Pasteur_BCG3442c          296  FA-WCPEAVTAAAELGERCAFGLQLI--APRLPPFDVPDGHTEDSWLRSLVMAGA-RERYG---PPKSAP  358
Mcanettii_DnaE2               299  FA-WCPEAVTAAAELGERCAFGLQLI--APRLPPFDVPDGHTEDSWLRSLVLAGA-RERYG---PPESAP  361
MintA_DnaE2                   312  FA-GGPDVVTAAAELGEQCAFGLALI--APQLPPFDVPDGHTEDSWLRQLTMAGA-RDRYG---P--NAP  372
Mkansasii_MkanA1_DnaE2        313  FA-WRPEAVTAAAELGGQCAFELALI--APQLPPFDVPGGHTEDSWLRQLVMAGA-RHRYG---PPDGAP  375
Mmarinum_MMAR1158             314  FA-QRPDVVTAAAELGEQCAFGLALI--APRLPPFDVPDGHTEDSWLRSLVMSGA-RERYG---STERSP  376
Mulcerans_MUL0923             302  FA-QRPDVVTAAAELGEQCAFGLALI--APRLPPFDVPDGHTEDSWLRSLVMSGA-RERYG---STERSP  364
Myco_sp_Mmcs_1183             314  FA-HCPQVVTAAADLGEQCAFGLALI--APKLPPFDVPAGHTEDSWLRHLVMLGA-RNRYG---PPERAP  376
Myco_sp_Mjls_1210             314  FA-HCPQVVTAAADLGEQCAFGLALI--APKLPPFDVPAGHTEDSWLRHLVMLGA-RNRYG---PPERAP  376
Mvanbaalenii_Mvan1530         313  FA-HCPEVVTAAAELGEQCAFGLALI--APQLPPFDVPDGHTESSWLRHLVMQGA-RERYG---PPERAS  375
Mgilvum_Mflv4893              302  FA-HCPEVVTAAADLGEQCAFGLALI--APQLPPFEVPAGHTENSWLRHLVMQGA-RERYG---PPERAS  364
MSMEG_DnaE2_MSMEG_1633        300  FD---PEFVAAAADLGEQCAFGLALI--APQLPPFDVPDGHTEDSWLRHLAMAGA-ARRYG---PPERAP  360
Mabscessus_MAB3703c           292  FS-HYPEVVTAAAELGEECAFDLRLI--APQLPPFDVPSGHTEDSWLRHLALEGA-ARRYG---PRAGAQ  354
Maviumparatub_MAP1257         271  WDAEVPGACDSTLLIAERVQSYAEVWTPRDRMPVFPVPEGHDQASWLHHEVMAGL-RRRFP----DGVGQ  335
Mavium_104_MAV3224            271  WDAEVPGACDSTLLIAERVQSYAEVWTPRDRMPVFPVPEGHDQASWLHHEVMAGL-RRRFP----DGVGQ  335
MintA_DnaE1                   271  WDDEVPGACDSTLLIAERVQPYDEVWAPRDRMPIFPVPEGHDQASWLHHEVMAGL-ARRFP----SGVGQ  335
Mabscessus_MAB2696c           273  WDDAVPGACDNTLLIAERVQSYEDVWNFKDRMPIFPVPEGEDQDSWLRKEVDRGL-ERRFEG-VPGGVPD  340
Myco_sp_Mmcs_3087             277  WDSQVPGACDSTLLIAERVQSYADVWAPRDRMPIFPVPEGHDQASWLHHEVMAGL-KRRFSAVSGGVVPN  345
Myco_sp_Mjls_3107             277  WDSQVPGACDSTLLIAERVQSYADVWAPRDRMPIFPVPEGHDQASWLHHEVMAGL-KRRFSAVSGGVVPN  345
Mgilvum_Mflv3636              270  WDGQVPGACDSTLLIAERVQPYTDVWAPKDRMPVFPVPEGHDQGSWLTHEVMAGL-ERRFSAAAGGQVPP  338
Mvanbaalenii_Mvan2777         270  WDGQVPGACDSTLLIAERVQSYADVWTPKDRMPIFPVPEGHDQGSWLTHEVMTGLQERRFH---GAPVPQ  336
MSM_MSMEG3178                 277  WDSQVPGACDSTLLIGERVQSYADVWEPRDRMPVFPVPEGHDQASWLTHEVKAGL-ERRFR---GGPVPE  342
Mmarinum_MMAR2369             274  WDDDIPGACDSTLLIAERVQSYADVWTPRDRMPIFPVPEGHDPASWLHHEVMAGL-RRRFP----SGLGQ  338
Mulcerans_MUL1545             274  WDDDIPGACDSTLLIAERVQSYADVWTPRDRMPIFPVPEGHDPASWLHHEVMAGL-RRRFP----SGLGQ  338
ML_ML1207                     271  WDDEVPGACDSTLLIAERVQSYADVWEPRNRMPVFPVPVGHDQASWLRHEVDAGL-KRRFP----DGPPN  335
Mkansasii_MkanA1_DnaE1        274  WDDEVPGACDSTLLIAERVQSYADVWTPRDRMPVFPVPDGHDPASWLRHEVAAGL-RRRFP----DGPPA  338
MTB_DnaE1_Rv1547              275  WDDEVPGACDSTLLIAERVQSYADVWTPRDRMPVFPVPDGHDQASWLRHEVDAGL-RRRFP----AGPPD  339
Mbovis_Mb1574                 275  WDDEVPGACDSTLLIAERVQSYADVWTPRDRMPVFPVPDGHDQASWLRHEVDAGL-RRRFP----AGPPD  339
BCG_Pasteur_BCG1600           275  WDDEVPGACDSTLLIAERVQSYADVWTPRDRMPVFPVPDGHDQASWLRHEVDAGL-RRRFP----AGPPD  339
Mcanettii_DnaE1               275  WDDEVPGACDSTLLIAERVQSYADVWTPRDRMPVFPVPDGHDQASWLRHEVDAGL-RRRFP----AGPPD  339
CDC1551_MT1598                275  WDDEVPGACDSTLLIAERVQSYADVWTPRDRMPVFPVPDGHDQASWLRHEVDAGL-RRRFP----AGPPD  339
```

		Palm	
Conservation:			
Mavium_MaviaA2_DnaE2	371	R---AYAQIEHELKVIAQLQFPGYFLVVHDIARFCRENNI-LCQGRGSAANSAVCYALGVTAVDPVANEL	436
Maviumparatub_MAP3476c	371	R---AYAQIEHELKVIAQLQFPGYFLVVHDIARFCRENNI-LCQGRGSAANSAVCYALGVTAVDPVANEL	436
Mavium_104_MAV4335	371	R---AYAQIEHELKVIAQLQFPGYFLVVHDIARFCRENNI-LCQGRGSAANSAVCYALGVTAVDPVANEL	436
CDC1551_MT3480	378	R---AYSQIEHELKVIAQLRFPGYFLVVHDITRFCRDNDI-LCQGRGSAANSAVCYALGVTAVDPVANEL	443
MTB_DnaE2_Rv3370c	359	R---AYSQIEHELKVIAQLRFPGYFLVVHDITRFCRDNDI-LCQGRGSAANSAVCYALGVTAVDPVANEL	424
Mbovis_Mb3405c	371	R---AYSQIEHELKVIAQLRFPGYFLVVHDITRFCRDNDI-LCQGRGSAANSAVCYALGVTAVDPVANEL	436
BCG_Pasteur_BCG3442c	359	R---AYSQIEHELKVIAQLRFPGYFLVVHDITRFCRDNDI-LCQGRGSAANSAVCYALGVTAVDPVANEL	424
Mcanettii_DnaE2	362	R---AYSQIEHELKVIAQLRFPGYFLVVHDITRFCRDNDI-LCQGRGSAANSAVCYALGVTAVDPVANEL	427
MintA_DnaE2	373	R---AYAQIEHELKVIAQLQFPGYFLVVHDIARFCRDNNI-LCQGRGSAANSAVCYALGVTAVDPVANEL	438
Mkansasii_MkanA1_DnaE2	376	R---AYAQIGHELKVIAQLKFPGYFLVVHDITQFCRRNNI-LCQGRGSAANSAVCYALGITAVDPVANEL	441
Mmarinum_MMAR1158	377	RARKAYTQIEHELKVIAQLSFPGYFLVVHDITQFCRRNDI-LCQGRGSAANSAVCYALGVTAVDPIANEL	445
Mulcerans_MUL0923	365	RARKAYAQIEHELKVIAQLSFPGYFLVVHDITQFCRRNDI-LCQGRGSAANSAVCYALGVTAVDPIANEL	433
Myco_sp_Mmcs_1183	377	Q---AYAQIEHELRVIEKLSFPGYFLVVHDITRFCKENGI-LAQGRGSAANSAVCYALGVTNVDPVANEL	442
Myco_sp_Mjls_1210	377	Q---AYAQIEHELRVIEKLSFPGYFLVVHDITRFCKENGI-LAQGRGSAANSAVCYALGVTNVDPVANEL	442
Mvanbaalenii_Mvan1530	376	R---AYAQIEHELAVIEQLNFPGYFLVVHDITRFCRDNNI-LSQGRGSAANSAVCYALKVTNVDPIANDL	441
Mgilvum_Mflv4893	365	R---AYAQIEHELRVIEQLNFPGYFLVVHDITRFCRENAI-LSQGRGSAANSAVCYALKVTNVDPIANGL	430
MSMEG_DnaE2_MSMEG_1633	361	K---AYAQIEHELRIIEQLRFPGYFLVVHDITQFCRDNNI-LCQGRGSAANSAVCYALGVTNVDPIANDL	426
Mabscessus_MAB3703c	355	K---AYAQIERELEIIAQLNFPGYFLVVHDITQFCRRSDI-LCQGRGSAANSAVCYALGVTNVDPVDNGL	420
Maviumparatub_MAP1257	336	D---YIDRAEYEIKVICDKGFPSYFLIVADLINYARSVDIRVGPGRGSAAGSLVAYALGITNIDPIPHGL	402
Mavium_104_MAV3224	336	D---YIDRAEYEIKVICDKGFPSYFLIVADLINYARSVDIRVGPGRGSAAGSLVAYALGITNIDPIPHGL	402
MintA_DnaE1	336	D---YIDRAEYEIKVICDKGFPSYFLIVADLINYARSVDIRVGPGRGSAAGSLVAYALGITNIDPIPHGL	402
Mabscessus_MAB2696c	341	E---YFTRAHYELDVIKQKGFPAYFLVVGDLVSHAKEVGIRVGPGRGSAAGSLVAYALGITNIDPIPHGL	407
Myco_sp_Mmcs_3087	346	D---YIERAEYEIKVICDKGFPSYFLIVADLINYAKSVDIRVGPGRGSAAGSLVAYALGITNIDPIPHGL	412
Myco_sp_Mjls_3107	346	D---YIERAEYEIKVICDKGFPSYFLIVADLINYAKSVDIRVGPGRGSAAGSLVAYALGITNIDPIPHGL	412
Mgilvum_Mflv3636	339	E---YIERADYEIKVICDKGFPAYFLIVADLINYAKSVGIRVGPGRGSAAGSLVAYAMGITNIDPIPHGL	405
Mvanbaalenii_Mvan2777	337	E---YIDRAQYEIKVICDKGFPAYFLIVADLINYARSVGIRVGPGRGSAAGSLVAYAMGITNIDPIPHGL	403
MSM_MSMEG3178	343	E---YTTRAEYEIKVICDKGFPSYFLIVADLINYARSVGIRVGPGRGSAAGSLVAYALGITNIDPIPHGL	409
Mmarinum_MMAR2369	339	D---YIDRAEYEIKVICDKGFPSYFLIVADLINYARSVDIRVGPGRGSAAGSLVAYAMGITNIDPIPHGL	405
Mulcerans_MUL1545	339	D---YIDRAEYEIKVICDKGFPSYFLIVADLINYARSVDIRVGPGRGSAAGSLVAYAMGITNIDPIPHGL	405
ML_ML1207	336	G---YVERAAYEIDVICDKGFPAYFLIVADLVNHARSVGIRVGPGRGSAAGSLAAYALGITDIDPIPHGL	402
Mkansasii_MkanA1_DnaE1	339	S---YVTRADYEIDVICAKGFPSYFLIVADLIGYAKSVGIRVGPGRGSAAGSLVAYALGITDIDPIPHGL	405
MTB_DnaE1_Rv1547	340	G---YRERAAYEIDVICSKGFPSYFLIVADLISYARSAGIRVGPGRGSAAGSLVAYALGITDIDPIPHGL	406
Mbovis_Mb1574	340	G---YRERAAYEIDVICSKGFPSYFLIVADLISYARSAGIRVGPGRGSAAGSLVAYALGITDIDPIPHGL	406
BCG_Pasteur_BCG1600	340	G---YRERAAYEIDVICSKGFPSYFLIVADLISYARSAGIRVGPGRGSAAGSLVAYALGITDIDPIPHGL	406
Mcanettii_DnaE1	340	G---YRERAAYEIDVICSKGFPSYFLIVADLISYARSAGIRVGPGRGSAAGSLVAYALGITDIDPIPHGL	406
CDC1551_MT1598	340	G---YRERAAYEIDVICSKGFPSYFLIVADLISYARSAGIRVGPGRGSAAGSLVAYALGITDIDPIPHGL	406


```
Conservation:
Mavium_MaviaA2_DnaE2          437  LFERFLSPARDGPPDIDMDIESDQREKVIQYVYDRYGRDYAAQVANVITYRGKIAVRDMARALG--YSQG  504
Maviumparatub_MAP3476c        437  LFERFLSPARDGPPDIDMDIESDQREKVIQYVYDRYGRDYAAQVANVITYRGKIAVRDMARALG--YSQG  504
Mavium_104_MAV4335            437  LFERFLSPARDGPPDIDMDIESDQREKVIQYVYDRYGRDYAAQVANVITYRGKIAVRDMARALG--FSQG  504
CDC1551_MT3480                444  LFERFLSPARDGPPDIDIDIESDQREKVIQYVYHKYGRDYAAQVANVITYRGRSAVRDMARALG--FSPG  511
MTB_DnaE2_Rv3370c             425  LFERFLSPARDGPPDIDIDIESDQREKVIQYVYHKYGRDYAAQVANVITYRGRSAVRDMARALG--FSPG  492
Mbovis_Mb3405c                437  LFERFLSPARDGPPDIDIDIESDQREKVIQYVYHKYGRDYAAQVANVITYRGRSAVRDMARALG--FSPG  504
BCG_Pasteur_BCG3442c          425  LFERFLSPARDGPPDIDIDIESDQREKVIQYVYHKYGRDYAAQVANVITYRGRSAVRDMARALG--FSPG  492
Mcanettii_DnaE2               428  LFERFLSPARDGPPDIDIDIESDQREKVIQYVYDKYGRDYAAQVANVITYRGRSAVRDMARALG--FSPG  495
MintA_DnaE2                   439  LFERFLSPARDGPPDIDMDIESDQREKVIQYVYDKYGRDYAAQVANVITYRGKIAVRDMARALG--FSQG  506
Mkansasii_MkanA1_DnaE2        442  LFERFLSPARDGPPDIDIDIESDRREKVIQYVYDKYGRDYAAQVANVITYRGRSAVRDMARALG--FSQG  509
Mmarinum_MMAR1158             446  LFERFLSPERDGPPDIDIDIESDQREKVIQYVYNKYGRDYAAQVANVITYRGRSAVRDMARALG--FSQG  513
Mulcerans_MUL0923             434  LFERFLSPERDGPPDIDIDIESDQREKVIQYVYNKYGRDYAAQVANVITYRGRSAVRDMARALG--FSQG  501
Myco_sp_Mmcs_1183             443  LFERFLSPARDGPPDIDIDIESDLREQAIQYVYDRYGRDYAAQVANVITYRGRSAIRDMARALG--FSQG  510
Myco_sp_Mjls_1210             443  LFERFLSPARDGPPDIDIDIESDLREQAIQYVYDRYGRDYAAQVANVITYRGRSAIRDMARALG--FSQG  510
Mvanbaalenii_Mvan1530         442  LFERFLSPARDGPPDIDIDIESDLRENAIQYVYQRYGREYAAQVANVITYRGRSAVRDMARALG--FSQG  509
Mgilvum_Mflv4893              431  LFERFLSPARDGPPDIDIDIESDLREKAIQYVYERYGREYAAQVANVITYRGRSAVRDMARALG--FSQG  498
MSMEG_DnaE2_MSMEG_1633        427  LFERFLSPARDGPPDIDIDIESDLRENVIQYVYERYGRDYAAQVANVITYRGRSAIRDMARALG--FSQG  494
Mabscessus_MAB3703c           421  LFERFLSPARDGPPDIDIDIESDEREQAIQYVYNKYGREYAAQVANVITYRGRMAVRDMAKALG--FAQG  488
Maviumparatub_MAP1257         403  LFERFLNPERPSAPDIDIDFDDRRRGEMVRYAADKWGSDRVAQVITFGTIKTKAALKDSARIHYGQPGFA  472
Mavium_104_MAV3224            403  LFERFLNPERPSAPDIDIDFDDRRRGEMVRYAADKWGSDRVAQVITFGTIKTKAALKDSARIHYGQPGFA  472
MintA_DnaE1                   403  LFERFLNPERPSAPDIDIDFDDRRRGEMVRYAADKWGSDRVAQVITFGTIKTKAALKDSARIHYGQPGFA  472
Mabscessus_MAB2696c           408  LFERFLNPERPSAPDIDIDFDDRRRGEMVRYATEKWGSDRVAQVITFGTIKTKAALKDSARIHYGQPGFA  477
Myco_sp_Mmcs_3087             413  LFERFLNPERPSAPDIDIDFDDRRRGEMLRYAANKWGSDRVAQVITFGTIKTKAALKDSARVHYGQPGFA  482
Myco_sp_Mjls_3107             413  LFERFLNPERPSAPDIDIDFDDRRRGEMLRYAANKWGSDRVAQVITFGTIKTKAALKDSARVHYGQPGFA  482
Mgilvum_Mflv3636              406  LFERFLNPERPSAPDIDIDFDDRRRGEMLRYAANKWGSDRVAQVITFGTIKTKAALKDSARVHYGQPGFA  475
Mvanbaalenii_Mvan2777         404  LFERFLNPERPSAPDIDIDFDDRRRGEMLRYAANKWGSDRVAQVITFGTIKTKAALKDSARVHYGQPGFA  473
MSM_MSMEG3178                 410  LFERFLNPERPSAPDIDIDFDDRRRGEMLRYAANRWGSERVAQVITFGTIKTKAALKDSARVNFGQPGFA  479
Mmarinum_MMAR2369             406  LFERFLNPERPSAPDIDIDFDDRRRGEMVRYAADKWGQDRVAQVITFGTIKTKAALKDSARIHYGQPGFA  475
Mulcerans_MUL1545             406  LFERFLNPERPSAPDIDIDFDDRRRGEMVRYAADKWGQDRVAQVITFGTIKTKAALKDSARIHYGQPGFA  475
ML_ML1207                     403  LFERFLNPERTSMPDIDIDFDDRRRGEMVRYAADKWGHDRVAQVITFGTIKTKAALKDSARIHYGQPGFA  472
Mkansasii_MkanA1_DnaE1        406  LFERFLNPERTSMPDIDIDFDDRRRGEMVRYAADKWGHDRVAQVITFGTIKTKAALKDSARIHYGQPGFA  475
MTB_DnaE1_Rv1547              407  LFERFLNPERTSMPDIDIDFDDRRRGEMVRYAADKWGHDRVAQVITFGTIKTKAALKDSARIHYGQPGFA  476
Mbovis_Mb1574                 407  LFERFLNPERTSMPDIDIDFDDRRRGEMVRYAADKWGHDRVAQVITFGTIKTKAALKDSARIHYGQPGFA  476
BCG_Pasteur_BCG1600           407  LFERFLNPERTSMPDIDIDFDDRRRGEMVRYAADKWGHDRVAQVITFGTIKTKAALKDSARIHYGQPGFA  476
Mcanettii_DnaE1               407  LFERFLNPERTSMPDIDIDFDDRRRGEMVRYAADKWGHDRVAQVITFGTIKTKAALKDSARIHYGQPGFA  476
CDC1551_MT1598                407  LFERFLNPERTSMPDIDIDFDDRRRGEMVRYAADKWGHDRVAQVITFGTIKTKAALKDSARIHYGQPGFA  476
```


```
Conservation:
Mavium_MaviaA2_DnaE2                505  QQDAWSKQISSWSGL--------------------ADSPDVEGIPPQVIDLANQVRNLPRHLGIHSGG  552
Maviumparatub_MAP3476c              505  QQDAWSKQISSWSGP--------------------ADSPDVEGIPPQVIDLANQVRNLPRHLGIHSGG  552
Mavium_104_MAV4335                  505  QQDAWSKQISSWSGL--------------------ADSPDVEGIPPQVIDLANQVRNLPRHLGIHSGG  552
CDC1551_MT3480                      512  QQDAWSKQVSHWTGQ-----------------------ADDVDGIPEQVIDLATQIRNLPRHLGIHSGG  557
MTB_DnaE2_Rv3370c                   493  QQDAWSKQVSHWTGQ-----------------------ADDVDGIPEQVIDLATQIRNLPRHLGIHSGG  538
Mbovis_Mb3405c                      505  QQDAWSKQVSHWTGQ-----------------------ADDVDGIPEQVIDLATQIRNLPRHLGIHSGG  550
BCG_Pasteur_BCG3442c                493  QQDAWSKQVSHWTGQ-----------------------ADDVDGIPEQVIDLATQIRSLPRHLGIHSGG  538
Mcanettii_DnaE2                     496  QQDAWSKQIGHWTGR-----------------------ADDTDGIPEQVLDLATQIRNLPRHLGIHSGG  541
MintA_DnaE2                         507  QQDAWSKQISHWGGL---------------------ADSPEVEDIPEPVIDLANQIRNLPRHMGIHSGG  554
Mkansasii_MkanA1_DnaE2              510  QQDAWSKQISHWNGQ-----------------------AAEVDGIPEQVVDLATQIRNLPRHMGIHSGG  555
Mmarinum_MMAR1158                   514  QQDAWSKQIGHWNAT-----------------------PDDVEGIPEQVIDLAAQIRNLPRHMGIHSGG  559
Mulcerans_MUL0923                   502  QQDAWSKQIGHWNAT-----------------------LDDVEGIPEQVIDLAAQIRNLPRHMGIHSGG  547
Myco_sp_Mmcs_1183                   511  QQDAWSKQLSKWNGL---------------------ADSPDIEGIPEPVVDLALQISNLPRHMGIHSGG  558
Myco_sp_Mjls_1210                   511  QQDAWSKQLSKWNGL---------------------ADSPDIEGIPEPVVDLALQISNLPRHMGIHSGG  558
Mvanbaalenii_Mvan1530               510  QQDAWSKQVSQWGNL---------------------ADATHVEDIPGPVVDLAKQISNLPRHMGIHSGG  557
Mgilvum_Mflv4893                    499  QQDAWSKQISQWGNL---------------------ADATHVEDIPEPVIDLAMQISHLPRHMGIHSGG  546
MSMEG_DnaE2_MSMEG_1633              495  QQDAWSKHLSRWDGR---------------------PDSPDVAEIPEQVIELANQIANLPRHMGIHSGG  542
Mabscessus_MAB3703c                 489  QQDAWSKQMGHWGGL---------------------ADAAAVDGIPPQVIQLARQIKDFPRHMGIHSGG  536
Maviumparatub_MAP1257               473  IADRITKALPPPIMAKDIPLSGITDPSHERYKEAAEVRGLIETDPDVRTIYQTARGLEGLIRNAGVHACA  542
Mavium_104_MAV3224                  473  IADRITKALPPPIMAKDIPLSGITDPSHERYKEAAEVRGLIETDPDVRTIYQTARGLEGLIRNAGVHACA  542
MintA_DnaE1                         473  IADRITKALPPPIMAKDIPLSGITDPAHERYKEAAEVRGLIETDPDVRTIYQTARGLEGLIRNAGVHACA  542
Mabscessus_MAB2696c                 478  IADRITKALPPPIMAKDIPVWGITNPEHERYKEAAEVRTLIDTDPDVRNIYQTALGLEGLIRNAGVHACA  547
Myco_sp_Mmcs_3087                   483  IADRITKALPPPIMAKDIPVSGITDPTHERYKEAAEVRALIDTDPDVRTIYETARGLEGLVRNAGVHACA  552
Myco_sp_Mjls_3107                   483  IADRITKALPPPIMAKDIPVSGITDPTHERYKEAAEVRALIDTDPDVRTIYETARGLEGLVRNAGVHACA  552
Mgilvum_Mflv3636                    476  IADRITKALPPPIMAKDIPVSGITDPTHERYKEAAEVRALIDTDPDVRTIYETARGLEGLVRNAGVHACA  545
Mvanbaalenii_Mvan2777               474  IADRITKALPPPIMAKDIPVSGITDPTHERYKEAAEVRALIDTDPDVRTIYETARGLEGLVRNAGVHACA  543
MSM_MSMEG3178                       480  IADRITKALPPPIMAKDIPLSGITDPNHERYKEAAEVRGLIDTDPDVRTIYETARGLEGLVRNAGVHACA  549
Mmarinum_MMAR2369                   476  IADRITKALPPPIMAKDIPLSGITDPAHERFKEAAEVRGLIETDPDVRTIYQTARGLEGLIRNAGVHACA  545
Mulcerans_MUL1545                   476  IADRITKALPPPIMAKDIPLSGITDPAHERFKEAAEVRGLIETDPDVRTIYQTARGLEGLIRNAGVHACA  545
ML_ML1207                           473  MADRITKALPPAIMAKDIPLSGITDPAHERFKEAAEVRSLIETDSDVRIIYQTARGLEGLIRNAGVHACA  542
Mkansasii_MkanA1_DnaE1              476  IADRITKALPPAIMAKDIPLSGITDPAHERYKEAAEVRGLIETDPDVRTIYQTALGLEGLIRNAGVHACA  545
MTB_DnaE1_Rv1547                    477  IADRITKALPPAIMAKDIPLSGITDPSHERYKEAAEVRGLIETDPDVRTIYQTARGLEGLIRNAGVHACA  546
Mbovis_Mb1574                       477  IADRITKALPPAIMAKDIPLSGITDPSHERYKEAAEVRGLIETDPDVRTIYQTARGLEGLIRNAGVHACA  546
BCG_Pasteur_BCG1600                 477  IADRITKALPPAIMAKDIPLSGITDPSHERYKEAAEVRGLIETDPDVRTIYQTARGLEGLIRNAGVHACA  546
Mcanettii_DnaE1                     477  IADRITKALPPAIMAKDIPLSGITDPSHERYKEAAEVRGLIETDPDVRTIYQTARGLEGLIRNAGVHACA  546
CDC1551_MT1598                      477  IADRITKALPPAIMAKDIPLSGITDPSHERYKEAAEVRGLIETDPDVRTIYQTARGLEGLIRNAGVHACA  546
```


Palm
Fingers
Conservation:
Mavium_MaviaA2_DnaE2 553 MVICDRPIADVCPVEWARMENRSVLQWDKDDCAAIGLVKFDLLGLGMLSALHYAIDLVAEHKGIEVDLSR 622
Maviumparatub_MAP3476c 553 MVICDRPIADVCPVEWARMENRSVLQWDKDDCAAIGLVKFDLLGLGMLSALHYAIDLVAEHKGIEVDLAR 622
Mavium_104_MAV4335 553 MVICDRPIADVCPVEWARMENRSVLQWDKDDCAAIGLVKFDLLGLGMLSALHYAIDLVAEHKGIEVDLAR 622
CDC1551_MT3480 558 MVICDRPIADVCPVEWARMANRSVLQWDKDDCAAIGLVKFDLLGLGMLSALHYAKDLVAEHKGIEVDLAR 627
MTB_DnaE2_Rv3370c 539 MVICDRPIADVCPVEWARMANRSVLQWDKDDCAAIGLVKFDLLGLGMLSALHYAKDLVAEHKGIEVDLAR 608
Mbovis_Mb3405c 551 MVICDRPIADVCPVEWARMANRSVLQWDKDDCAAIGLVKFDLLGLGMLSALHYAKDLVAEHKGIEVDLAR 620
BCG_Pasteur_BCG3442c 539 MVICDRPIADVCPVEWARMANRSVLQWDKDDCAAIGLVKFDLLGLGMLSALHYAKDLVAEHKGIEVDLAR 608
Mcanettii_DnaE2 542 MVICDRPIADVCPVEWARMANRSVLQWDKDDCAAIGLVKFDLLGLGMLSALHYAKDLVAEHKGIEVDLAR 611
MintA_DnaE2 555 MVICDRPIADVCPVEWARMENRSVLQWDKDDCAAIGLVKFDLLGLGMLSALHYAIDLVAEHKGIEVDLAR 624
Mkansasii_MkanA1_DnaE2 556 MVICDRPIADVCPVEWARMENRSVLQWDKDDCAAIGLVKFDLLGLGMLSALHYAMDLVAEHKGLEVDLAK 625
Mmarinum_MMAR1158 560 MVICDRPIADVCPVEWARMADRSVLQWDKDDCAAIGLVKFDLLGLGMLSALHYARDLVAEHKGIEVDLAR 629
Mulcerans_MUL0923 548 MVICDRPIADVCPVEWARMADRSVLQWDKDDCAAIGLVKFDLLGLGMLSALHYARDLVAEHKGIEVDLAR 617
Myco_sp_Mmcs_1183 559 MVICDRPIADVCPVEWARMANRSVLQWDKDDCAAIGLVKFDMLGLGMLSALHYCIDLVREHKGLDVDLAK 628
Myco_sp_Mjls_1210 559 MVICDRPIADVCPVEWARMANRSVLQWDKDDCAAIGLVKFDMLGLGMLSALHYCIDLVREHKGLDVDLAK 628
Mvanbaalenii_Mvan1530 558 MVICDRPIADVCPVEWARMENRSVLQWDKDDCAAIGLVKFDLLGLGMLSALHYAIDLVAEHKGIEVDLAK 627
Mgilvum_Mflv4893 547 MVICDRPIADVCPVEWARMENRSVLQWDKDDCAAIGLVKFDLLGLGMLSALHYAIDLVAEHKGLEVDLAK 616
MSMEG_DnaE2_MSMEG_1633 543 MVICDRPIADVCPVEWARMENRSVLQWDKDDCAAIGLVKFDMLGLGMLSALHYAIDLVAEHKGIEVDLAT 612
Mabscessus_MAB3703c 537 MVICDRPLADVVPVEWARMENRSVLQWDKDDCAAVGLVKFDLLGLGMLSALHYCIDLVREHKGIDVDLAH 606
Maviumparatub_MAP1257 543 VIMSSEPLTEAIPLWKRPQDGAIITGWDYPSCEAIGLLKMDFLGLRNLTIIGDALDNIKANRGIDLDLES 612
Mavium_104_MAV3224 543 VIMSSEPLTEAIPLWKRPQDGAIITGWDYPSCEAIGLLKMDFLGLRNLTIIGDALDNIKANRGIDLDLES 612
MintA_DnaE1 543 VIMSSEPLTEAIPLWKRPQDGAIITGWDYPSCEAIGLLKMDFLGLRNLTIIGDALDNIKANRGIDLDLES 612
Mabscessus_MAB2696c 548 VIMSSEPLTDAIPLWKRAQDGAIITGWDYPSCEAIGLLKMDFLGLRNLTVIGDALDNIKANRGIDLDLDH 617
Myco_sp_Mmcs_3087 553 VIMSSEPLIDAIPLWRRPQDGAVITGWDYPSCEAIGLLKMDFLGLRNLTIIGDCIENIKANRGVDLDLES 622
Myco_sp_Mjls_3107 553 VIMSSEPLIDAIPLWRRPQDGAVITGWDYPSCEAIGLLKMDFLGLRNLTIIGDCIENIKANRGVDLDLES 622
Mgilvum_Mflv3636 546 VIMSSEPLIDAIPLWKRPQDGAVITGWDYPSCEAIGLLKMDFLGLRNLTIIGDCIENIKANRGVDVDLES 615
Mvanbaalenii_Mvan2777 544 VIMSSEPLIDAIPLWKRPQDGAVITGWDYPSCEAIGLLKMDFLGLRNLTIIGDCIENIKANRGIDVDLES 613
MSM_MSMEG3178 550 VIMSSEPLIDAIPLWKRPQDGAIITGWDYPSCEAIGLLKMDFLGLRNLTIIGDAIANIKANRGIDLDLET 619
Mmarinum_MMAR2369 546 VIMSSEPLTEAIPLWKRPQDGAIITGWDYPSCEAIGLLKMDFLGLRNLTIIGDAIENIKANRGIDLDLES 615
Mulcerans_MUL1545 546 VIMSSEPLTEAIPLWKRPQDGAIITGWDYPSCEAIGLLKMDFLGLRNLTIIGDAIENIEANRGIDLDLES 615
ML_ML1207 543 VILSSEPLTEAIPLWKRPQDGAIITGWDYPACEAIGLLKMDFLGLRNLTIIGDAIENIKTNRGIDLDLES 612
Mkansasii_MkanA1_DnaE1 546 VIMSSEPLTEAIPLWKRPQDGAIITGWDYPSCEAIGLLKMDFLGLRNLTIIGDAIDNIKANKGIDLDLES 615
MTB_DnaE1_Rv1547 547 VIMSSEPLTEAIPLWKRPQDGAIITGWDYPACEAIGLLKMDFLGLRNLTIIGDAIDNVRANRGIDLDLES 616
Mbovis_Mb1574 547 VIMSSEPLTEAIPLWKRPQDGAIITGWDYPACEAIGLLKMDFLGLRNLTIIGDAIDNVRANRGIDLDLES 616
BCG_Pasteur_BCG1600 547 VIMSSEPLTEAIPLWKRPQDGAIITGWDYPACEAIGLLKMDFLGLRNLTIIGDAIDNVRANRGIDLDLES 616
Mcanettii_DnaE1 547 VIMSSEPLTEAIPLWKRPQDGAIITGWDYPACEAIGLLKMDFLGLRNLTIIGDAIDNVRANRGIDLDLES 616
CDC1551_MT1598 547 VIMSSEPLTEAIPLWKRPQDGAIITGWDYPACEAIGLLKMDFLGLRNLTIIGDAIDNVRANRGIDLDLES 616

		Fingers	
Conservation:			
Mavium_MaviaA2_DnaE2	623	LDLSEPAVYEMLARADSVGVFQVESRAQMATLPRLKPRVFYDLVVEVALIRPGPIQGGSVHPYIRRRNGV	692
Maviumparatub_MAP3476c	623	LDLSEPAVYEMLARADSVGVFQVESRAQMATLPRLKPRVFYDLVVEVALIRPGPIQGGSVHPYIRRRNGV	692
Mavium_104_MAV4335	623	LDLSEPAVYEMLARADSVGVFQVESRAQMATLPRLKPRVFYDLVVEVALIRPGPIQGGSVHPYIRRRNGV	692
CDC1551_MT3480	628	LDLSEPAVYEMLARADSVGVFQVESRAQMATLPRLKPRVFYDLVVEVALIRPGPIQGGSVHPYIRRRNGV	697
MTB_DnaE2_Rv3370c	609	LDLSEPAVYEMLARADSVGVFQVESRAQMATLPRLKPRVFYDLVVEVALIRPGPIQGGSVHPYIRRRNGV	678
Mbovis_Mb3405c	621	LDLSEPAVYEMLARADSVGVFQVESRAQMATLPRLKPRVFYDLVVEVALIRPGPIQGGSVHPYIRRRNGV	690
BCG_Pasteur_BCG3442c	609	LDLSEPAVYEMLARADSVGVFQVESRAQMATLPRLKPRVFYDLVVEVALIRPGPIQGGSVHPYIRRRNGV	678
Mcanettii_DnaE2	612	LDLSEPAVYEMLARADSVGVFQVESRAQMATLPRLRPRVFYDLVVEVALIRPGPIQGGSVHPYIRRRNGV	681
MintA_DnaE2	625	LDLSEPAVYEMLARADSVGVFQVESRAQMATLPRLKPRVFYDLVVEVALIRPGPIQGGSVHPYIRRRNGE	694
Mkansasii_MkanA1_DnaE2	626	LDLSEPAVYEMLARADSVGVFQVESRAQMATLPRLRPRVFYDLVVEVALIRPGPIQGGSVHPYIRRRNGI	695
Mmarinum_MMAR1158	630	LDLSEPAVYEMLARADSVGVFQVESRAQMATLPRLKPRVFYDLVVEVALIRPGPIQGGSVHPYIRRRNGI	699
Mulcerans_MUL0923	618	LDLSEPAVYEMLARADSVGVFQVESRAQMATLPRLKPRVFYDLVVEVALIRPGPIQGGSVHPYIRRRSGI	687
Myco_sp_Mmcs_1183	629	LDLSEPAVYEMLQRADSVGVFQVESRAQMATLPRLKPRVFYDLVVEVALIRPGPIQGGSVHPYIKRRNGE	698
Myco_sp_Mjls_1210	629	LDLSEPAVYEMLQRADSVGVFQVESRAQMATLPRLKPRVFYDLVVEVALIRPGPIQGGSVHPYIKRRNGE	698
Mvanbaalenii_Mvan1530	628	LDLSEPAVYEMLARADSVGVFQVESRAQMATLPRLKPRMFYDLVVEVALIRPGPIQGGSVHPYIKRRNGQ	697
Mgilvum_Mflv4893	617	LDLSEPAVYEMLQRADSVGVFQVESRAQMATLPRLKPRMFYDLVVEVALIRPGPIQGGSVHPYIKRRNGQ	686
MSMEG_DnaE2_MSMEG_1633	613	LDLSEPAVYEMLQRADSVGVFQVESRAQMATLPRLKPREFYDLVVEVALIRPGPIQGGSVHPYIKRRNGQ	682
Mabscessus_MAB3703c	607	INLKEQAVYEMLARADSVGVFQVESRAQMATLPRLKPKCFYDLVVEVALIRPGPIQGGSVHPYIRRYNKI	676
Maviumparatub_MAP1257	613	VPLDDKATYELLGRGDTLGVFQLDGGPMRDLLRRMQPTEFNDIVAVLALYRPGPMGMNAHNDYADRKNGR	682
Mavium_104_MAV3224	613	VPLDDKATYELLGRGDTLGVFQLDGGPMRDLLRRMQPTEFNDIVAVLALYRPGPMGMNAHNDYADRKNGR	682
MintA_DnaE1	613	VPLDDGPTYELLGRGDTLGVFQLDGGPMRDLLRRMQPTEFNDIVAVLALYRPGPMGMNAHNDYADRKNGR	682
Mabscessus_MAB2696c	618	LPLDDPATYELLSRGDTLGVFQLDGGPMRDLLRRMQPTGFNDIVAVLALYRPGPMGMNAHNDYADRKNGR	687
Myco_sp_Mmcs_3087	623	LALDDPKAYELLGRGDTLGVFQLDGGPMRDLLRRMQPTEFNDIVAVLALYRPGPMGMNAHNDYADRKNGR	692
Myco_sp_Mjls_3107	623	LALDDPKAYELLGRGDTLGVFQLDGGPMRDLLRRMQPTEFNDIVAVLALYRPGPMGMNAHNDYADRKNGR	692
Mgilvum_Mflv3636	616	LALDDPKAYELLGRGDTLGVFQLDGSAMRDLLRRMQPTEFNDIVAVLALYRPGPMGMNAHNDYADRKNGR	685
Mvanbaalenii_Mvan2777	614	LTLDDPKAYELLGRGDTLGVFQLDGGPMRDLLRRMQPTEFNDIVAVLALYRPGPMGMNAHNDYADRKNGR	683
MSM_MSMEG3178	620	LPLDDPAAYELLSRGDTLGVFQLDGGPMRDLLRRMQPTGFNDIVAVLALYRPGPMGMNAHNDYADRKNGR	689
Mmarinum_MMAR2369	616	LPLDDKPTYELLGRGETLGVFQLDGGPMRDLLRRMQPTEFNDIVAVLALYRPGPMGMNAHNDYADRKNNR	685
Mulcerans_MUL1545	616	LPLDDKPTYELLGRGETLGVFQLDGGPMRDLLRRMQPTEFNDIVAVLALYRPGPMGMNAHNDYADRKNNR	685
ML_ML1207	613	VPLDDQATYELLGRGDTLGVFQLDGGPMRDLLRRMQPTGFEDIVAVLALYRPGPMGMNAHNDYADRKNNR	682
Mkansasii_MkanA1_DnaE1	616	VPLDDKPTYELLGRGDTLGVFQLDGGPMRDLLRRMQPTGFEDIVAVLALYRPGPMGQNAHNDYADRKNGR	685
MTB_DnaE1_Rv1547	617	VPLDDKATYELLGRGDTLGVFQLDGGPMRDLLRRMQPTGFEDVVAVIALYRPGPMGMNAHNDYADRKNNR	686
Mbovis_Mb1574	617	VPLDDKATYELLGRGDTLGVFQLDGGPMRDLLRRMQPTGFEDVVAVIALYRPGPMGMNAHNDYADRKNNR	686
BCG_Pasteur_BCG1600	617	VPLDDKATYELLGRGDTLGVFQLDGGPMRDLLRRMQPTGFEDVVAVIALYRPGPMGMNAHNDYADRKNNR	686
Mcanettii_DnaE1	617	VPLDDKATYELLGRGDTLGVFQLDGGPMRDLLRRMQPTGFEDVVAVIALYRPGPMGMNAHNDYADRKNNR	686
CDC1551_MT1598	617	VPLDDKATYELLGRGDTLGVFQLDGGPMRDLLRRMQPTGFEDVVAVIALYRPGPMGMNAHNDYADRKNNR	686

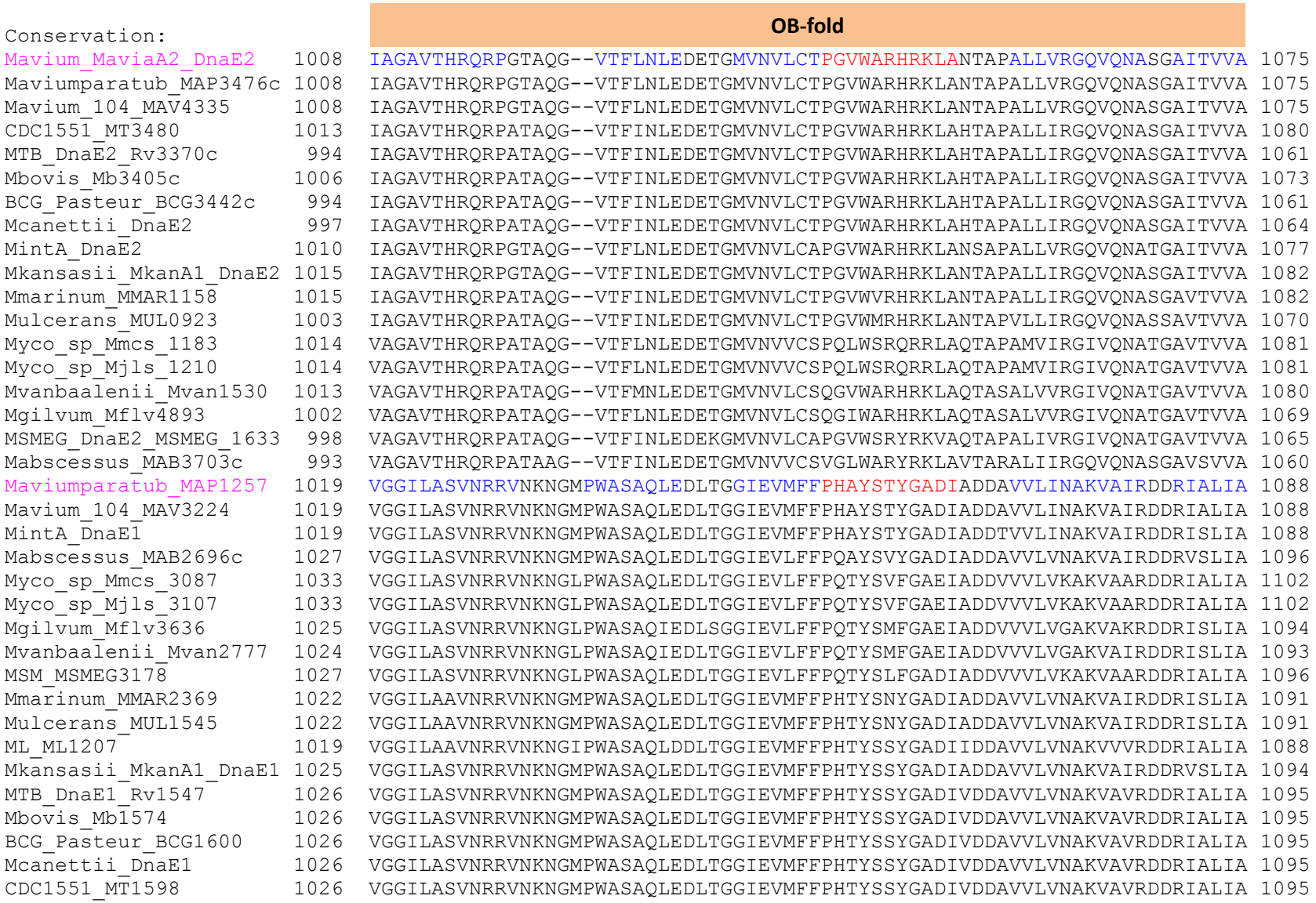
		Fingers	
Conservation:			
Mavium_MaviaA2_DnaE2	693	DP-VLYDH----PSMEPALRKTLGVPLFQEQLMQLAVDCAGFSAAEADQLRRAMGSKRSTERMRRLRGRF	757
Maviumparatub_MAP3476c	693	DP-VLYDH----PSMEPALRKTLGVPLFQEQLMQLAVDCAGFSAAEADQLRRAMGSKRSTERMRRLRSRF	757
Mavium_104_MAV4335	693	DP-VLYDH----PSMEPALRKTLGVPLFQEQLMQLAVDCAGFSAAEADQLRRAMGSKRSTERMRRLRGRF	757
CDC1551_MT3480	698	DP-VIYEH----PSMAPALRKTLGVPLFQEQLMQLAVDCAGFSAAEADQLRRAMGSKRSTERMRRLRGRF	762
MTB_DnaE2_Rv3370c	679	DP-VIYEH----PSMAPALRKTLGVPLFQEQLMQLAVDCAGFSAAEADQLRRAMGSKRSTERMRRLRGRF	743
Mbovis_Mb3405c	691	DP-VIYEH----PSMAPALRKTLGVPLFQEQLMQLAVDCAGFSAAEADQLRRAMGSKRSTERMRRLRGRF	755
BCG_Pasteur_BCG3442c	679	DP-VIYEH----PSMAPALRKTLGVPLFQEQLMQLAVDCAGFSAAEADQLRRAMGSKRSTERMRRLRGRF	743
Mcanettii_DnaE2	682	DP-VIYEH----PSMAPALRKTLGVPLFQEQLMQLAVDCAGFSAAEADQLRRAMGSKRSTERMRRLRGRF	746
MintA_DnaE2	695	DP-VVYDH----PSMEPALRKTLGVPLFQEQLMQLAVDCAGFSAAEADQLRRAMGSKRSTERMRRLRGRF	759
Mkansasii_MkanA1_DnaE2	696	DP-VVYEH----PSMAPALRKTLGVPLFQEQLMQLAVDCAGFTAAEADQLRRAMGSKRSTERMQRLRGRF	760
Mmarinum_MMAR1158	700	DP-VVYEH----PSMESALRKTLGVPLFQEQLMQLAVDCAGFSAAEADQLRRAIGSKRSTERMRRLRGRF	764
Mulcerans_MUL0923	688	DP-VVYEH----PSMESALRKTLGVPLFQEQLMQLAVDCAGFSAAEADQLRRAIGSKRSTERMRRLRGRF	752
Myco_sp_Mmcs_1183	699	EP-VTYDH----PSMANALRKTLGVPLFQEQLMQLAVDCAGFTAAEADQLRRAMGSKRSTEKMRRLRGRF	763
Myco_sp_Mjls_1210	699	EP-VTYDH----PSMANALRKTLGVPLFQEQLMQLAVDCAGFTAAEADQLRRAMGSKRSTEKMRRLRGRF	763
Mvanbaalenii_Mvan1530	698	EA-VTYDH----PSMESALRKTLGVPLFQEQLMQLAVDCAGFTPAEADQLRRAMGSKRSTEKMRRLRGRF	762
Mgilvum_Mflv4893	687	EP-VTYEH----PSMERALRKTLGVPLFQEQLMQLAVDCAGFSAAEADQLRRAMGSKRSTEKMRRLRGRF	751
MSMEG_DnaE2_MSMEG_1633	683	EP-VTYDH----PSMEPALKKTLGVPLFQEQLMQLAVDCAGFSAAEADQLRRAMGSKRSTAKMRRLRSRF	747
Mabscessus_MAB3703c	677	DKDWQHDH----PSMAAALDKTLGVPLFQEQLMQLAVDVAGFSPAESDQLRRAMGSKRSPERMERLRNRF	742
Maviumparatub_MAP1257	683	QP-IKPIHPELEEPLREILAETYGLIVYQEQIMFIAQKVASYTMGKADALRKAMGKKK-LEVLEAEYKGF	750
Mavium_104_MAV3224	683	QP-IKPIHPELEEPLREILAETYGLIVYQEQIMFIAQKVASYTMGKADALRKAMGKKK-LEVLEAEYKGF	750
MintA_DnaE1	683	QP-IKPIHPELEEPLREILSETYGLIVYQEQIMFIAQKVASYTMGKADALRKAMGKKK-LEVLEAEYKGF	750
Mabscessus_MAB2696c	688	QP-IKPIHPELEEPLKDILAETFGLIVYQEQIMFIAQKVASYSMGKADALRKAMGKKK-LEVLEAEYKGF	755
Myco_sp_Mmcs_3087	693	QP-IKPIHPELEEPLKEILAETYGLIVYQEQIMFIAQKVASYTMGKADALRKAMGKKK-LEVLEAEYQGF	760
Myco_sp_Mjls_3107	693	QP-IKPIHPELEEPLKEILAETYGLIVYQEQIMFIAQKVASYTMGKADALRKAMGKKK-LEVLEAEYQGF	760
Mgilvum_Mflv3636	686	QP-IKPIHPELEEPLKDILAETYGLIVYQEQIMFIAQKVASYSMGKADALRKAMGKKK-LEVLEAEYKGF	753
Mvanbaalenii_Mvan2777	684	QA-IKPIHPELEEPLKDILSETYGLIVYQEQIMFIAQKVASYSMGKADALRKAMGKKK-LEVLEAEYQGF	751
MSM_MSMEG3178	690	QP-IKPIHPELEEPLKDILAETYGLIVYQEQIMFIAQKVASYTMGKADALRKAMGKKK-LEVLEAEYKGF	757
Mmarinum_MMAR2369	686	QA-VKPIHPELEEPLREILSETYGLIVYQEQIMFIAQKVASYTMGKADALRKAMGKKK-LEVLEAEYKGF	753
Mulcerans_MUL1545	686	QA-VKPIHPELEEPLREILSETYGLIVYQEQIMFIAQKVASYTMGKADALRKAMGKKK-LEVLEAEYKGF	753
ML_ML1207	683	QV-IKPIHPELEEPLREILAETYGLIVYQEQIMRIAQKVAGYSLARADILRKAMGKKK-REVLEKEFEGF	750
Mkansasii_MkanA1_DnaE1	686	QA-IKPIHPELAEPLEEILAETYGLIVYQEQIMRIAQKVAGYSLARADILRKAMGKKK-REVLDKEYEGF	753
MTB_DnaE1_Rv1547	687	QA-IKPIHPELEEPLREILAETYGLIVYQEQIMRIAQKVASYSLARADILRKAMGKKK-REVLEKEFEGF	754
Mbovis_Mb1574	687	QA-IKPIHPELEEPLREILAETYGLIVYQEQIMRIAQKVASYSLARADILRKAMGKKK-REVLEKEFEGF	754
BCG_Pasteur_BCG1600	687	QA-IKPIHPELEEPLREILAETYGLIVYQEQIMRIAQKVASYSLARADILRKAMGKKK-REVLEKEFEGF	754
Mcanettii_DnaE1	687	QA-IKPIHPELEEPLREILAETYGLIVYQEQIMRIAQKVASYSLARADILRKAMGKKK-REVLEKEFEGF	754
CDC1551_MT1598	687	QA-IKPIHPELEEPLREILAETYGLIVYQEQIMRIAQKVASYSLARADILRKAMGKKK-REVLEKEFEGF	754

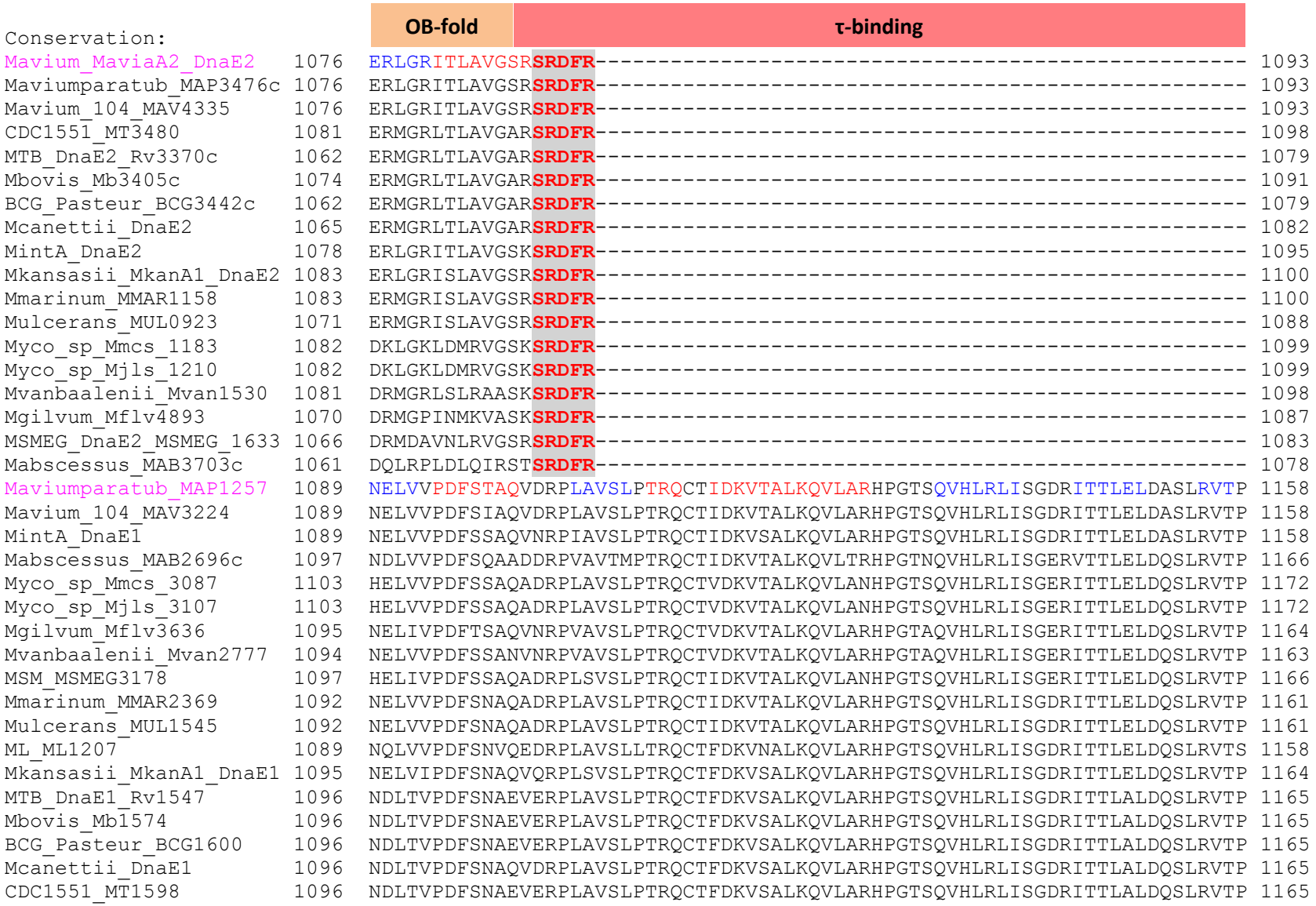

```
Conservation:
Mavium_MaviaA2_DnaE2          758  YDGMRALHGAPDEVIDRTYEKLEAFANFGFPESHALSFASLVFYSSWFKLHHPAAFCAALLRAQPMGFYS  827
Maviumparatub_MAP3476c        758  YDGMRALHGAPDEVIDRTYEKLEAFANFGFPESHALSFASLVFYSSWFKLHHPAAFCAALLRAQPMGFYS  827
Mavium_104_MAV4335            758  YDGMRALHGAPDEVIDRTYEKLEAFANFGFPESHALSFASLVFYSSWFKLHHPAAFCAALLRAQPMGFYS  827
CDC1551_MT3480                763  YDGMRALHGAPDEVIDRIYEKLEAFANFGFPESHALSFASLVFYSAWFKLHHPAAFCAALLRAQPMGFYS  832
MTB_DnaE2_Rv3370c             744  YDGMRALHGAPDEVIDRIYEKLEAFANFGFPESHALSFASLVFYSAWFKLHHPAAFCAALLRAQPMGFYS  813
Mbovis_Mb3405c                756  YDGMRALHGAPDEVIDRIYEKLEAFANFGFPESHALSFASLVFYSAWFKLHHPAAFCAALLRAQPMGFYS  825
BCG_Pasteur_BCG3442c          744  YDGMRALHGAPDEVIDRIYEKLEAFANFGFPESHALSFASLVFYSAWFKLHHPAAFCAALLRAQPMGFYS  813
Mcanettii_DnaE2               747  YDGMRALHGAPDEVIDRIYEKLEAFANFGFPESHALSFASLVFYSAWFKLHHPAAFCAALLRAQPMGFYS  816
MintA_DnaE2                   760  YEGMRALHGAPDEVIDRTYEKLEAFANFGFPESHALSFASLVFYSSWFKLHHPAAFCAALLRAQPMGFYS  829
Mkansasii_MkanA1_DnaE2        761  YDGMRALHGTRDEVIDRIYEKLEAFANFGFPESHALSFASLVFYSAWFKLHHPAAFCAALLRAQPMGFYS  830
Mmarinum_MMAR1158             765  YDGMRALHGAPDEVIDRIYEKLEAFANFGFPESHALSFASLVFYSSWFKLHHPAAFCAALLRAQPMGFYS  834
Mulcerans_MUL0923             753  YDGMRALHGAPDEVIDRIYEKLEAFANFGFPESHALSFASLVFYSSWFKLHHPAAFCAALLRAQPMGFYS  822
Myco_sp_Mmcs_1183             764  YDGMRQRHGITGEVADRIYDKLEAFANFGFPESHSLSFASLVFYSSWFKLHHPAAFCAALLRAQPMGFYS  833
Myco_sp_Mjls_1210             764  YDGMRQRHGITGEVADRIYDKLEAFANFGFPESHSLSFASLVFYSSWFKLHHPAAFCAALLRAQPMGFYS  833
Mvanbaalenii_Mvan1530         763  FDGMAELHGVTGDVAQRIYEKLEAFANFGFPESHSLSFASLVFYSSWFKLHHPAAFCAALLRAQPMGFYS  832
Mgilvum_Mflv4893              752  FEGMAELHGISGDVARRIYEKLEAFANFGFPESHSLSFASLVFYSSWFKLHHPAAFCAALLRAQPMGFYS  821
MSMEG_DnaE2_MSMEG_1633        748  YDGMRERHGITGVVADRIYEKLEAFANFGFPESHSLSFASLVFYSSWFKLHHPAAFCAALLRAQPMGFYS  817
Mabscessus_MAB3703c           743  YEGMRNLHGITGELADRIYEKLYAFANFGFPESHAQSFASLVFYSSWFKLHHPAAFCAALLRAQPMGFYS  812
Maviumparatub_MAP1257         751  YEGMT-ANGFSEKAVKALWDTILPFAGYAFNKSHAAGYGLVSYWTAYLKANYPAEYMAGLLTSVGDDKDK  819
Mavium_104_MAV3224            751  YEGMT-ANGFSEKAVKALWDTILPFAGYAFNKSHAAGYGLVSYWTAYLKANYPAEYMAGLLTSVGDDKDK  819
MintA_DnaE1                   751  YEGMT-ANGFSEKAVKALWDTILPFAGYAFNKSHAAGYGLVSYWTAYLKANYPAEYMAGLLTSVGDDKDK  819
Mabscessus_MAB2696c           756  YEGMT-NNGFSEKAVKALWDTILPFAGYAFNKSHAAGYGLVSFWTAYLKANYPAEYMAGLLTSVGDDKDK  824
Myco_sp_Mmcs_3087             761  REGMT-ANGFSEAAVKALWDTILPFAGYAFNKSHAAGYGLVSYWTAYLKANYPAEYMAGLLTSVGDDKDK  829
Myco_sp_Mjls_3107             761  REGMT-ANGFSEAAVKALWDTILPFAGYAFNKSHAAGYGLVSYWTAYLKANYPAEYMAGLLTSVGDDKDK  829
Mgilvum_Mflv3636              754  KEGMT-ANGFSEGAVKALWDTILPFAGYAFNKSHAAGYGLVSYWTAYLKANYQAEYMAGLLTSVGDDKDK  822
Mvanbaalenii_Mvan2777         752  KEGMT-ANGFSEKAVKALWDTILPFAGYAFNKSHAAGYGLVSYWTAYLKANYPAEYMAGLLTSVGDDKDK  820
MSM_MSMEG3178                 758  KEGMT-ANGFSEGAVKALWDTILPFAGYAFNKSHAAGYGLVSYWTAYLKANYPAEYMAGLLTSVGDDKDK  826
Mmarinum_MMAR2369             754  YEGMT-ANGFSEKAVKALWDTILPFAGYAFNKSHAAGYGLVSYWTAYLKANYPAEYMAGLLTSVGDDKDK  822
Mulcerans_MUL1545             754  YEGMT-ANGFSEKAVKALWDTILPFAGYAFNKSHAAGYGLVSYWTAYLKANYPAEYMAGLLTSVGDDKDK  822
ML_ML1207                     751  SEGMQ-ANGFSVHAIKALWDIILPFADYAFNKSHAAGYGLISYWTAYLKANFAGEYMAGLLTSVGDDKDK  819
Mkansasii_MkanA1_DnaE1        754  AQGMQ-ANGFSSGAIKALWDTILPFADYAFNKSHAAGYGLVSYWTAYLKANYPAEYMAGLLTSVGDDKDK  822
MTB_DnaE1_Rv1547              755  SDGMQ-ANGFSPAAIKALWDTILPFADYAFNKSHAAGYGMVSYWTAYLKANYPAEYMAGLLTSVGDDKDK  823
Mbovis_Mb1574                 755  SDGMQ-ANGFSPAAIKALWDTILPFADYAFNKSHAAGYGMVSYWTAYLKANYPAEYMAGLLTSVGDDKDK  823
BCG_Pasteur_BCG1600           755  SDGMQ-ANGFSPAAIKALWDTILPFADYAFNKSHAAGYGMVSYWTAYLKANYPAEYMAGLLTSVGDDKDK  823
Mcanettii_DnaE1               755  SDGMQ-ANGFSPAAIKALWDTILPFADYAFNKSHAAGYGMVSYWTAYLKANYPAEYMAGLLTSVGDDKDK  823
CDC1551_MT1598                755  SDGMQ-ANGFSPAAIKALWDTILPFADYAFNKSHAAGYGMVSYWTAYLKANYPAEYMAGLLTSVGDDKDK  823
```

		β-binding	
Conservation:			
Mavium_MaviaA2_DnaE2	828	PQSLVADARRHGVTVHGPDVNASLAHATLENAGTEVRLGLGAVRHIGDDLAEKLVQERKANGPFASLLDL	897
Maviumparatub_MAP3476c	828	PQSLVADARRHGVTVHGPDVNASLAHATLENAGTEVRLGLGAVRHIGDDLAEKLVQERKANGPFTSLLDL	897
Mavium_104_MAV4335	828	PQSLVADARRHGVTVHGPDVNASLAHATLENAGTEVRLGLGAVRHIGDDLAEKLVQERKANGPFASLLDL	897
CDC1551_MT3480	833	PQSLVADARRHGVAVHGPCVNASLAHATCENAGTEVRLGLGAVRYLGAELAEKLVAERTANGPFTSLPDL	902
MTB_DnaE2_Rv3370c	814	PQSLVADARRHGVAVHGPCVNASLAHATCENAGTEVRLGLGAVRYLGAELAEKLVAERTANGPFTSLPDL	883
Mbovis_Mb3405c	826	PQSLVADARRHGVAVHGPCVNASLAHATCENAGTEVRLGLGAVRYLGAELAEKLVAERTANGPFTSLPDL	895
BCG_Pasteur_BCG3442c	814	PQSLVADARRHGVAVHGPCVNASLAHATCENAGTEVRLGLGAVRYLGAELAEKLVAERTANGPFTSLPDL	883
Mcanettii_DnaE2	817	PQSLVADARRHGVVVHGPCVNASLAHATCENAGTEVRLGLGAVRYLGAKLAEKLVAERTANGPFTSLPDL	886
MintA_DnaE2	830	PQSLVADARRHGVTVHGPDVNASLAHATLENTGTEVRLGLGAVRHIGDDLAEQLVEERKANGPFASLLDV	899
Mkansasii_MkanA1_DnaE2	831	PQSLVADARRHGVVVHGPDVNASLAHATLENAGMEVRLGLGAVRHIGDDLAEKLVGERNANGPFASLLDL	900
Mmarinum_MMAR1158	835	PQSLVADARRHGVLVHGPDVNASLAHATLENAGMQVRLGLGAVRHIGDDLAESLVAERNDNGPFASLLNL	904
Mulcerans_MUL0923	823	PQSLVADARRHGVLVHGPDVNASLAHATLENAGMQVRLGLGAVRHIGDDLAESLVAERNDNGPFASLLNL	892
Myco_sp_Mmcs_1183	834	PQSLVADARRHGVTVHGPDVNASLAYATLESAGTEVRLGLGAVRHIGDDLAERIVEERKANGPFESLLDL	903
Myco_sp_Mjls_1210	834	PQSLVADARRHGVTVHGPDVNASLAYATLESAGTEVRLGLGAVRHIGDDLAERIVEERKANGPFESLLDL	903
Mvanbaalenii_Mvan1530	833	PQSLVADARRHGVVVHGPDVNAGLAHATLENHGLDVRLGLGGVRHIGDELAERLVGERKAHGPFTSLTDL	902
Mgilvum_Mflv4893	822	PQTLVADARRHGVDVHGPDVNASLAHATLENHGLDVRLGLGSIRHIGDELAQRLVEDRKLNGPFVSLTDL	891
MSMEG_DnaE2_MSMEG_1633	818	PQSLVADARRHGVTVHGPDVNASLAHAGLENRGLDVRLGLGSVRHIGDDLAQRIVDEREANGQFTSLLDL	887
Mabscessus_MAB3703c	813	PQSLVADARRHGVTVHGPDVNASLSYATLENAGLEVRIGLGAVRHIGDDLAQAIVEERKVRGPFVSLLDL	882
Maviumparatub_MAP1257	820	AAVYLADCRKLGITVLPPDVNESLVNFASVG--QDIRFGLGAVRNVGANVVGSLIKTRNEKGKFTDFSDY	887
Mavium_104_MAV3224	820	AAVYLADCRKLGITVLPPDVNESLVNFASVG--QDIRFGLGAVRNVGANVVGSLIKTREEKGKFTDFSDY	887
MintA_DnaE1	820	AAVYLADCRKLGITVLPPDVNESLVNFASVG--TDIRFGLGAVRNVGANVVGSLISTRNGKGKFTDFSDY	887
Mabscessus_MAB2696c	825	AAVYLADCRRLGITVLPPDVNESVHNFASVG--EDIRYGLGGVRNVGANVVQSLIATREEKGAFTDFSDY	892
Myco_sp_Mmcs_3087	830	AAVYLADCRRLGITVLPPDVNESVQNFASVG--DDIRFGLGAVRNVGANVVASLVNTRAEKGKYSDFSDY	897
Myco_sp_Mjls_3107	830	AAVYLADCRRLGITVLPPDVNESVQNFASVG--DDIRFGLGAVRNVGANVVASLVNTRAEKGKYSDFSDY	897
Mgilvum_Mflv3636	823	AAIYLADCRRLGITVLPPDVNESVQNFASVG--DDIRFGLGAVRNVGANVVASLVSTREEKGKYTDFSDY	890
Mvanbaalenii_Mvan2777	821	AAVYLADCRRLGITVLPPDVNESVQNFASVG--DDIRFGLGAVRNVGANVVSSLVATRKEKGKFADFSDY	888
MSM_MSMEG3178	827	AAVYLADCRRLGITVLPPDVNESEHNFASVG--DDIRFGLGAIRNVGANVVSSLIKTRTEKGKFTDFSDY	894
Mmarinum_MMAR2369	823	AAVYLADCRKLGITVLPPDVNESLVNFASVG--QDIRFGLGAVRNVGANVVGSLIGTRSGKGKFTDFSDY	890
Mulcerans_MUL1545	823	AAVYLADCRKLGITVLPPDVNESLVNFASVG--QDIRFGLGAVRNVGANVVGSLIGTRSGKGKFTDFSDY	890
ML_ML1207	820	AAVYLADCRKFGITVLPPDVNESVLDFASVG--ADIRYGLGAVRNVGANVVGSLIKTRNAKGKFADFSDY	887
Mkansasii_MkanA1_DnaE1	823	AAVYLADCRKLGITVLPPDVNESGLNFASVG--EDIRYGLGAVRNVGANVVGSLIQTRSDKGKFTDFSDY	890
MTB_DnaE1_Rv1547	824	AAVYLADCRKLGITVLPPDVNESGLNFASVG--QDIRYGLGAVRNVGANVVGSLLQTRNDKGKFTDFSDY	891
Mbovis_Mb1574	824	AAVYLADCRKLGITVLPPDVNESGLNFASVG--QDIRYGLGAVRNVGANVVGSLLQTRNDKGKFTDFSDY	891
BCG_Pasteur_BCG1600	824	AAVYLADCRKLGITVLPPDVNESGLNFASVG--QDIRYGLGAVRNVGANVVGSLLQTRNDKGKFTDFSDY	891
Mcanettii_DnaE1	824	AAVYLADCRKLGITVLPPDVNESGLNFASVG--QDIRYGLGAVRNVGANVVGSLLQTRNDKGKFTDFSDY	891
CDC1551_MT1598	824	AAVYLADCRKLGITVLPPDVNESGLNFASVG--QDIRYGLGAVRNVGANVVGSLLQTRNDKGKFTDFSDY	891

		β-binding	
Conservation:			
Mavium_MaviaA2_DnaE2	898	TARLH---LSVPQTEALATAGAFGCFGMSRREALWAAGAAATQRPDRLPG----------	944
Maviumparatub_MAP3476c	898	TARLQ---LSVQQTEALATAGAFGCFGMSRREALWAAGAAATQRPDRLPG----------	944
Mavium_104_MAV4335	898	TARLQ---LSVPQTEALATAGAFGCFGMSRREALWAAGAAATQRPDRLPG----------	944
CDC1551_MT3480	903	TSRVQ---LSVPQVEALATAGALGCFGMSRREALWAAGAAATGRPDRLPG----------	949
MTB_DnaE2_Rv3370c	884	TSRVQ---LSVPQVEALATAGALGCFGMSRREALWAAGAAATGRPDRLPG----------	930
Mbovis_Mb3405c	896	TSRVQ---LSVPQVEALATAGALGCFGMSRREALWAAGAAATGRPDRLPG----------	942
BCG_Pasteur_BCG3442c	884	TSRVQ---LSVPQVEALATAGALGCFGMSRREALWAAGAAATGRPDRLPG----------	930
Mcanettii_DnaE2	887	TSRVQ---LSVPQVEALATAGALGCFGMSRREALWAAGAAATGRPDRLPG----------	933
MintA_DnaE2	900	TSRLQ---LSVPQTEALATAGAFACFGMSRREGLWAAGAAATQRPGRLPG----------	946
Mkansasii_MkanA1_DnaE2	901	ASRVQ---LSVPQTEALATAGALGCFGMSRREALWAAGAAATQRPDRLPG----------	947
Mmarinum_MMAR1158	905	TSRVQ---LSVPQTEALATAGALGCFGMSRREALWAAGAAATQRPDRLPG----------	951
Mulcerans_MUL0923	893	TSRVQ---LSVPQTEALATTGALGCFGMSRREALWAAGAAATQRPDRLPG----------	939
Myco_sp_Mmcs_1183	904	TGRVQ---LSVPQTEALATAGALGCFGVTRREGLWAAGAAAAERPDRLPG----------	950
Myco_sp_Mjls_1210	904	TGRVQ---LSVPQTEALATAGALGCFGVTRREGLWAAGAAAAERPDRLPG----------	950
Mvanbaalenii_Mvan1530	903	TRRVQ---LSVPQTEALATAGALGCFGITRREGLWAAGAAATERPDRLPG----------	949
Mgilvum_Mflv4893	892	TRRVQ---LTVPQTEALATAGALGCFGITRREGLWAAGAAATERPDRLPG----------	938
MSMEG_DnaE2_MSMEG_1633	888	TSRVQ---LTVPQTEALATAGALGCFGITRREGLWAAGAAATQRPDRLPG----------	934
Mabscessus_MAB3703c	883	TGRIQ---LTTAQVEALATGGALGCFDMSRREALWVAGAAAAQRPDRLPG----------	929
Maviumparatub_MAP1257	888	LNKIDISACNKKVTESLIKAGAFDSLKHARKGLFLVHTDAVDSVLGTKKAEAMGQFDLFG-GDGGC-T--	953
Mavium_104_MAV3224	888	LNKIDISACNKKVTESLIKAGAFDSLKHARKGLFLVHTDAVDSVLGTKKAEAMGQFDLFG-GDGGC-T--	953
MintA_DnaE1	888	LNKIDIAACNKKVTESLIKAGAFDSLGHARKGLFLVHTDAVDSVLGTKKAEAIGQFDLFG-GDDGC-T--	953
Mabscessus_MAB2696c	893	LHKIDIAACNKKVTESLVKAGAFDSLGHPRKGLFLVQSDAVDSVLGTKKAEAMGQFDLFGGGGEDS-VPA	961
Myco_sp_Mmcs_3087	898	LNKIDIAACNKKVTESLIKAGAFDSLGHPRKGLFLIHTDAVDSVLGTKKAEAMGQFDLFGSGDGSG-ADA	966
Myco_sp_Mjls_3107	898	LNKIDIAACNKKVTESLIKAGAFDSLGHPRKGLFLIHTDAVDSVLGTKKAEAMGQFDLFGSGDGSG-ADA	966
Mgilvum_Mflv3636	891	LNKIDITACTKKVTESLIKAGAFDSLGHPRKGLFLVHTDAVDSVLGTKKAEAMGQFDLFGGGDDDG-D-T	958
Mvanbaalenii_Mvan2777	889	LNKIDIAACNKKVTESLVKAGAFDSLGHPRKGLFLVHTDAVDSVLGTKKAEAMGQFDLFGGADGSGAD-A	957
MSM_MSMEG3178	895	LNKIDITACNKKVTESLIKAGAFDSLGHPRKGLFLVHTDAVDSVLGTKKAEAMGQFDLFGGGEDTG-T--	961
Mmarinum_MMAR2369	891	LNKIDISACNKKVTESLIKAGAFDSLGHPRKGLFLVHTDAVDSVLGTKKAEAMGQFDLFG-GSDSG-A--	956
Mulcerans_MUL1545	891	LNKIDISACNKKVTESLIKAGAFDSLGHPRKGLFLVHTDAVDSVLGTKKAEAMGQFDLFG-GSDSG-A--	956
ML_ML1207	888	LNKIDITSCNKKVTESLIKAGAFDSLGHSRKGLFLVHADAVDSVLGTKKAEAIGQFDLFG-GTDGG-T--	953
Mkansasii_MkanA1_DnaE1	891	LNKIDISACNKKVTESLIKAGAFDSLGHARKGLFLVHTDAVDSVLGTKKAEAMGQFDLFG-GSDDG-TGA	958
MTB_DnaE1_Rv1547	892	LNKIDISACNKKVTESLIKAGAFDSLGHARKGLFLVHSDAVDSVLGTKKAEALGQFDLFG-SNDDG-TGT	959
Mbovis_Mb1574	892	LNKIDISACNKKVTESLIKAGAFDSLGHARKGLFLVHSDAVDSVLGTKKAEALGQFDLFG-SNDDG-TGT	959
BCG_Pasteur_BCG1600	892	LNKIDISACNKKVTESLIKAGAFDSLGHARKGLFLVHSDAVDSVLGTKKAEALGQFDLFG-SNDDG-TGT	959
Mcanettii_DnaE1	892	LNKIDISACNKKVTESLIKAGAFDSLGHARKGLFLVHSDAVDSVLGTKKAEALGQFDLFG-SNDDG-TGT	959
CDC1551_MT1598	892	LNKIDISACNKKVTESLIKAGAFDSLGHARKGLFLVHSDAVDSVLGTKKAEALGQFDLFG-SNDDG-TGT	959


```
Conservation:
Mavium_MaviaA2_DnaE2                945  -VGS----SSHIPALPGMSELELAAADVWATGISPDSYPTQFLRDDLDAMGVVPAARLGSVP--DGDRVL 1007
Maviumparatub_MAP3476c              945  -VGS----SSHIPALPGMSELELAAADVWATGISPDSYPTQFLRDDLDAMGVVPAARLGSVP--DGDRVL 1007
Mavium_104_MAV4335                  945  -VGS----SSHIPALPGMSELELAAADVWATGISPDSYPTQFLRDDLDAMGVVPAARLGSVP--DGDRVL 1007
CDC1551_MT3480                      950  -VGS----SSHIPALPGMSELELAAADVWATGVSPDSYPTQFLRADLDAMGVLPAERLGSVS--DGDRVL 1012
MTB_DnaE2_Rv3370c                   931  -VGS----SSHIPALPGMSELELAAADVWATGVSPDSYPTQFLRADLDAMGVLPAERLGSVS--DGDRVL  993
Mbovis_Mb3405c                      943  -VGS----SSHIPALPGMSELELAAADVWATGVSPDSYPTQFLRADLDAMGVLPAERLGSVS--DGDRVL 1005
BCG_Pasteur_BCG3442c                931  -VGS----SSHIPALPGMSELELAAADVWATGVSPDSYPTQFLRADLDAMGVLPAERLGSVS--DGDRVL  993
Mcanettii_DnaE2                     934  -VGS----SSHIPALPGMSELELAAADVWATGVSPDSYPTQFLRADLDAMGVLPAERLGSVP--DGDRVL  996
MintA_DnaE2                         947  -VGS----SSHVPALPGMSELELAAADVWATGISPDSYPTQFLREDLDAMGVVPAEKLGSVP--DGDRVL 1009
Mkansasii_MkanA1_DnaE2              948  -VGAQGADGSHIPVLPGMSELELAAADVWATGISPDSYPTQFLRADLDAMGVLPADALGSVP--DGDRVL 1014
Mmarinum_MMAR1158                   952  -VGS----SSHIPTLPGMSELELAASDVWATGISPDSYPTQFLRADLDALGVLAAAALLSVP--DGERVL 1014
Mulcerans_MUL0923                   940  -VGS----SSHIPTLPGMSELELAASDVWATGISPDSYPTQFLRADLDALGVLAAAALLSVP--DGERVL 1002
Myco_sp_Mmcs_1183                   951  -VGS----AGQVPSLPGMSKLELAAADVWATGVSPDSYPTQFLRENLDAMGVVPAGRLLDVP--DGTRVL 1013
Myco_sp_Mjls_1210                   951  -VGS----AGQVPSLPGMSKLELAAADVWATGVSPDSYPTQFLRENLDAMGVIPAGRLLDVP--DGTRVL 1013
Mvanbaalenii_Mvan1530               950  -VGS----SSHVPSLPGMTELELTVADVWATGVSPDRYPTEFLREDLDAMGVVPADQLLSLP--DGTRVL 1012
Mgilvum_Mflv4893                    939  -VGS----SSQVPSLPGMTELELTAADVWATGVSPDRYPTEFLREDLDAMGVVPADRLLSVR--DGTRVL 1001
MSMEG_DnaE2_MSMEG_1633              935  -VGS----STHIPPLPGMSALELSAADVWATGISPDSYPTEYLRKHLDSLGVVPTDRLLDVA--DGTRIL  997
Mabscessus_MAB3703c                 930  -VGV----SSRIPALPEMGEVTRAAADVWATGVSPDSYPTQFLRARLDAMGVVPAKGLFGVP--DGSRVL  992
Maviumparatub_MAP1257               954  -ESV----FTIKVPDDEWEDKHKLALEREMLGLYVSGHPLNGVAHLLAAQVDTQIPAILDGDVPNETQVR 1018
Mavium_104_MAV3224                  954  -ESV----FTIKVPDDEWEDKHKLALEREMLGLYVSGHPLNGVAHLLAAQVDTQIPAILDGDVPNETQVR 1018
MintA_DnaE1                         954  -EAV----FTIKVPDDEWEDKHKLALEREMLGLYVSGHPLNGVAHLLAAQVDTQIPAILDGDVPNETQVR 1018
Mabscessus_MAB2696c                 962  SDI-----FAIKVPDDEWEEKHKLALEREMLGLYVSGHPLNGVAHLLSRQTDTQIPAILDGEVANDAVVK 1026
Myco_sp_Mmcs_3087                   967  GDSA----FSIKVPDEEWEDKHKLALEREMLGLYVSGHPLNGVAHLLANQVDTQIPAILDGDVANDAQVL 1032
Myco_sp_Mjls_3107                   967  GDSA----FSIKVPDEEWEDKHKLALEREMLGLYVSGHPLNGVAHLLANQVDTQIPAILDGDVANDAQVL 1032
Mgilvum_Mflv3636                    959  GDSA----FTIKVPDEEWEDKHKLALEREMLGLYVSGHPLNGIAHLLTAQVDTQIPAILEGAIPNDAQVR 1024
Mvanbaalenii_Mvan2777               958  GDSA----FTIKVPDEEWEDKHKLALEREMLGLYVSGHPLNGVAHLLAAQVDTQIPAILDGDVANDTQVR 1023
MSM_MSMEG3178                       962  -DAV----FTIKVPDEEWEDKHKLALEREMLGLYVSGHPLNGVAHLLANQVDTQIPAILDGDVANDAQVL 1026
Mmarinum_MMAR2369                   957  -DAV----FTIKVPEDEWEDKHKLALEREMLGLYVSGHPLNGVAHLLAAQVDTAIPAILDGDVSNDAQVR 1021
Mulcerans_MUL1545                   957  -DAV----FTIKVPEDEWEDKHKLALEREMLGLYVSGHPLNGVAHLLATQVDTAIPAILDGDVSNDAQVR 1021
ML_ML1207                           954  -DAV----FTIKVPDDEWEDKHKLALEREMLGLYVSGHPLNGVAHLLAAQVDTAIPTILDGGVSNDTQVR 1018
Mkansasii_MkanA1_DnaE1              959  ADPV----FTIKVPDDEWEDKHKLALEREMLGLYVSGHPLNGVAHLLAGQVDTAIPAILDGDVPNDTQVR 1024
MTB_DnaE1_Rv1547                    960  ADPV----FTIKVPDDEWEDKHKLALEREMLGLYVSGHPLNGVAHLLAAQVDTAIPAILDGDVPNDAQVR 1025
Mbovis_Mb1574                       960  ADPV----FTIKVPDDEWEDKHKLALEREMLGLYVSGHPLNGVAHLLAAQVDTAIPAILDGDVPNDAQVR 1025
BCG_Pasteur_BCG1600                 960  ADPV----FTIKVPDDEWEDKHKLALEREMLGLYVSGHPLNGVAHLLAAQVDTAIPAILDGDVPNDAQVR 1025
Mcanettii_DnaE1                     960  ADPV----FTIKVPDDEWEDKHKLALEREMLGLYVSGHPLNGVAHLLAAQVDTAIPAILDGDVPNDAQVR 1025
CDC1551_MT1598                      960  ADPV----FTIKVPDDEWEDKHKLALEREMLGLYVSGHPLNGVAHLLAAQVDTAIPAILDGDVPNDAQVR 1025
```


OB-fold
Conservation:
Mavium_MaviaA2_DnaE2 1008 IAGAVTHRQRPGTAQG--VTFLNLEDETGMVNVLCTPGVWARHRKLANTAPALLVRGQVQNASGAITVVA 1075
Maviumparatub_MAP3476c 1008 IAGAVTHRQRPGTAQG--VTFLNLEDETGMVNVLCTPGVWARHRKLANTAPALLVRGQVQNASGAITVVA 1075
Mavium_104_MAV4335 1008 IAGAVTHRQRPGTAQG--VTFLNLEDETGMVNVLCTPGVWARHRKLANTAPALLVRGQVQNASGAITVVA 1075
CDC1551_MT3480 1013 IAGAVTHRQRPATAQG--VTFINLEDETGMVNVLCTPGVWARHRKLAHTAPALLIRGQVQNASGAITVVA 1080
MTB_DnaE2_Rv3370c 994 IAGAVTHRQRPATAQG--VTFINLEDETGMVNVLCTPGVWARHRKLAHTAPALLIRGQVQNASGAITVVA 1061
Mbovis_Mb3405c 1006 IAGAVTHRQRPATAQG--VTFINLEDETGMVNVLCTPGVWARHRKLAHTAPALLIRGQVQNASGAITVVA 1073
BCG_Pasteur_BCG3442c 994 IAGAVTHRQRPATAQG--VTFINLEDETGMVNVLCTPGVWARHRKLAHTAPALLIRGQVQNASGAITVVA 1061
Mcanettii_DnaE2 997 IAGAVTHRQRPATAQG--VTFINLEDETGMVNVLCTPGVWARHRKLAHTAPALLIRGQVQNASGAITVVA 1064
MintA_DnaE2 1010 IAGAVTHRQRPGTAQG--VTFLNLEDETGMVNVLCAPGVWARHRKLANSAPALLVRGQVQNATGAITVVA 1077
Mkansasii_MkanA1_DnaE2 1015 IAGAVTHRQRPGTAQG--VTFINLEDETGMVNVLCTPGVWARHRKLANTAPALLIRGQVQNASGAITVVA 1082
Mmarinum_MMAR1158 1015 IAGAVTHRQRPATAQG--VTFINLEDETGMVNVLCTPGVWVRHRKLANTAPALLIRGQVQNASGAVTVVA 1082
Mulcerans_MUL0923 1003 IAGAVTHRQRPATAQG--VTFINLEDETGMVNVLCTPGVWMRHRKLANTAPVLLIRGQVQNASSAVTVVA 1070
Myco_sp_Mmcs_1183 1014 VAGAVTHRQRPATAQG--VTFLNLEDETGMVNVVCSPQLWSRQRRLAQTAPAMVIRGIVQNATGAVTVVA 1081
Myco_sp_Mjls_1210 1014 VAGAVTHRQRPATAQG--VTFLNLEDETGMVNVVCSPQLWSRQRRLAQTAPAMVIRGIVQNATGAVTVVA 1081
Mvanbaalenii_Mvan1530 1013 VAGAVTHRQRPATAQG--VTFMNLEDETGMVNVLCSQGVWARHRKLAQTASALVVRGIVQNATGAVTVVA 1080
Mgilvum_Mflv4893 1002 VAGAVTHRQRPATAQG--VTFLNLEDETGMVNVLCSQGIWARHRKLAQTASALVVRGIVQNATGAVTVVA 1069
MSMEG_DnaE2_MSMEG_1633 998 VAGAVTHRQRPATAQG--VTFINLEDEKGMVNVLCAPGVWSRYRKVAQTAPALIVRGIVQNATGAVTVVA 1065
Mabscessus_MAB3703c 993 VAGAVTHRQRPATAAG--VTFINLEDETGMVNVVCSVGLWARYRKLAVTARALIIRGQVQNASGAVSVVA 1060
Maviumparatub_MAP1257 1019 VGGILASVNRRVNKNGMPWASAQLEDLTGGIEVMFFPHAYSTYGADIADDAVVLINAKVAIRDDRIALIA 1088
Mavium_104_MAV3224 1019 VGGILASVNRRVNKNGMPWASAQLEDLTGGIEVMFFPHAYSTYGADIADDAVVLINAKVAIRDDRIALIA 1088
MintA_DnaE1 1019 VGGILASVNRRVNKNGMPWASAQLEDLTGGIEVMFFPHAYSTYGADIADDTVVLINAKVAIRDDRISLIA 1088
Mabscessus_MAB2696c 1027 VGGILASVNRRVNKNGMPWASAQLEDLTGGIEVLFFPQAYSVYGADIADDAVVLVNAKVAIRDDRVSLIA 1096
Myco_sp_Mmcs_3087 1033 VGGILASVNRRVNKNGLPWASAQLEDLTGGIEVLFFPQTYSVFGAEIADDVVVLVKAKVAARDDRIALIA 1102
Myco_sp_Mjls_3107 1033 VGGILASVNRRVNKNGLPWASAQLEDLTGGIEVLFFPQTYSVFGAEIADDVVVLVKAKVAARDDRIALIA 1102
Mgilvum_Mflv3636 1025 VGGILASVNRRVNKNGLPWASAQIEDLSGGIEVLFFPQTYSMFGAEIADDVVVLVGAKVAKRDDRISLIA 1094
Mvanbaalenii_Mvan2777 1024 VGGILASVNRRVNKNGLPWASAQIEDLTGGIEVLFFPQTYSMFGAEIADDVVVLVGAKVAIRDDRISLIA 1093
MSM_MSMEG3178 1027 VGGILASVNRRVNKNGLPWASAQLEDLTGGIEVLFFPQTYSLFGADIADDVVVLVKAKVAARDDRIALIA 1096
Mmarinum_MMAR2369 1022 VGGILAAVNRRVNKNGMPWASAQLEDLTGGIEVMFFPHTYSNYGADIADDAVVLVNAKVAIRDDRISLIA 1091
Mulcerans_MUL1545 1022 VGGILAAVNRRVNKNGMPWASAQLEDLTGGIEVMFFPHTYSNYGADIADDAVVLVNAKVAIRDDRISLIA 1091
ML_ML1207 1019 VGGILAAVNRRVNKNGIPWASAQLDDLTGGIEVMFFPHTYSSYGADIIDDAVVLVNAKVVVRDDRIALIA 1088
Mkansasii_MkanA1_DnaE1 1025 VGGILASVNRRVNKNGMPWASAQLEDLTGGIEVMFFPHTYSSYGADIADDAVVLVNAKVAIRDDRVSLIA 1094
MTB_DnaE1_Rv1547 1026 VGGILASVNRRVNKNGMPWASAQLEDLTGGIEVMFFPHTYSSYGADIVDDAVVLVNAKVAVRDDRIALIA 1095
Mbovis_Mb1574 1026 VGGILASVNRRVNKNGMPWASAQLEDLTGGIEVMFFPHTYSSYGADIVDDAVVLVNAKVAVRDDRIALIA 1095
BCG_Pasteur_BCG1600 1026 VGGILASVNRRVNKNGMPWASAQLEDLTGGIEVMFFPHTYSSYGADIVDDAVVLVNAKVAVRDDRIALIA 1095
Mcanettii_DnaE1 1026 VGGILASVNRRVNKNGMPWASAQLEDLTGGIEVMFFPHTYSSYGADIVDDAVVLVNAKVAVRDDRIALIA 1095
CDC1551_MT1598 1026 VGGILASVNRRVNKNGMPWASAQLEDLTGGIEVMFFPHTYSSYGADIVDDAVVLVNAKVAVRDDRIALIA 1095


OB-fold
τ-binding
Conservation:
Mavium_MaviaA2_DnaE2 1076 ERLGRITLAVGSRSRDFR---------- 1093
Maviumparatub_MAP3476c 1076 ERLGRITLAVGSRSRDFR---------- 1093
Mavium_104_MAV4335 1076 ERLGRITLAVGSRSRDFR---------- 1093
CDC1551_MT3480 1081 ERMGRLTLAVGARSRDFR---------- 1098
MTB_DnaE2_Rv3370c 1062 ERMGRLTLAVGARSRDFR---------- 1079
Mbovis_Mb3405c 1074 ERMGRLTLAVGARSRDFR---------- 1091
BCG_Pasteur_BCG3442c 1062 ERMGRLTLAVGARSRDFR---------- 1079
Mcanettii_DnaE2 1065 ERMGRLTLAVGARSRDFR---------- 1082
MintA_DnaE2 1078 ERLGRITLAVGSKSRDFR---------- 1095
Mkansasii_MkanA1_DnaE2 1083 ERLGRISLAVGSRSRDFR---------- 1100
Mmarinum_MMAR1158 1083 ERMGRISLAVGSRSRDFR---------- 1100
Mulcerans_MUL0923 1071 ERMGRISLAVGSRSRDFR---------- 1088
Myco_sp_Mmcs_1183 1082 DKLGKLDMRVGSKSRDFR---------- 1099
Myco_sp_Mjls_1210 1082 DKLGKLDMRVGSKSRDFR---------- 1099
Mvanbaalenii_Mvan1530 1081 DRMGRLSLRAASKSRDFR---------- 1098
Mgilvum_Mflv4893 1070 DRMGPINMKVASKSRDFR---------- 1087
MSMEG_DnaE2_MSMEG_1633 1066 DRMDAVNLRVGSRSRDFR---------- 1083
Mabscessus_MAB3703c 1061 DQLRPLDLQIRSTSRDFR---------- 1078
Maviumparatub_MAP1257 1089 NELVVPDFSTAQVDRPLAVSLPTRQCTIDKVTALKQVLARHPGTSQVHLRLISGDRITTLELDASLRVTP 1158
Mavium_104_MAV3224 1089 NELVVPDFSIAQVDRPLAVSLPTRQCTIDKVTALKQVLARHPGTSQVHLRLISGDRITTLELDASLRVTP 1158
MintA_DnaE1 1089 NELVVPDFSSAQVNRPIAVSLPTRQCTIDKVSALKQVLARHPGTSQVHLRLISGDRITTLELDASLRVTP 1158
Mabscessus_MAB2696c 1097 NDLVVPDFSQAADDRPVAVTMPTRQCTIDKVTALKQVLTRHPGTNQVHLRLISGERVTTLELDQSLRVTP 1166
Myco_sp_Mmcs_3087 1103 HELVVPDFSSAQADRPLAVSLPTRQCTVDKVTALKQVLANHPGTSQVHLRLISGERITTLELDQSLRVTP 1172
Myco_sp_Mjls_3107 1103 HELVVPDFSSAQADRPLAVSLPTRQCTVDKVTALKQVLANHPGTSQVHLRLISGERITTLELDQSLRVTP 1172
Mgilvum_Mflv3636 1095 NELIVPDFTSAQVNRPVAVSLPTRQCTVDKVTALKQVLARHPGTAQVHLRLISGERITTLELDQSLRVTP 1164
Mvanbaalenii_Mvan2777 1094 NELVVPDFSSANVNRPVAVSLPTRQCTVDKVTALKQVLARHPGTAQVHLRLISGERITTLELDQSLRVTP 1163
MSM_MSMEG3178 1097 HELIVPDFSSAQADRPLSVSLPTRQCTIDKVTALKQVLANHPGTSQVHLRLISGERITTLELDQSLRVTP 1166
Mmarinum_MMAR2369 1092 NELVVPDFSNAQADRPLAVSLPTRQCTIDKVTALKQVLARHPGTSQVHLRLISGDRITTLELDQSLRVTP 1161
Mulcerans_MUL1545 1092 NELVVPDFSNAQADRPLAVSLPTRQCTIDKVTALKQVLARHPGTSQVHLRLISGDRITTLELDQSLRVTP 1161
ML_ML1207 1089 NQLVVPDFSNVQEDRPLAVSLLTRQCTFDKVNALKQVLARHPGTSQVHLRLISGDRITTLELDQSLRVTS 1158
Mkansasii_MkanA1_DnaE1 1095 NELVIPDFSNAQVQRPLSVSLPTRQCTFDKVSALKQVLARHPGTSQVHLRLISGDRITTLELDQSLRVTP 1164
MTB_DnaE1_Rv1547 1096 NDLTVPDFSNAEVERPLAVSLPTRQCTFDKVSALKQVLARHPGTSQVHLRLISGDRITTLALDQSLRVTP 1165
Mbovis_Mb1574 1096 NDLTVPDFSNAEVERPLAVSLPTRQCTFDKVSALKQVLARHPGTSQVHLRLISGDRITTLALDQSLRVTP 1165
BCG_Pasteur_BCG1600 1096 NDLTVPDFSNAEVERPLAVSLPTRQCTFDKVSALKQVLARHPGTSQVHLRLISGDRITTLALDQSLRVTP 1165
Mcanettii_DnaE1 1096 NDLTVPDFSNAQVDRPLAVSLPTRQCTFDKVSALKQVLARHPGTSQVHLRLISGDRITTLALDQSLRVTP 1165
CDC1551_MT1598 1096 NDLTVPDFSNAEVERPLAVSLPTRQCTFDKVSALKQVLARHPGTSQVHLRLISGDRITTLALDQSLRVTP 1165

τ-binding

```
Conservation:
Mavium_MaviaA2_DnaE2          ------------------
Maviumparatub_MAP3476c        ------------------
Mavium_104_MAV4335            ------------------
CDC1551_MT3480                ------------------
MTB_DnaE2_Rv3370c             ------------------
Mbovis_Mb3405c                ------------------
BCG_Pasteur_BCG3442c          ------------------
Mcanettii_DnaE2               ------------------
MintA_DnaE2                   ------------------
Mkansasii_MkanA1_DnaE2        ------------------
Mmarinum_MMAR1158             ------------------
Mulcerans_MUL0923             ------------------
Myco_sp_Mmcs_1183             ------------------
Myco_sp_Mjls_1210             ------------------
Mvanbaalenii_Mvan1530         ------------------
Mgilvum_Mflv4893              ------------------
MSMEG_DnaE2_MSMEG_1633        ------------------
Mabscessus_MAB3703c           ------------------
Maviumparatub_MAP1257  1159   SPALMGDLKELLGPGCLGG 1177
Mavium_104_MAV3224     1159   SPALMGDLKELLGPGCLGG 1177
MintA_DnaE1            1159   SPALMGDLKELLGPGCLGG 1177
Mabscessus_MAB2696c    1167   SSALMGDLKALLGPGCLGG 1185
Myco_sp_Mmcs_3087      1173   SSALMGDLKALLGPGCLGG 1191
Myco_sp_Mjls_3107      1173   SSALMGDLKALLGPGCLGG 1191
Mgilvum_Mflv3636       1165   SSALMGDLKELLGPGCLGG 1183
Mvanbaalenii_Mvan2777  1164   SSALMGDLKELLGPGCLGG 1182
MSM_MSMEG3178          1167   SSALMGDLKALLGPGCLG- 1184
Mmarinum_MMAR2369      1162   SPALMGDLKELLGPGCLGS 1180
Mulcerans_MUL1545      1162   SPALMGDLKELLGPGCLGS 1180
ML_ML1207              1159   SPALMGDLKALLGPGCLGD 1177
Mkansasii_MkanA1_DnaE1 1165   SPALMGDLKELLGPGCLGG 1183
MTB_DnaE1_Rv1547       1166   SPALMGDLKELLGPGCLGS 1184
Mbovis_Mb1574          1166   SPALMGDLKELLGPGCLGS 1184
BCG_Pasteur_BCG1600    1166   SPALMGDLKELLGPGCLGS 1184
Mcanettii_DnaE1        1166   SPALMGDLKELLGPGCLGS 1184
CDC1551_MT1598         1166   SPALMGDLKELLGPGCLGS 1184
```

Figure S1: Multiple sequence alignment of representative mycobacterial DnaE subunits. DnaE-type sequences from *M. avium* 104 [DnaE1 (MAV_3224) and DnaE2 (MAV_4335)], *M. avium paratuberculosis* [DnaE1 (MAP1257) and DnaE2 (MAP3476c)], Mtb CDC1551 [DnaE1 (MT1598) and DnaE2 (MT3480)], Mtb [DnaE1 (Rv1547) and DnaE2 (Rv3370c)], *M. bovis* [DnaE1 (Mb1574) and DnaE2 (Mb3405c)], *M. bovis* BCG_Pasteur [DnaE1 (BCG1600) and DnaE2 (BCG3442c)], *M. canettii* [DnaE1 (WP_015303087) and DnaE2 (WP_015303851)], *M. intracellulare* [DnaE1 (WP_014385147) and DnaE2 (AFC45412)], *M. kansasii* [DnaE1 (KEP42088) and DnaE2 (AIR29751)], *M. marinum* [DnaE1 (MMAR2369) and DnaE2 (MMAR1158)], *M. ulcerans* [DnaE1 (MUL1545) and DnaE2 (MUL0923*)], Mycobacterium sp*. MCS DnaE1 (Mmcs_3087) and DnaE2 (Mmcs_1183)], *M. vanbaalenii* [DnaE1 (Mvan2777) and DnaE2 (Mvan1530)], *M. gilvum* [DnaE1 (Mflv3636) and DnaE2 (Mflv4893)], Msm [DnaE1 (MSMEG_3718) and DnaE2 (MSMEG_1633)], *M. abscessus* [DnaE1 (MAB2696c) and DnaE2 (MAB3703c)], *Mycobacterium sp*. JLS [DnaE1 (Mjls_3107) and DnaE2 (Mjls_1210)] and *M. leprae* (ML1207) were aligned using the PROMALS3D multiple sequence and structure alignment server (http://prodata.swmed.edu/promals3d/promals3d.php). (1) Bold red residues shaded in green are the conserved metal co-ordinating residues [H H D/H H E H C/H D/N H] (Barros *et al.*, 2013), (2) bold red residues shaded in grey are C-terminal amino acid motif -[S/T/G]R[D/N]F[D/R/H]- conserved in DnaE2-type proteins (Timinskas *et al.*, 2014, Warner *et al.*, 2010), (3) bold blue residues inside red boxes are conserved in essential DnaE-type polymerases and absent in non-essential DnaE-type polymerases. The annotation of the polymerase domains is derived from crystal structures of *E. coli* (Lamers *et al.*, 2006) and *T. aquaticus* (Bailey *et al.*, 2006) PolIIIα subunits.

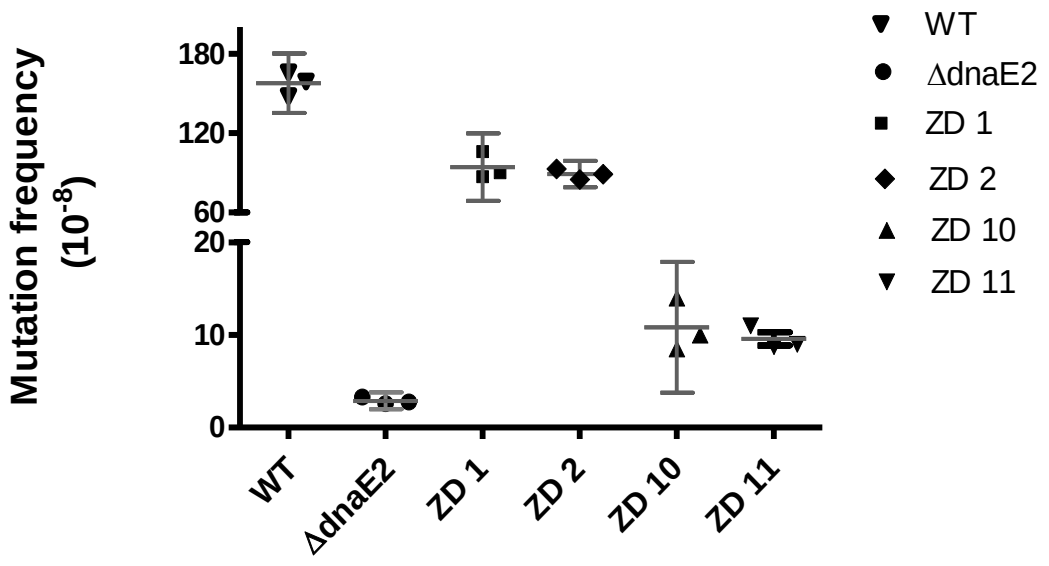


Figure S2: The conserved DnaE1 PHP domain residues are not required for UV-induced mutagenesis. UV-induced mutation frequencies of WT, Δ*dnaE2*, ZD 1, ZD 2, ZD 10 and ZD 11 following UV-irradiation. Mycobacterial cultures of WT mc^2155 and mutant strains were grown until log phase harvested, re-suspended and UV irradiated at 25 mJ/cm^2. After irradiation, cells were rescued at 37°C for 4 h and plated on 7H10 OADC containing 200 µg/ml Rif. Data are from a representative experiment performed in triplicate. Error bars represent the 95% confidence interval.

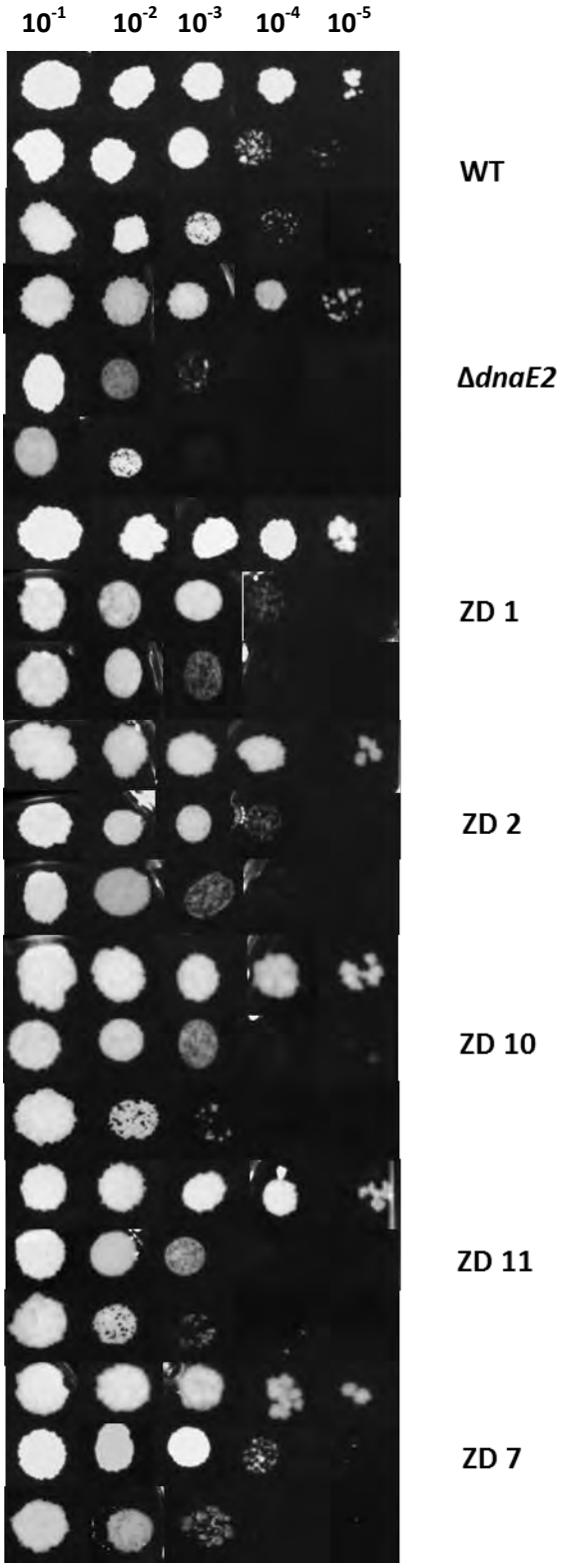


Figure S3: The conserved DnaE1 PHP domain residues are not required for damage tolerance. Log_{10}-fold dilutions of the mycobacterial cultures (10^{-1}-10^{-5}) were spotted on 7H10 OADC media without MMC (top row for each strain), with 0.02 µg/ml MMC (second row for each strain); and 0.04 µg/ml MMC (third row for each strain).

REFERENCES

ADAMS, K. N., SZUMOWSKI, J. D. & RAMAKRISHNAN, L. 2014. Verapamil, and its metabolite norverapamil, inhibit macrophage-induced, bacterial efflux pump-mediated tolerance to multiple anti-tubercular drugs. *J Infect Dis,* 210**,** 456-66.

ADAMS, K. N., TAKAKI, K., CONNOLLY, L. E., WIEDENHOFT, H., WINGLEE, K., HUMBERT, O., EDELSTEIN, P. H., COSMA, C. L. & RAMAKRISHNAN, L. 2011. Drug tolerance in replicating mycobacteria mediated by a macrophage-induced efflux mechanism. *Cell,* 145**,** 39-53.

AKERLEY, B. J., RUBIN, E. J., NOVICK, V. L., AMAYA, K., JUDSON, N. & MEKALANOS, J. J. 2002. A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. *Proc Natl Acad Sci USA,* 99**,** 966-71.

ALDRIDGE, B. B., FERNANDEZ-SUAREZ, M., HELLER, D., AMBRAVANESWARAN, V., IRIMIA, D., TONER, M. & FORTUNE, S. M. 2012. Asymmetry and aging of mycobacterial cells lead to variable growth and antibiotic susceptibility. *Science,* 335**,** 100-4.

ALMEIDA DA SILVA, P. E. & PALOMINO, J. C. 2011. Molecular basis and mechanisms of drug resistance in *Mycobacterium tuberculosis*: classical and new drugs. *J Antimicrob Chemother,* 66**,** 1417-30.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

ANDERSSON, D. I., KOSKINIEMI, S. & HUGHES, D. 2010. Biological roles of translesion synthesis DNA polymerases in eubacteria. *Mol Microbiol,* 77**,** 540-8.

ANH, D. D., BORGDORFF, M. W., VAN, L. N., LAN, N. T., VAN GORKOM, T., KREMER, K. & VAN SOOLINGEN, D. 2000. *Mycobacterium tuberculosis* Beijing genotype emerging in Vietnam. *Emerg Infect Dis,* 6**,** 302-5.

ARAVIND, L. & KOONIN, E. V. 1998. Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Res,* 26**,** 3746-52.

ARJES, H. A., KRIEL, A., SORTO, N. A., SHAW, J. T., WANG, J. D. & LEVIN, P. A. 2014. Failsafe mechanisms couple division and DNA replication in bacteria. *Curr Biol,* 24**,** 2149-55.

BAILEY, S., WING, R. A. & STEITZ, T. A. 2006. The structure of *T. aquaticus* DNA polymerase III is distinct from eukaryotic replicative DNA polymerases. *Cell,* 126**,** 893-904.

BALCELLS, M. E., THOMAS, S. L., GODFREY-FAUSSETT, P. & GRANT, A. D. 2006. Isoniazid preventive therapy and risk for resistant tuberculosis. *Emerg Infect Dis,* 12**,** 744-51.

BANOS, B., LAZARO, J. M., VILLAR, L., SALAS, M. & DE VEGA, M. 2008. Characterization of a *Bacillus subtilis* 64-kDa DNA polymerase X potentially involved in DNA repair. *J Mol Biol,* 384**,** 1019-28.

BAÑOS, B., LÁZARO, J. M., VILLAR, L., SALAS, M. & DE VEGA, M. 2008. Editing of misaligned 3′-termini by an intrinsic 3′–5′ exonuclease activity residing in the PHP domain of a family X DNA polymerase. *Nucleic Acids Res.,* 36**,** 5736-5749.

BARROS, T., GUENTHER, J., KELCH, B., ANAYA, J., PRABHAKAR, A., O'DONNELL, M., KURIYAN, J. & LAMERS, M. H. 2013. A structural role for the PHP domain in *E. coli* DNA polymerase III. *BMC Struct Biol,* 13**,** 1472-6807.

BARRY, C. E., 3RD, BOSHOFF, H. I., DARTOIS, V., DICK, T., EHRT, S., FLYNN, J., SCHNAPPINGER, D., WILKINSON, R. J. & YOUNG, D. 2009. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nat Rev Microbiol,* 7**,** 845-55.

BERTRAM, J. G., BLOOM, L. B., O'DONNELL, M. & GOODMAN, M. F. 2004. Increased dNTP binding affinity reveals a nonprocessive role for *Escherichia coli* beta clamp with DNA polymerase IV. *J Biol Chem,* 279**,** 33047-50.

BISWAS, T., RESTO-ROLDAN, E., SAWYER, S. K., ARTSIMOVITCH, I. & TSODIKOV, O. V. 2013. A novel non-radioactive primase-pyrophosphatase activity assay and its application to the discovery of inhibitors of *Mycobacterium tuberculosis* primase DnaG. *Nucleic Acids Res,* 41**,** 24.

BJEDOV, I., DASGUPTA, C. N., SLADE, D., LE BLASTIER, S., SELVA, M. & MATIC, I. 2007. Involvement of *Escherichia coli* DNA polymerase IV in tolerance of cytotoxic alkylating DNA lesions in vivo. *Genetics,* 176**,** 1431-40.

BLANCO, D., PEREZ-HERRAN, E., CACHO, M., BALLELL, L., CASTRO, J., GONZALEZ DEL RIO, R., LAVANDERA, J. L., REMUINAN, M. J., RICHARDS, C., RULLAS, J., VAZQUEZ-MUNIZ, M. J., WOLDU, E., ZAPATERO-GONZALEZ, M. C., ANGULO-BARTUREN, I., MENDOZA, A. & BARROS, D. 2015. *Mycobacterium tuberculosis* Gyrase Inhibitors (MGI) as a New Class of Antitubercular Drugs. *Antimicrob Agents Chemother,* 12**,** 03913-14.

BORDEN, A., O'GRADY, P. I., VANDEWIELE, D., FERNANDEZ DE HENESTROSA, A. R., LAWRENCE, C. W. & WOODGATE, R. 2002. *Escherichia col*i DNA polymerase III can

replicate efficiently past a T-T cis-syn cyclobutane dimer if DNA polymerase V and the 3' to 5' exonuclease proofreading function encoded by *dnaQ* are inactivated. *J Bacteriol,* 184**,** 2674-81.

BORRELL, S. & GAGNEUX, S. 2011. Strain diversity, epistasis and the evolution of drug resistance in *Mycobacterium tuberculosis*. *Clin Microbiol Infect,* 17**,** 815-20.

BOSHOFF, H. I. M., REED, M. B., BARRY III, C. E. & MIZRAHI, V. 2003. DnaE2 contributes to *in vivo* survival and the emergence of drug resistance in *Mycobacterium tuberculosis Cell,* 133**,** 183-193.

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. *J. Bacteriol.,* 183**,** 4459-4467.

BRUCK, I., GOODMAN, M. F. & O'DONNELL, M. 2003. The essential C family DnaE polymerase is error-prone and efficient at lesion bypass. *J. Biol. Chem.,* 278**,** 44361-44368.

CIRZ, R. T., O'NEILL, B. M., HAMMOND, J. A., HEAD, S. R. & ROMESBERG, F. E. 2006. Defining the *Pseudomonas aeruginosa* SOS response and its role in the global response to the antibiotic ciprofloxacin. *J Bacteriol,* 188**,** 7101-10.

CLAUSEN, A. R., MURRAY, M. S., PASSER, A. R., PEDERSEN, L. C. & KUNKEL, T. A. 2013. Structure-function analysis of ribonucleotide bypass by B family DNA replicases. *Proc Natl Acad Sci U S A,* 110**,** 16802-7.

COLANGELI, R., ARCUS, V. L., CURSONS, R. T., RUTHE, A., KARALUS, N., COLEY, K., MANNING, S. D., KIM, S., MARCHIANO, E. & ALLAND, D. 2014. Whole genome sequencing of *Mycobacterium tuberculosis* reveals slow growth and low mutation rates during latent infections in humans. *PLoS One,* 9.

COLE, S. T., EIGLMEIER, K., PARKHILL, J., JAMES, K. D., THOMSON, N. R., WHEELER, P. R., HONORE, N., GARNIER, T., CHURCHER, C., HARRIS, D., MUNGALL, K., BASHAM, D., BROWN, D., CHILLINGWORTH, T., CONNOR, R., DAVIES, R. M., DEVLIN, K., DUTHOY, S., FELTWELL, T., FRASER, A., HAMLIN, N., HOLROYD, S., HORNSBY, T., JAGELS, K., LACROIX, C., MACLEAN, J., MOULE, S., MURPHY, L., OLIVER, K., QUAIL, M. A., RAJANDREAM, M. A., RUTHERFORD, K. M., RUTTER, S., SEEGER, K., SIMON, S., SIMMONDS, M., SKELTON, J., SQUARES, R., SQUARES, S., STEVENS, K., TAYLOR, K.,

WHITEHEAD, S., WOODWARD, J. R. & BARRELL, B. G. 2001. Massive gene decay in the leprosy bacillus. *Nature,* 409**,** 1007-11.

COSCOLLA, M. & GAGNEUX, S. 2010. Does *M. tuberculosis* genomic diversity explain disease diversity? *Drug Discov Today Dis Mech,* 7**,** e43-e59.

COSMA, C. L., HUMBERT, O. & RAMAKRISHNAN, L. 2004. Superinfecting mycobacteria home to established tuberculous granulomas. *Nat Immunol,* 5**,** 828-35.

COURCELLE, J., KHODURSKY, A., PETER, B., BROWN, P. O. & HANAWALT, P. C. 2001. Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient Escherichia coli. *Genetics,* 158**,** 41-64.

DALRYMPLE, B. P. 2001. A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. *Proc. Natl. Acad. Sci. U.S.A,* 98**,** 11627-11632.

DARWIN, K. H. & NATHAN, C. F. 2005. Role for nucleotide excision repair in virulence of Mycobacterium tuberculosis. *Infect Immun,* 73**,** 4581-7.

DAVIS, E. O., SPRINGER, B., GOPAUL, K. K., PAPAVINASASUNDARAM, K. G., SANDER, P. & BOTTGER, E. C. 2002. DNA damage induction of recA in Mycobacterium tuberculosis independently of RecA and LexA. *Mol Microbiol,* 46**,** 791-800.

DE JONG, B. C., HILL, P. C., AIKEN, A., AWINE, T., ANTONIO, M., ADETIFA, I. M., JACKSON-SILLAH, D. J., FOX, A., DERIEMER, K., GAGNEUX, S., BORGDORFF, M. W., MCADAM, K. P., CORRAH, T., SMALL, P. M. & ADEGBOLA, R. A. 2008. Progression to active tuberculosis, but not transmission, varies by *Mycobacterium tuberculosis* lineage in The Gambia. *J Infect Dis,* 198**,** 1037-43.

DEGNEN, G. E. & COX, E. C. 1974. Conditional mutator gene in *Escherichia coli*: isolation, mapping, and effector studies. *J Bacteriol,* 117**,** 477-87.

DOS VULTOS, T., MESTRE, O., RAUZIER, J., GOLEC, M., RASTOGI, N., RASOLOFO, V., TONJUM, T., SOLA, C., MATIC, I. & GICQUEL, B. 2008. Evolution and diversity of clonal bacteria: the paradigm of *Mycobacterium tuberculosis*. *PLoS One,* 3**,** 0001538.

DOS VULTOS, T., MESTRE, O., TONJUM, T. & GICQUEL, B. 2009. DNA repair in Mycobacterium tuberculosis revisited. *FEMS Microbiol Rev,* 33**,** 471-87.

DUONG, D. A., NGUYEN, T. H., NGUYEN, T. N., DAI, V. H., DANG, T. M., VO, S. K., DO, D. A., NGUYEN, V. V., NGUYEN, H. D., DINH, N. S., FARRAR, J. & CAWS, M. 2009. Beijing genotype of *Mycobacterium tuberculosis* is significantly associated with high-level fluoroquinolone resistance in Vietnam. *Antimicrob Agents Chemother,* 53**,** 4835-9.

DUTTA, N. K., MEHRA, S., DIDIER, P. J., ROY, C. J., DOYLE, L. A., ALVAREZ, X., RATTERREE, M., BE, N. A., LAMICHHANE, G., JAIN, S. K., LACEY, M. R., LACKNER, A. A. & KAUSHAL, D. 2010. Genetic requirements for the survival of tubercle bacilli in primates. *J Infect Dis,* 201**,** 1743-52.

EBRAHIMI-RAD, M., BIFANI, P., MARTIN, C., KREMER, K., SAMPER, S., RAUZIER, J., KREISWIRTH, B., BLAZQUEZ, J., JOUAN, M., VAN SOOLINGEN, D. & GICQUEL, B. 2003. Mutations in putative mutator genes of *Mycobacterium tuberculosis* strains of the W-Beijing family. *Emerg Infect Dis,* 9**,** 838-45.

EHRT, S., GUO, X. V., HICKEY, C. M., RYOU, M., MONTELEONE, M., RILEY, L. W. & SCHNAPPINGER, D. 2005. Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Res,* 33.

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

ERNST, J. D. 2012. The immunological life cycle of tuberculosis. *Nat Rev Immunol,* 12**,** 581-91.

EVANS, R. J., DAVIES, D. R., BULLARD, J. M., CHRISTENSEN, J., GREEN, L. S., GUILES, J. W., PATA, J. D., RIBBLE, W. K., JANJIC, N. & JARVIS, T. C. 2008. Structure of PolC reveals unique DNA binding and fidelity determinants. *Proc Natl Acad Sci U S A,* 105**,** 20695-700.

FARHAT, M. R., SHAPIRO, B. J., KIESER, K. J., SULTANA, R., JACOBSON, K. R., VICTOR, T. C., WARREN, R. M., STREICHER, E. M., CALVER, A., SLOUTSKY, A., KAUR, D., POSEY, J. E., PLIKAYTIS, B., OGGIONI, M. R., GARDY, J. L., JOHNSTON, J. C., RODRIGUES, M., TANG, P. K., KATO-MAEDA, M., BOROWSKY, M. L., MUDDUKRISHNA, B., KREISWIRTH, B. N., KUREPINA, N., GALAGAN, J., GAGNEUX, S., BIRREN, B., RUBIN, E. J., LANDER, E. S., SABETI, P. C. & MURRAY, M. 2013. Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. *Nat Genet,* 45**,** 1183-9.

FIJALKOWSKA, I. J. & SCHAAPER, R. M. 1995. Effects of *Escherichia coli dnaE* antimutator alleles in a proofreading-deficient mutD5 strain. *J Bacteriol,* 177**,** 5979-86.

FIJALKOWSKA, I. J. & SCHAAPER, R. M. 1996. Mutants in the Exo I motif of *Escherichia coli dnaQ*: defective proofreading and inviability due to error catastrophe. *Proc Natl Acad Sci U S A,* 93**,** 2856-61.

FIJALKOWSKA, I. J., SCHAAPER, R. M. & JONCZYK, P. 2012. DNA replication fidelity in *Escherichia coli*: a multi-DNA polymerase affair. *FEMS Microbiol Rev,* 36**,** 1105-21.

FORD, C. B. 2012. *The evolution of drug resistant Mycobacterium tuberculosis.* PhD thesis, Harvard School of Public Health.

FORD, C. B., LIN, P. L., CHASE, M. R., SHAH, R. R., IARTCHOUK, O., GALAGAN, J., MOHAIDEEN, N., IOERGER, T. R., SACCHETTINI, J. C., LIPSITCH, M., FLYNN, J. L. & FORTUNE, S. M. 2011. Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. *Nat Genet,* 43**,** 482-6.

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. *Nat Genet,* 45**,** 784-90.

FOTI, J. J., DEVADOSS, B., WINKLER, J. A., COLLINS, J. J. & WALKER, G. C. 2012. Oxidation of the guanine nucleotide pool underlies cell death by bactericidal antibiotics. *Science,* 336**,** 315-9.

GALHARDO, R. S., ROCHA, R. P., MARQUES, M. V. & MENCK, C. F. M. 2005. An SOS-regulated operon involved in damage-inducible mutagenesis in *Caulobacter crescentus*. *Nucleic Acids Res.,* 33**,** 2603-2614.

GAMULIN, V., CETKOVIC, H. & AHEL, I. 2004. Identification of a promoter motif regulating the major DNA damage response mechanism of *Mycobacterium tuberculosis*. *FEMS Microbiol Lett,* 238**,** 57-63.

GAWEL, D., HAMILTON, M. D. & SCHAAPER, R. M. 2008. A novel mutator of *Escherichia coli* carrying a defect in the dgt gene, encoding a dGTP triphosphohydrolase. *J. Bacteriol.,* 190**,** 6931-6939.

GEORGESCU, R. E., KURTH, I. & O'DONNELL, M. E. 2011. Single-molecule studies reveal the function of a third polymerase in the replisome. *Nat Struct Mol Biol,* 19**,** 113-6.

GEORGESCU, R. E., YURIEVA, O., KIM, S. S., KURIYAN, J., KONG, X. P. & O'DONNELL, M. 2008. Structure of a small-molecule inhibitor of a DNA polymerase sliding clamp. *Proc Natl Acad Sci U S A,* 105**,** 11116-21.

GERDES, S. Y., SCHOLLE, M. D., CAMPBELL, J. W., BALAZSI, G., RAVASZ, E., DAUGHERTY, M. D., SOMERA, A. L., KYRPIDES, N. C., ANDERSON, I., GELFAND, M. S., BHATTACHARYA, A., KAPATRAL, V., D'SOUZA, M., BAEV, M. V., GRECHKIN, Y., MSEEH, F., FONSTEIN, M.

Y., OVERBEEK, R., BARABASI, A. L., OLTVAI, Z. N. & OSTERMAN, A. L. 2003. Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J Bacteriol,* 185**,** 5673-84.

GILL, W. P., HARIK, N. S., WHIDDON, M. R., LIAO, R. P., MITTLER, J. E. & SHERMAN, D. R. 2009. A replication clock for *Mycobacterium tuberculosis*. *Nat Med,* 15**,** 211-4.

GILLESPIE, S. H. 2001. Antibiotic resistance in the absence of selective pressure. *Int J Antimicrob Agents,* 17**,** 171-6.

GILLESPIE, S. H. 2002. Evolution of drug resistance in *Mycobacterium tuberculosis*: clinical and molecular perspective. *Antimicrob Agents Chemother,* 46**,** 267-74.

GLYNN, J. R., WHITELEY, J., BIFANI, P. J., KREMER, K. & VAN SOOLINGEN, D. 2002. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. *Emerg Infect Dis,* 8**,** 843-9.

GON, S., NAPOLITANO, R., ROCHA, W., COULON, S. & FUCHS, R. P. 2011. Increase in dNTP pool size during the DNA damage response plays a key role in spontaneous and induced-mutagenesis in Escherichia coli. *Proc Natl Acad Sci U S A,* 108**,** 19311-6.

GONG, C., BONGIORNO, P., MARTINS, A., STEPHANOU, N. C., ZHU, H., SHUMAN, S. & GLICKMAN, M. S. 2005. Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. *Nat Struct Mol Biol,* 12**,** 304-12.

GONG, C., MARTINS, A., BONGIORNO, P., GLICKMAN, M. & SHUMAN, S. 2004. Biochemical and genetic analysis of the four DNA ligases of mycobacteria. *J Biol Chem,* 279**,** 20594-606.

GOODMAN, M. F. 2002. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu Rev Biochem,* 71**,** 17-50.

GORDHAN, B. G., ANDERSEN, S. J., DE MEYER, A. R. & MIZRAHI, V. 1996. Construction by homologous recombination and phenotypic characterization of a DNA polymerase domain polA mutant of *Mycobacterium smegmatis*. *Gene,* 178**,** 125-30.

GORNA, A. E., BOWATER, R. P. & DZIADEK, J. 2010. DNA repair systems and the pathogenesis of *Mycobacterium tuberculosis*: varying activities at different stages of infection. *Clin Sci,* 119**,** 187-202.

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). *Proc Natl Acad Sci U S A,* 96**,** 11554-9.

GRIFFIN, J. E., GAWRONSKI, J. D., DEJESUS, M. A., IOERGER, T. R., AKERLEY, B. J. & SASSETTI, C. M. 2011. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog,* 7**,** 29.

GUPTA, A., KAUL, A., TSOLAKI, A. G., KISHORE, U. & BHAKTA, S. 2012. *Mycobacterium tuberculosis*: immune evasion, latency and reactivation. *Immunobiology,* 217**,** 363-74.

GUPTA, R., BARKAN, D., REDELMAN-SIDI, G., SHUMAN, S. & GLICKMAN, M. S. 2011. Mycobacteria exploit three genetically distinct DNA double-strand break repair pathways. *Mol Microbiol,* 79**,** 316-30.

GUPTA, S., TYAGI, S. & BISHAI, W. R. 2015. Verapamil increases the bactericidal activity of bedaquiline against *Mycobacterium tuberculosis* in a mouse model. *Antimicrob Agents Chemother,* 59**,** 673-6.

GUTHLEIN, C., WANNER, R. M., SANDER, P., DAVIS, E. O., BOSSHARD, M., JIRICNY, J., BOTTGER, E. C. & SPRINGER, B. 2009. Characterization of the mycobacterial NER system reveals novel functions of the uvrD1 helicase. *J Bacteriol,* 191**,** 555-62.

HAMEED, P. S., PATIL, V., SOLAPURE, S., SHARMA, U., MADHAVAPEDDI, P., RAICHURKAR, A., CHINNAPATTU, M., MANJREKAR, P., SHANBHAG, G., PUTTUR, J., SHINDE, V., MENASINAKAI, S., RUDRAPATANA, S., ACHAR, V., AWASTHY, D., NANDISHAIAH, R., HUMNABADKAR, V., GHOSH, A., NARAYAN, C., RAMYA, V. K., KAUR, P., SHARMA, S., WERNGREN, J., HOFFNER, S., PANDUGA, V., KUMAR, C. N., REDDY, J., KUMAR, K. N. M., GANGULY, S., BHARATH, S., BHEEMARAO, U., MUKHERJEE, K., ARORA, U., GAONKAR, S., COULSON, M., WATERSON, D., SAMBANDAMURTHY, V. K. & DE SOUSA, S. M. 2014a. Novel N-linked aminopiperidine-based gyrase inhibitors with improved hERG and in vivo efficacy against *Mycobacterium tuberculosis*. *J Med Chem,* 57**,** 4889-905.

HAMEED, P. S., RAICHURKAR, A., MADHAVAPEDDI, P., MENASINAKAI, S., SHARMA, S., KAUR, P., NANDISHAIAH, R., PANDUGA, V., REDDY, J., SAMBANDAMURTHY, V. K. & SRIRAM, D. 2014b. Benzimidazoles: novel mycobacterial gyrase inhibitors from scaffold morphing. *ACS Med Chem Lett,* 5**,** 820-5.

HARRIS, K. K., FAY, A., YAN, H. G., KUNWAR, P., SOCCI, N. D., POTTABATHINI, N., JUVENTHALA, R. R., DJABALLAH, H. & GLICKMAN, M. S. 2014. Novel imidazoline antimicrobial scaffold that inhibits DNA replication with activity against mycobacteria and drug resistant Gram-positive cocci. *ACS Chem Biol,* 9**,** 2572-83.

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. *Infect Immun,* 82**,** 3177-85.

HERSHBERG, R., LIPATOV, M., SMALL, P. M., SHEFFER, H., NIEMANN, S., HOMOLKA, S., ROACH, J. C., KREMER, K., PETROV, D. A., FELDMAN, M. W. & GAGNEUX, S. 2008. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol,* 6**,** 0060311.

HORIUCHI, T., MAKI, H. & SEKIGUCHI, M. 1978. A new conditional lethal mutator (dnaQ49) in *Escherichia col*i K12. *Mol Gen Genet,* 163**,** 277-83.

HUANG, H. Y., TSAI, Y. S., LEE, J. J., CHIANG, M. C., CHEN, Y. H., CHIANG, C. Y., LIN, N. T. & TSAI, P. J. 2010. Mixed infection with Beijing and non-Beijing strains and drug resistance pattern of *Mycobacterium tuberculosis*. *J Clin Microbiol,* 48**,** 4474-80.

HUANG, Y. P. & ITO, J. 1999. DNA polymerase C of the thermophilic bacterium *Thermus aquaticus*: classification and phylogenetic analysis of the family C DNA polymerases. *J Mol Evol,* 48**,** 756-69.

ITO, J. & BRAITHWAITE, D. K. 1991. Compilation and alignment of DNA polymerase sequences. *Nucleic Acids Res,* 19**,** 4045-57.

JAROSZ, D. F., BEUNING, P. J., COHEN, S. E. & WALKER, G. C. 2007a. Y-family DNA polymerases in *Escherichia coli*. *Trends Microbiol,* 15**,** 70-7.

JAROSZ, D. F., GODOY, V. G., DELANEY, J. C., ESSIGMANN, J. M. & WALKER, G. C. 2006. A single amino acid governs enhanced activity of DinB DNA polymerases on damaged templates. *Nature,* 439**,** 225-8.

JAROSZ, D. F., GODOY, V. G. & WALKER, G. C. 2007b. Proficient and accurate bypass of persistent DNA lesions by DinB DNA polymerases. *Cell Cycle,* 6**,** 817-22.

JIANG, Q., KARATA, K., WOODGATE, R., COX, M. M. & GOODMAN, M. F. 2009. The active form of DNA polymerase V is UmuD'(2)C-RecA-ATP. *Nature,* 460**,** 359-63.

JOHNSON, A. & O'DONNELL, M. 2005. Cellular DNA replicases: Components and dynamics at the replication fork. *Annu. Rev. Biochem.,* 74**,** 283-315.

JOHNSON, R., WARREN, R. M., VAN DER SPUY, G. D., GEY VAN PITTIUS, N. C., THERON, D., STREICHER, E. M., BOSMAN, M., COETZEE, G. J., VAN HELDEN, P. D. & VICTOR, T. C. 2010. Drug-resistant tuberculosis epidemic in the Western Cape driven by a virulent Beijing genotype strain. *Int J Tuberc Lung Dis,* 14**,** 119-21.

JOYCE, G., WILLIAMS, K. J., ROBB, M., NOENS, E., TIZZANO, B., SHAHREZAEI, V. & ROBERTSON, B. D. 2012. Cell division site placement and asymmetric growth in mycobacteria. *PLoS One,* 7**,** 10.

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. & MIZRAHI, V. 2010. Role of the *dinB* homologs Rv1537 and Rv3056 in *Mycobacterium tuberculosis*. *J. Bacteriol.,* 192**,** 2220-2227.

KARKARE, S., CHUNG, T. T., COLLIN, F., MITCHENALL, L. A., MCKAY, A. R., GREIVE, S. J., MEYER, J. J., LALL, N. & MAXWELL, A. 2013. The naphthoquinone diospyrin is an inhibitor of DNA gyrase with a novel mechanism of action. *J Biol Chem,* 288**,** 5149-56.

KATO-MAEDA, M., SHANLEY, C. A., ACKART, D., JARLSBERG, L. G., SHANG, S., OBREGON-HENAO, A., HARTON, M., BASARABA, R. J., HENAO-TAMAYO, M., BARROZO, J. C., ROSE, J., KAWAMURA, L. M., COSCOLLA, M., FOFANOV, V. Y., KOSHINSKY, H., GAGNEUX, S., HOPEWELL, P. C., ORDWAY, D. J. & ORME, I. M. 2012. Beijing sublineages of *Mycobacterium tuberculosis* differ in pathogenicity in the guinea pig. *Clin Vaccine Immunol,* 19**,** 1227-37.

KESAVAN, A. K., BROOKS, M., TUFARIELLO, J., CHAN, J. & MANABE, Y. C. 2009. Tuberculosis genes expressed during persistence and reactivation in the resistant rabbit model. *Tuberculosis (Edinb),* 89**,** 17-21.

KESTER, J. C. & FORTUNE, S. M. 2014. Persisters and beyond: mechanisms of phenotypic drug resistance and drug tolerance in bacteria. *Crit Rev Biochem Mol Biol,* 49**,** 91-101.

KIESER, K. J. & RUBIN, E. J. 2014. How sisters grow apart: mycobacterial growth and division. *Nat Rev Microbiol,* 12**,** 550-62.

KLANN, A. G., BELANGER, A. E., ABANES-DE MELLO, A., LEE, J. Y. & HATFULL, G. F. 1998. Characterization of the *dnaG* locus in *Mycobacterium smegmatis* reveals linkage of DNA replication and cell division. *J Bacteriol,* 180**,** 65-72.

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**,** 863-8.

KORNBERG, A. & BAKER, T. A. 2005. *DNA Replication* Sausalito, CA, University Science Books.

KOSER, C. U., BRYANT, J. M., BECQ, J., TOROK, M. E., ELLINGTON, M. J., MARTI-RENOM, M. A., CARMICHAEL, A. J., PARKHILL, J., SMITH, G. P. & PEACOCK, S. J. 2013. Whole-genome sequencing for rapid susceptibility testing of *M. tuberculosis*. *N Engl J Med.,* 369**,** 290-2.

KUMARI, A., MINKO, I. G., HARBUT, M. B., FINKEL, S. E., GOODMAN, M. F. & LLOYD, R. S. 2008. Replication bypass of interstrand cross-link intermediates by *Escherichia coli* DNA polymerase IV. *J Biol Chem,* 283**,** 27433-7.

KURTH, I. & O'DONNELL, M. 2013. New insights into replisome fluidity during chromosome replication. *Trends Biochem Sci,* 38**,** 195-203.

KURTHKOTI, K., KUMAR, P., JAIN, R. & VARSHNEY, U. 2008. Important role of the nucleotide excision repair pathway in Mycobacterium smegmatis in conferring protection against commonly encountered DNA-damaging agents. *Microbiology,* 154**,** 2776-85.

KYSELA, D. T., BROWN, P. J., HUANG, K. C. & BRUN, Y. V. 2013. Biological consequences and advantages of asymmetric bacterial growth. *Annu Rev Microbiol,* 67**,** 417-35.

LAMERS, M. H., GEORGESCU, R. E., LEE, S.-G., O'DONNELL, M. & KURIYAN, J. 2006. Crystal structure of the catalytic [alpha] subunit of *E. coli* replicative DNA polymerase III. *Cell,* 126**,** 881-892.

LAMERS, M. H. & O'DONNELL, M. 2008. A consensus view of DNA binding by the C family of replicative DNA polymerases. *Proc Natl Acad Sci U S A,* 105**,** 20565-6.

LARSEN, M. H. 2000. *Some common methods in Mycobacterial genetics. In Molecular Genetics of Mycobacteria (2nd edition): Hatfull G, Jacobs WR Jr (eds)* Washington D. C, ASM Press.

LE CHATELIER, E., BÉCHEREL, O. J., D'ALENÇON, E., CANCEILL, D., EHRLICH, S. D., FUCHS, R. P. & JANNIÈRE, L. 2004. Involvement of DnaE, the second replicative DNA polymerase from *Bacillus subtilis*, in DNA mutagenesis. *J. Biol. Chem.,* 279**,** 1757-1767.

LEONARD, A. C. & GRIMWADE, J. E. 2011. Regulation of DnaA assembly and activity: taking directions from the genome. *Annu Rev Microbiol,* 65**,** 19-35.

LI, X. Z. & NIKAIDO, H. 2009. Efflux-mediated drug resistance in bacteria: an update. *Drugs,* 69**,** 1555-623.

LILLEBAEK, T., DIRKSEN, A., BAESS, I., STRUNGE, B., THOMSEN, V. O. & ANDERSEN, A. B. 2002. Molecular evidence of endogenous reactivation of *Mycobacterium tuberculosis* after 33 years of latent infection. *J Infect Dis,* 185**,** 401-4.

LIN, P. L., FORD, C. B., COLEMAN, M. T., MYERS, A. J., GAWANDE, R., IOERGER, T., SACCHETTINI, J., FORTUNE, S. M. & FLYNN, J. L. 2014. Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing. *Nat Med,* 20**,** 75-9.

LINCOLN, E. M. 1954. The effect of antimicrobial therapy on the prognosis of primary tuberculosis in children. *Am Rev Tuberc,* 69**,** 682-9.

LLERES, D., SWIFT, S. & LAMOND, A. I. 2007. Detecting protein-protein interactions in vivo with FRET using multiphoton fluorescence lifetime imaging microscopy (FLIM). *Curr Protoc Cytom,* 12.

MACHADO, D., COUTO, I., PERDIGAO, J., RODRIGUES, L., PORTUGAL, I., BAPTISTA, P., VEIGAS, B., AMARAL, L. & VIVEIROS, M. 2012. Contribution of efflux to the emergence of isoniazid and multidrug resistance in *Mycobacterium tuberculosis*. *PLoS One,* 7**,** 6.

MAKAROVA, A. V., NICK MCELHINNY, S. A., WATTS, B. E., KUNKEL, T. A. & BURGERS, P. M. 2014. Ribonucleotide incorporation by yeast DNA polymerase zeta. *DNA Repair,* 18**,** 63-7.

MATHEWS, C. K. 2014. Deoxyribonucleotides as genetic and metabolic regulators. *FASEB J.,* 28**,** 3832-3840.

MCHENRY, C. S. 2011a. Bacterial replicases and related polymerases. *Curr Opin Chem Biol,* 15**,** 587-94.

MCHENRY, C. S. 2011b. DNA replicases from a bacterial perspective. *Annu Rev Biochem,* 80**,** 403-36.

MCINERNEY, P., JOHNSON, A., KATZ, F. & O'DONNELL, M. 2007. Characterization of a triple DNA polymerase replisome. *Mol Cell,* 27**,** 527-38.

MCKENZIE, G. J., LEE, P. L., LOMBARDO, M. J., HASTINGS, P. J. & ROSENBERG, S. M. 2001. SOS mutator DNA polymerase IV functions in adaptive mutation and not adaptive amplification. *Mol Cell,* 7**,** 571-9.

MERENS, A., MATRAT, S., AUBRY, A., LASCOLS, C., JARLIER, V., SOUSSY, C. J., CAVALLO, J. D. & CAMBAU, E. 2009. The pentapeptide repeat proteins MfpAMt and QnrB4 exhibit opposite effects on DNA gyrase catalytic reactions and on the ternary gyrase-DNA-quinolone complex. *J Bacteriol,* 191**,** 1587-94.

MESTRE, O., LUO, T., DOS VULTOS, T., KREMER, K., MURRAY, A., NAMOUCHI, A., JACKSON, C., RAUZIER, J., BIFANI, P., WARREN, R., RASOLOFO, V., MEI, J., GAO, Q. & GICQUEL, B. 2011. Phylogeny of *Mycobacterium tuberculosis* Beijing strains constructed from polymorphisms in genes involved in DNA replication, recombination and repair. *PLoS One,* 6**,** 0016020.

MINIAS, A. E., BRZOSTEK, A. M., MINIAS, P. & DZIADEK, J. 2015. The Deletion of rnhB in *Mycobacterium smegmatis* does not affect the level of RNase HII substrates or influence genome stability. *PLoS One,* 10.

MINKO, I. G., YAMANAKA, K., KOZEKOV, I. D., KOZEKOVA, A., INDIANI, C., O'DONNELL, M. E., JIANG, Q., GOODMAN, M. F., RIZZO, C. J. & LLOYD, R. S. 2008. Replication bypass of the acrolein-mediated deoxyguanine DNA-peptide cross-links by DNA polymerases of the DinB family. *Chem Res Toxicol,* 21**,** 1983-90.

MIZRAHI, V. & ANDERSEN, S. J. 1998. DNA repair in *Mycobacterium tuberculosis*. What have we learnt from the genome sequence? *Mol Microbiol,* 29**,** 1331-9.

MIZRAHI, V. & HUBERTS, P. 1996. Deoxy- and dideoxynucleotide discrimination and identification of critical 5' nuclease domain residues of the DNA polymerase I from *Mycobacterium tuberculosis*. *Nucleic Acids Res,* 24**,** 4845-52.

MOKROUSOV, I. 2014. Widely-used laboratory and clinical *Mycobacterium tuberculosis* strains: to what extent they are representative of their phylogenetic lineages? *Tuberculosis (Edinb).* 94**,** 355-6.

MONOT, M., HONORE, N., GARNIER, T., ZIDANE, N., SHERAFI, D., PANIZ-MONDOLFI, A., MATSUOKA, M., TAYLOR, G. M., DONOGHUE, H. D., BOUWMAN, A., MAYS, S., WATSON, C., LOCKWOOD, D., KHAMESIPOUR, A., DOWLATI, Y., JIANPING, S., REA, T. H., VERA-CABRERA, L., STEFANI, M. M., BANU, S., MACDONALD, M., SAPKOTA, B. R., SPENCER, J. S., THOMAS, J., HARSHMAN, K., SINGH, P., BUSSO, P., GATTIKER, A.,

ROUGEMONT, J., BRENNAN, P. J. & COLE, S. T. 2009. Comparative genomic and phylogeographic analysis of Mycobacterium leprae. *Nat Genet,* 41**,** 1282-9.

MURLI, S. & OPPERMAN, T. 2000. A role for the *umuDC* gene products of *Escherichia coli* in increasing resistance to DNA damage in stationery phase by inhibiting the transition to exponential growth. *J. Bacteriol.,* 182**,** 1127-1135.

NAHID, P. & DALEY, C. L. 2006. Prevention of tuberculosis in HIV-infected patients. *Curr Opin Infect Dis,* 19**,** 189-93.

NAIR, N., DZIEDZIC, R., GREENDYKE, R., MUNIRUZZAMAN, S., RAJAGOPALAN, M. & MADIRAJU, M. V. 2009. Synchronous replication initiation in novel *Mycobacterium tuberculosis dnaA* cold-sensitive mutants. *Mol Microbiol,* 71**,** 291-304.

NAKANE, S., NAKAGAWA, N., KURAMITSU, S. & MASUI, R. 2009. Characterization of DNA polymerase X from *Thermus thermophilus* HB8 reveals the POLXc and PHP domains are both required for 3'-5' exonuclease activity. *Nucleic Acids Res,* 37**,** 2037-52.

NICK MCELHINNY, S. A., KUMAR, D., CLARK, A. B., WATT, D. L., WATTS, B. E., LUNDSTROM, E. B., JOHANSSON, E., CHABES, A. & KUNKEL, T. A. 2010. Genome instability due to ribonucleotide incorporation into DNA. *Nat Chem Biol,* 6**,** 774-81.

NIIMI, N., SASSA, A., KATAFUCHI, A., GRUZ, P., FUJIMOTO, H., BONALA, R. R., JOHNSON, F., OHTA, T. & NOHMI, T. 2009. The steric gate amino acid tyrosine 112 is required for efficient mismatched-primer extension by human DNA polymerase kappa. *Biochemistry,* 48**,** 4239-46.

O'DONNELL, M. 2006. Replisome architecture and dynamics in *Escherichia coli*. *J. Biol. Chem.,* 281**,** 10653-10656.

OHMORI, H., FRIEDBERG, E. C., FUCHS, R. P., GOODMAN, M. F., HANAOKA, F., HINKLE, D., KUNKEL, T. A., LAWRENCE, C. W., LIVNEH, Z., NOHMI, T., PRAKASH, S., TODO, T., WALKER, G. C., WANG, Z. & WOODGATE, R. 2001. The Y-family of DNA polymerases. *Mol. Cell,* 8**,** 7-8.

OPPERMAN, T., MURLI, S., SMITH, B. T. & WALKER, G. C. 1999. A model of a *umuDC*-dependent prokaryotic DNA-damage checkpoint. *Proc Natl Acad Sci USA,* 96**,** 9218-9223.

ORDONEZ, H., USON, M. L. & SHUMAN, S. 2014. Characterization of three mycobacterial DinB (DNA polymerase IV) paralogs highlights DinB2 as naturally adept at ribonucleotide incorporation. *Nucleic Acids Res*.

PARISH, T. & STOKER, N. G. 2000. Use of a flexible cassette method to generate a double unmarked *Mycobacterium tuberculosis tlyA* plcABC mutant by gene replacement. *Microbiology,* 146**,** 1969-75.

PARWATI, I., VAN CREVEL, R. & VAN SOOLINGEN, D. 2010. Possible underlying mechanisms for successful emergence of the *Mycobacterium tuberculosis* Beijing genotype strains. *Lancet Infect Dis,* 10**,** 103-11.

PIETRASZEWSKA-BOGIEL, A. & GADELLA, T. W. 2011. FRET microscopy: from principle to routine technology in cell biology. *J Microsc,* 241**,** 111-8.

PRAMMANANAN, T., PHUNPRUCH, S., JAITRONG, S. & PALITTAPONGARNPIM, P. 2012. *Mycobacterium tuberculosis uvrC* essentiality in response to UV-induced cell damage. *Southeast Asian J Trop Med Public Health,* 43**,** 370-5.

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. *Mol. Microbiol.,* 50**,** 1031-1042.

RAO, V. S., SRINIVAS, K., SUJINI, G. N. & KUMAR, G. N. 2014. Protein-protein interaction detection: methods and analysis. *Int J Proteomics,* 147648**,** 17.

REYERS-LAMOTHE, R., SHERRATT, D. J. & LEAKE, M. C. 2010. Stoichiometry and architecture of active DNA replication machinery in *Escherichia coli*. *Science,* 382**,** 498-501.

REYES-LAMOTHE, R., NICOLAS, E. & SHERRATT, D. J. 2012. Chromosome replication and segregation in bacteria. *Annu Rev Genet,* 46**,** 121-43.

REYRAT, J.-M. & KAHN, D. 2001. *Mycobacterium smegmatis*: an absurd model for tuberculosis? *Trends in Microbiol.,* 9**,** 472-473.

ROBINSON, A., CAUSER, R. J. & DIXON, N. E. 2012. Architecture and conservation of the bacterial DNA replication machinery, an underexploited drug target. *Curr Drug Targets,* 13**,** 352-72.

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. *Nat Genet,* 47**,** 677-81.

ROSCHE, W. A. & FOSTER, P. L. 2000. Determining mutation rates in bacterial populations. *Methods,* 20**,** 4-17.

SAFI, H., LINGARAJU, S., AMIN, A., KIM, S., JONES, M., HOLMES, M., MCNEIL, M., PETERSON, S. N., CHATTERJEE, D., FLEISCHMANN, R. & ALLAND, D. 2013. Evolution of high-level ethambutol-resistant tuberculosis through interacting mutations in decaprenylphosphoryl-beta-D-arabinose biosynthetic and utilization pathway genes. *Nat Genet,* 45**,** 1190-7.

SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. 1989. *Molecular cloning: A laboratory manual* New York, Cold Spring Harbour. Laboratory Press.

SAMBROOK, J. & RUSSELL, D. W. 2001. *Molecular cloning: A laboratory manual,* New York, Cold Spring Harbor.

SANDERS, G. M., DALLMAN, H. G. & MCHENRY, C. S. 2010. Reconstitution of the *B. subtilis* replisome with 13 proteins including two distinct replicases. *Mol. Cell,* 37**,** 273-281.

SANDERS, L. H., ROCKEL, A., LU, H., WOZNIAK, D. J. & SUTTON, M. D. 2006. Role of *Pseudomonas aeruginosa dinB*-encoded DNA polymerase IV in mutagenesis. *J Bacteriol,* 188**,** 8573-85.

SANDGREN, A., STRONG, M., MUTHUKRISHNAN, P., WEINER, B. K., CHURCH, G. M. & MURRAY, M. B. 2009. Tuberculosis drug resistance mutation database. *PLoS Med,* 6**,** e1000002.

SANTI, I., DHAR, N., BOUSBAINE, D., WAKAMOTO, Y. & MCKINNEY, J. D. 2013. Single-cell dynamics of the chromosome replication and cell division cycles in mycobacteria. *Nat Commun,* 4.

SANYAL, G. & DOIG, P. 2012. Bacterial DNA replication enzymes as targets for antibacterial drug discovery. *Expert Opin Drug Discov,* 7**,** 327-39.

SASSETTI, C. M., BOYD, D. H. & RUBIN, E. J. 2001. Comprehensive identification of conditionally essential genes in mycobacteria. *Proc Natl Acad Sci U S A,* 98**,** 12712-7.

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

SASSETTI, C. M. & RUBIN, E. J. 2003. Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A,* 100**,** 12989-94.

SCHMALSTIEG, A. M., SRIVASTAVA, S., BELKAYA, S., DESHPANDE, D., MEEK, C., LEFF, R., VAN OERS, N. S. & GUMBO, T. 2012. The antibiotic resistance arrow of time: efflux pump induction is a general first step in the evolution of mycobacterial drug resistance. *Antimicrob Agents Chemother,* 56**,** 4806-15.

SHARMA, A. & NAIR, D. T. 2012. MsDpo4-a DinB Homolog from *Mycobacterium smegmatis*-Is an Error-Prone DNA Polymerase That Can Promote G:T and T:G Mismatches. *J Nucleic Acids,* 285481**,** 15.

SHURTLEFF, B. W., OLLIVIERRE, J. N., TEHRANI, M., WALKER, G. C. & BEUNING, P. J. 2009. Steric gate variants of UmuC confer UV hypersensitivity on *Escherichia coli*. *J Bacteriol,* 191**,** 4815-23.

SINGH, B., NITHARWAL, R. G., RAMESH, M., PETTERSSON, B. M., KIRSEBOM, L. A. & DASGUPTA, S. 2013. Asymmetric growth and division in *Mycobacterium spp*.: compensatory mechanisms for non-medial septa. *Mol Microbiol,* 88**,** 64-76.

SMITH, P. A. & ROMESBERG, F. E. 2007. Combating bacteria and drug resistance by inhibiting mechanisms of persistence and adaptation. *Nat Chem Biol,* 3**,** 549-56.

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*. *J Biol Chem,* 287**,** 22004-14.

SNAPPER, S. B., MELTON, R. E., MUSTAFA, S., KIESER, T. & JR, W. R. J. 1990. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol. Microbiol.,* 4**,** 1911-1919.

SRIVASTAVA, S. K., DUBE, D., KUKSHAL, V., JHA, A. K., HAJELA, K. & RAMACHANDRAN, R. 2007. NAD+-dependent DNA ligase (Rv3014c) from *Mycobacterium tuberculosis*: novel structure-function relationship and identification of a specific inhibitor. *Proteins,* 69**,** 97-111.

STANO, N. M., CHEN, J. & MCHENRY, C. S. 2006. A coproofreading Zn2+-dependent exonuclease within a bacterial replicase. *Nat Struct Mol Biol,* 13**,** 458-459.

STELTER, M., GUTSCHE, I., KAPP, U., BAZIN, A., BAJIC, G., GORET, G., JAMIN, M., TIMMINS, J. & TERRADOT, L. 2012. Architecture of a dodecameric bacterial replicative helicase. *Structure,* 20**,** 554-64.

SUN, G., LUO, T., YANG, C., DONG, X., LI, J., ZHU, Y., ZHENG, H., TIAN, W., WANG, S., BARRY, C. E., 3RD, MEI, J. & GAO, Q. 2012. Dynamic population changes in *Mycobacterium tuberculosis* during acquisition and fixation of drug resistance in patients. *J Infect Dis,* 206**,** 1724-33.

SUNDIN, G. W. & WEIGAND, M. R. 2007. The microbiology of mutability. *FEMS Microbiol Lett,* 277**,** 11-20.

SUTTON, M. D., FARROW, M. F., BURTON, B. M. & BURTON, G. C. 2001a. Genetic interactions between the *Escherichia coli umuDC* gene products and the beta processivity of the replicative DNA polymerase. *J. Bacteriol.,* 183**,** 2897-2909.

SUTTON, M. D., MURLI, S., OPPERMAN, T., KLEIN, C. & WALKER, G. C. 2001b. *umuDC-dnaQ* interaction and its implication for cell cycle regulation and SOS mutagenesis in *Escherichia coli*. *J. Bacteriol.,* 183**,** 1085-1089.

SUTTON, M. D., OPPERMAN, T. & WALKER, G. C. 1999. The *Escherichia coli* SOS mutagenesis proteins UmuD and UmuD' interact physically with the replicative DNA polymerase. *Proc Natl Acad Sci USA,* 96**,** 12373-12378.

TANG, M., PHAM, P., SHEN, X., TAYLOR, J. S., O'DONNELL, M., WOODGATE, R. & GOODMAN, M. F. 2000. Roles of *E. coli* DNA polymerases IV and V in lesion targeted and untargeted mutagenesis. *Nature,* 404**,** 1014-1018.

TIMINSKAS, K., BALVOCIUTE, M., TIMINSKAS, A. & VENCLOVAS, C. 2014. Comprehensive analysis of DNA polymerase III alpha subunits and their homologs in bacterial genomes. *Nucleic Acids Res,* 42**,** 1393-413.

TIPPIN, B., PHAM, P. & GOODMAN, M. F. 2004. Error-prone replication for better or worse. *Trends Microbiol.,* 12**,** 288-295.

TOMASZ, M. 1995. Mitomycin C: small, fast and deadly (but very selective). *Chemistry & biology,* 2**,** 575-579.

TOMPKINS, J. D., NELSON, J. L., HAZEL, J. C., LEUGERS, S. L., STUMPF, J. D. & FOSTER, P. L. 2003. Error-prone polymerase, DNA polymerase IV, is responsible for transient hypermutation during adaptive mutation in *Escherichia coli*. *J Bacteriol,* 185**,** 3469-72.

TSAI, H. H., SHU, H. W., YANG, C. C. & CHEN, C. W. 2012. Translesion-synthesis DNA polymerases participate in replication of the telomeres in *Streptomyces*. *Nucleic Acids Res,* 40**,** 1118-30.

TURCIOS, L., CASART, Y., FLOREZ, I., DE WAARD, J. & SALAZAR, L. 2009. Characterization of IS6110 insertions in the *dnaA-dnaN* intergenic region of *Mycobacterium tuberculosis* clinical isolates. *Clin Microbiol Infect,* 15**,** 200-3.

VAN OPIJNEN, T., BODI, K. L. & CAMILLI, A. 2009. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods,* 6**,** 767-72.

VANDEWIELE, D., BORDEN, A., O'GRADY, P. I., WOODGATE, R. & LAWRENCE, C. W. 1998. Efficient translesion replication in the absence of *Escherichia coli* Umu proteins and 3'-5' exonuclease proofreading function. *Proc Natl Acad Sci U S A,* 95**,** 15519-24.

WAGNER, J., FUJII, S., GRUZ, P., NOHMI, T. & FUCHS, R. P. 2000. The beta clamp targets DNA polymerase IV to DNA and strongly increases its processivity. *EMBO Rep,* 1**,** 484-8.

WANG, Y., CUI, T., ZHANG, C., YANG, M., HUANG, Y., LI, W., ZHANG, L., GAO, C., HE, Y., LI, Y., HUANG, F., ZENG, J., HUANG, C., YANG, Q., TIAN, Y., ZHAO, C., CHEN, H., ZHANG, H. & HE, Z. G. 2010. Global protein-protein interaction network in the human pathogen *Mycobacterium tuberculosis* H37Rv. *J Proteome Res,* 9**,** 6665-77.

WANNER, R. M., GUTHLEIN, C., SPRINGER, B., BOTTGER, E. C. & ACKERMANN, M. 2008. Stabilization of the genome of the mismatch repair deficient *Mycobacterium tuberculosis* by context-dependent codon choice. *BMC Genomics,* 9**,** 1471-2164.

WARNER, D. F. 2010. The role of DNA repair in *M. tuberculosis* pathogenesis. *Drug Discovery Today,* 7**,** e5-e7.

WARNER, D. F., EVANS, J. C. & MIZRAHI, V. 2014. *Nucleotide Metabolism and DNA Replication. In Molecular Genetics of Mycobacteria (2nd edition): Hatfull G, Jacobs WR Jr (eds)* Washington D. C, ASM Press.

WARNER, D. F., KOCH, A. & MIZRAHI, V. 2015. Diversity and disease pathogenesis in *Mycobacterium tuberculosis*. *Trends Microbiol.,* 23**,** 14-21.

WARNER, D. F. & MIZRAHI, V. 2006. Tuberculosis chemotherapy: The influence of bacillary stress and damage response pathways on drug efficacy. *Clin. Microbiol. Rev.,* 19**,** 558-570.

WARNER, D. F. & MIZRAHI, V. 2011. Making ends meet in mycobacteria. *Mol Microbiol,* 79**,** 283-7.

WARNER, D. F. & MIZRAHI, V. 2013. Complex genetics of drug resistance in *Mycobacterium tuberculosis*. *Nat Genet.,* 45**,** 1107-8.

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*. *Proc. Natl. Acad. USA,* 107**,** 13093-13098.

WARNER, D. F., TONJUM, T. & MIZRAHI, V. 2013. DNA metabolism in mycobacterial pathogenesis. *Curr Top Microbiol Immunol,* 374**,** 27-51.

WATKINS, H. A. & BAKER, E. N. 2010. Structural and functional characterization of an RNase HI domain from the bifunctional protein Rv2228c from *Mycobacterium tuberculosis*. *J Bacteriol,* 192**,** 2878-86.

WAYNE, L. G. 1977. Synchronized replication of *Mycobacterium tuberculosis*. *Infect Immun,* 17**,** 528-30.

WERNGREN, J. & HOFFNER, S. E. 2003. Drug-susceptible *Mycobacterium tuberculosis* Beijing genotype does not develop mutation-conferred resistance to rifampin at an elevated rate. *J Clin Microbiol,* 41**,** 1520-4.

WHO 2010. Multidrug and extensively drug-resistant TB (M/XDR-TB): GLOBAL REPORT ON SURVEILLANCE AND RESPONSE. Geneva: World Health Organization.

WHO 2012. 'Totally Drug-Resistant' tuberculosis: a WHO consultation on the diagnostic definition and treatment options.

WHO 2014. Global Tuberculosis Report. Geneva, Switzerland: World Health Organization.

WIECZOREK, A. & MCHENRY, C. S. 2006. The NH2-terminal PHP domain of the α subunit of the *Escherichia coli* replicase binds the ϵ proofreading subunit. *J. Biol. Chem.,* 281**,** 12561-12567.

WILLIAMS, A., GUTHLEIN, C., BERESFORD, N., BOTTGER, E. C., SPRINGER, B. & DAVIS, E. O. 2011. UvrD2 Is Essential in *Mycobacterium tuberculosis*, but Its Helicase Activity Is Not Required. *J. Bacteriol.,* 193**,** 4487-4494.

WING, R. A., BAILEY, S. & STEITZ, T. A. 2008. Insights into the replisome from the structure of a ternary complex of the DNA polymerase III alpha-subunit. *J Mol Biol,* 382**,** 859-69.

XIE, Y. & HE, Z. G. 2009. Characterization of physical interaction between replication initiator protein DnaA and replicative helicase from *Mycobacterium tuberculosis* H37Rv. *Biochemistry,* 74**,** 1320-7.

XU, W. C., WRIGHT, G. E., BROWN, N. C., LONG, Z. Y., ZHI, C. X., DVOSKIN, S., GAMBINO, J. J., BARNES, M. H. & BUTLER, M. M. 2011. 7-Alkyl-N(2)-substituted-3-deazaguanines. Synthesis, DNA polymerase III inhibition and antibacterial activity. *Bioorg Med Chem Lett,* 21**,** 4197-202.

YANG, Q., HUANG, F., HU, L. & HE, Z. G. 2012. Physical and functional interactions between 3-methyladenine DNA glycosylase and topoisomerase I in mycobacteria. *Biochemistry,* 77**,** 378-87.

YANG, W. & WOODGATE, R. 2007. What a difference a decade makes: Insights into translesion DNA synthesis. *Proc Natl Acad Sci USA,* 104**,** 15591-15598.

YEISER, B., PEPPER, E. D., GOODMAN, M. F. & FINKEL, S. E. 2002. SOS-induced DNA polymerases enhance long-term survival and evolutionary fitness. *Proc Natl Acad Sci U S A,* 99**,** 8737-41.

YIN, Z., WHITTELL, L. R., WANG, Y., JERGIC, S., LIU, M., HARRY, E. J., DIXON, N. E., BECK, J. L., KELSO, M. J. & OAKLEY, A. J. 2014. Discovery of lead compounds targeting the bacterial sliding clamp using a fragment-based approach. *J Med Chem,* 57**,** 2799-806.

ZHANG, H., LI, D., ZHAO, L., FLEMING, J., LIN, N., WANG, T., LIU, Z., LI, C., GALWEY, N., DENG, J., ZHOU, Y., ZHU, Y., GAO, Y., WANG, S., HUANG, Y., WANG, M., ZHONG, Q., ZHOU, L., CHEN, T., ZHOU, J., YANG, R., ZHU, G., HANG, H., ZHANG, J., LI, F., WAN, K., WANG, J., ZHANG, X. E. & BI, L. 2013. Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from China identifies genes and intergenic regions associated with drug resistance. *Nat Genet,* 45**,** 1255-60.

ZHAO, X. Q., HU, J. F. & YU, J. 2006. Comparative analysis of eubacterial DNA polymerase III alpha subunits. *Genomics Proteomics Bioinformatics,* 4**,** 203-11.

ZHI, C., LONG, Z. Y., GAMBINO, J., XU, W. C., BROWN, N. C., BARNES, M., BUTLER, M., LAMARR, W. & WRIGHT, G. E. 2003. Synthesis of substituted 6-anilinouracils and their inhibition of DNA polymerase IIIC and Gram-positive bacterial growth. *J Med Chem,* 46**,** 2731-9.