Structure/function analyses indicate novel roles for mycobacterial DnaQ homologs in genome maintenance.

Dimitri Louis Jonathan GRIFFAULT

DMTGRI001

Supervisor: Prof. Digby F. Warner

22th of April 2024

A thesis submitted to Department of Medical Microbiology in the Faculty of Health Sciences University of Cape Town, in fulfilment of the requirements for the degree of Doctor of Philosophy

SAMRC/NHLS/UCT Molecular Mycobacteriology Research Unit

Abstract:

Mycobacterial DNA metabolism is of increasing interest as both an underexplored source of new targets for anti-tuberculosis (TB) drug development and for its potential role in the emergence of drug-resistant *Mycobacterium tuberculosis* strains.

However, the redundancy implied by the sizable complement of DNA replication and repair pathways complicates investigations of gene function. There are, moreover, multiple examples in mycobacteria of apparent fusion – or hybrid – proteins in which N- and C-terminal domains appear to provide discrete functions. Both challenges apply to the mycobacterial DnaQ homologs – comprising separate DnaQ and DnaQ-UvrC hybrid proteins – which, by analogy to model organisms such as *E. coli*, have traditionally been assumed to fulfil proofreading roles in DNA replication owing to the presence of conserved exonuclease domains.

Phylogenetic analysis of DnaQ-like proteins revealed a unique domain composition specific to the *Mycobacterium* genus comprising a conserved BRCA1 C Terminus (BRCT) domain in DnaQ. Owing to the presence of the BRCT domain and based on the phenotypes observed in domain-targeted mutants, it appeared that the activity of DnaQ protein (*M. tuberculosis* Rv3711c; *M. smegmatis* MSMEG_6275) might be linked to the mycobacterial gyrases that are responsible for DNA negative supercoiling following replication. The phylogenetic analysis also revealed highly conserved nucleotide excision repair (NER) proteins among bacteria; however, some species like *Actinobacteria*, possess both a canonical UvrC along with a DnaQ-UvrC protein.

The mycobacterial DnaQ-UvrC (*M. tuberculosis* Rv2191; *M. smegmatis* MSMEG_4259) N-terminal was shown to be structurally very similar to that of DnaQ and its C-terminal to that of UvrC, giving the ability to bind either the β-clamp or the components of the UvrABC system. Opposing phenotypes between DnaQ-UvrC (*M. tuberculosis* Rv2191; *M. smegmatis* MSMEG_4259) and *uvrC* deletion imply a DNA-damage specific NER in mycobacteria. This was further confirmed using a combination of gene knockout, site-directed mutants and CRISPRi of NER genes, namely *uvrB* and *uvrC*, in DNA damaging conditions. However, further work is required to elucidate the precise functions of DnaQ and DnaQ-UvrC, and

their contribution to genome dynamics in a family of organisms that includes major human and animal pathogens.

Declaration of Authorship:

I, Dimitri Louis Jonathan GRIFFAULT, hereby declare that the work on which this thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I authorise the University to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate Date: February 12th 2024

Dedication:

For my late grandparents, Louis, Lucienne and Ali

Acknowledgements:

It was not my intention to pursue doctoral studies nor was it to live there for a few years when I first landed in Cape Town back in 2016, I was just a curious, naïve young man eager to discover the world. It's the people I met there who gave me the opportunity and the strength to explore a thrilling yet demanding facet of science.

To my scientific colleagues. First of all, Digby, you are a role model for me and for so many young students that were lucky enough to cross your path, both as a scientist and as a person. Your support, advice and kindness helped me go through these tough years. Thank you for trusting me and making me confident about my work. One of my best memories of South Africa, was in 2018, at the end of my Master 2 internship at the MMRU, when told me that I was ready to pursue doctoral studies. I never thought before that I would be able to do so and it still carries a lot of emotions.

Valerie, thank you for your always valuable inputs and advice. Your successes as a world-renowned scientist made me proud of being a part of the MMRU.

Saber, thank you for the support during my Master internships. You taught me all the skills I needed for this PhD, and you did so regardless of the time or day it was. Most of all, thank you for your friendship and the evenings at Beerhouse.

To Česlovas Venclovas, thank you for your help and advice regarding protein clustering of the DnaQ mycobacterial homologs.

Vinayak and Pooja, thank you for your friendship, for taking interest in my work and for the restaurants that we shared.

To Laurianne and Nina, thank you for your friendship, for the countless beers and runs that we had together with the ORCs.

The members of the MMRU, especially Mandy, Phia, Joseph, Gabriel and Terry thank you for your kind welcome at the MMRU, for your advices and contribution to my work. It was always a pleasure to share coffee and cakes with you.

Nomfundo, Avuyonke and Rudranil from CIDRI, Darien and Alexandre from Professor Blackburn team. Thank you for the always nice conversations and coffee breaks.

Lameeza, Ronnett, Kathryn and Rene, thank you for your kind welcome, your assistance to find my way through the mazes of administration and paper work I had to face throughout this PhD.

I also want to thank the Wellcome trust and CIDRI Africa for funding my work through the WUN CIDRI-PhD scholarship. It was the first major success of my young career, and I am grateful for the opportunity to have worked in a world-class institution recognized for its expertise in infectious diseases.

To my friends. Thank you Jerôme, Dina, Victor, Gaëtan, Pierre-Gilles and Alexandre for your friendship through the years and the distance. Thank you for your support, the good times and so much more.

To Audrey, Pierre, Florent and Corentin, we went through the same journey but on different countries. Knowing that you were there understanding what I was going through made my life easier during this PhD. Thank you for the conversations and advice. I wish you all of the best for our future career.

To all the other that have enlightened my too few, short stays in France, thank you Benjamin, Valentin, Mathias, Thibault, Fabien, and so much more. Thank you for the friendship and for all we have the adventures we have shared for so many years.

To Kenza, a special thank for my Master tutor without whom I would never have the idea nor the courage to go to South Africa in the first place.

To my family. My parents, Danielle and Didier. Thank you for everything. For your love, your support, for allowing to develop my fondness for learning and knowledge. Thank you for allowing me (and supporting) to pursue the studies and career I chose. Thank you for pushing me to give my best for this project and in life in general. This thesis is also your accomplishment.

To my brother and sisters, Baptiste, Marion and Mélaine, thank you for your constant love and support in the most difficult moments.

“The Road goes ever on and on

Down from the door where it began.

Now far ahead the Road has gone,

And I must follow, if I can,

Pursuing it with eager feet,

Until it joins some larger way

Where many paths and errands meet.

And whither then? I cannot say”

— J.R.R. Tolkien

Table of Contents:

List of Abbreviations:

4-Nitroquinoline1-oxide	4-NQO
Acetylaminofluorene-guanine	AAF-G
Acquired Immunodeficiency Syndrome	AIDS
Adenosine triphosphate	ATP
Arabinogalactan	AG
Base excision repair	BER
CLuster ANalysis of Sequences	CLANS
Clustered regularly interspaced short palindromic repeats interference	CRISPRi
Colony forming unit	CFU
Cryo-electron microscopy	cryoEM
Deoxyribonucleic acid	DNA
Deoxyribonucleotide triphosphate	dNTP
Double-stranded break	DSB
Double-stranded DNA	dsDNA
Ecole Polytechnique Fédérale de Lausanne	EPFL
Enhanced green fluorescent protein	eGFP
Extensively drug-resistant TB	XDR-TB
Global genomic repair	GGR

Griselimycin	Grs
Half maximal inhibitory concentration	IC_{50}
Helix-hairpin-helix	HhH
High-scoring Segment Pair	HSP
Homologous recombination	HR
Human Immunodeficiency Virus	HIV
Knock-out	KO
Lipoarabinomannan	LAM
Lipomannan	LM
Minimal inhibitory concentration	MIC
Mismatch repair	MMR
Mitomycin C	MMC
Molecular Mycobacteriology Research Unit	MMRU
Multidrug-resistant TB	MDR-TB
Mutation frequency decrease factor	Mfd
Mycobacterium smegmatis	*Msm*
Mycobacterium tuberculosis	*Mtb*
Mycolic acid	MA
New England Biolabs	NEB
Non-homologous end joining	NHEJ

Nucleotide excision repair	NER
Oleate-Albumin-Dextrose-Catalase	OADC
Oligonucleotide binding	OB
Optical density	OD
Polymerase Chain Reaction	PCR
People living with HIV	plHIV
Peptidoglycan	PG
Phosphatidylinositol mannosides	PIMs
Phthiocerol dimycocerosate	PDIM
Polymerase and histidinol phosphatase	PHP
Protospacer adjacent motif	PAM
Pyrimidine dimer	Pyr<>Pyr
Reactive nitrogen intermediates	RNI
Reactive oxygen intermediates	ROI
Ribonucleic acid	RNA
Rifampicin	Rif
RNA polymerase	RNAP
Single-guide RNA	sgRNA
Single-strand annealing	SSA
Tetracycline	Tet

Totally drug-resistant TB	TDR-TB
Transcription-coupled repair	TCR
Translesion synthesis	TLS
Tuberculosis	TB
Ultra-violet	UV
Wild type	Wt
World Health Organisation	WHO

List of Tables:

List of Figures:

CHAPTER I:

General Introduction:

Adapted from Active Tuberculosis (FL): StatPearls Publishing; 2023 Jan-. [Figure, Scanning electron micrograph, SEM, Rod-shaped...] Available from: https://www.ncbi.nlm.nih.gov/books/NBK513246/figure/article-17136.image.f4/

1-1- *Mycobacterium tuberculosis*:

Mycobacterium tuberculosis (*Mtb*), also known as Koch's bacillus, is an obligate pathogenic bacterium and the causative agent of tuberculosis (TB). TB is a major global health problem as one of the leading causes of death along with the Human Immunodeficiency Virus (HIV). Since 2005, the estimated number of TB deaths has diminished each year owing to the global effort to better detect and treat the millions of *Mtb*-infected people. In 2020, the number of new TB cases per year dropped below 6 million, which was an historical low[1] after a large increase between 2017 and 2019[2] in all WHO regions. The task nevertheless remains long and challenging for TB to be eradicated as a public health threat globally.[3]

Unfortunately, the recent COVID-19 pandemic seems to have reversed this trend.[3] Even though the effect of SARS-CoV-2 on *Mtb* infection is not well understood,[4] it is now established that the profound social and economic impact of COVID-19 led to reductions in the number of both reported deaths and newly diagnosed cases.[5] This is expected to result in an increase in transmission which will lead to an increased number of deaths in the coming years, especially if those people remained undiagnosed and, therefore-, untreated.[1]

Indeed, in 2021 there were an estimated 1.4 million deaths among HIV-negative people and an additional 187 000 deaths among people living with HIV (plHIV) which is a 6.67% increase, a similar level to 2017. In contrast, deaths from HIV/AIDS continued to decline, even during the COVID-19 pandemic. Now, TB mortality is more than twice that attributed to HIV/AIDS. The severe impact of COVID-19 calls for a renewed effort to fight TB globally.[3]

Although rightly considered a global threat, the TB burden is much greater in low-income regions like in Southern Africa and South-Eastern Asia, especially among the poorest populations. This is unacceptable considering that, when diagnosed and treated, TB is curable. So, while it is estimated that about one-quarter of the world's population has latent TB,[3] only 30 countries concentrate about 87% of TB global cases whereas Europe and America account for only about 6%. Despite the availability of effective drug treatment since the 1940s, TB remains a major public health problem in low-income regions of the world.

Mtb has been a human pathogen for the last 70 000 years and is thought to have co-evolved with humans following early out-of-Africa migration.[6,7,8] This assumption, however, was challenged by Sabin et al. Using ancient genomes of preserved calcified nodules as calibration points, they estimated that organism of the Mycobacterium Tuberculosis Complex (MTBC) derived from a common ancestor about 6000 years ago.[184] Evidence suggests that *Mtb* gained virulence and persistence mechanisms through a series of genetic events.[9–12] This, along with the increase in the human population size and density owing to the development of agriculture and more broadly, industrial civilization, drove the evolution of virulence and transmissibility traits that characterize today's *Mtb* strains.[13] As humans domesticated animals, they were able to transmit diseases to them too, enabling the emergence of animal pathogens such as *M. bovis*.

1-2- *Mycobacterium tuberculosis* infection:

Mtb is transmitted through aerosol droplets released when an infected person coughs or simply breathes.[14] Following inhalation, the bacilli access the lungs (Figure I) where they are taken up by resident alveolar macrophages. *Mtb* is able to prevent phagosome and lysosome fusion within the macrophages, thus surviving phagocytosis and allowing for bacterial replication. This quickly leads to a high bacterial burden that can disseminate to other parts of the organism. To prevent this, macrophages release various cytokines to induce the recruitment of more macrophages and granulocytes at the infection site, establishing the core of the granuloma. Over time, cytokines also draw T-cells that encircle and confine the infected macrophages to prevent bacterial dissemination. At this point, the host immune response to TB is to form granulomatous focal lesions with a caseous centre that is calcified on the outside to contain the infection.[15–17]

The conditions in the granuloma are harsh owing to the acidic pH, the presence of reactive oxygen (ROI) and nitrogen (RNI) intermediates, and the reduced availability of oxygen. It was long thought that granulomas provided protection to the host by trapping the mycobacteria under conditions that favoured their elimination. Yet, contradictory evidence suggests that *Mtb* may manipulate the granuloma to protect itself from host immune cells while allowing for dissemination.[18,19] Latent infection with *Mtb* may last

decades or even a lifetime. The strength of the host immune response will determine whether the infection is arrested or is allowed to progress. For this reason, plHIV are more susceptible to *Mtb* infection.

Granulomas can undergo necrosis, resulting in the formation of caseum, which may allow bacilli to replicate while also minimizing access of anti-TB drugs. Then, the caseum can undergo liquefaction leading to bacterial spread [19,20] and, ultimately, to active TB (Figure I). However, this assumption was challenged, and it is now thought that reactivation of the diseases does not originate from the primary site of infection but rather from an unaffected part of the host where *Mtb* bacilli seem to be able to survive in tissues not associated with granulomas.[21,22] This mechanism remains poorly understood and has to be further investigated. Because of the unusual composition and structure of the mycobacterial cell wall, effective TB treatment is difficult. It usually takes 6 months of antibiotic treatment to eliminate the disease.

1-3- Factors contributing to the emergence of drug resistance:

1-3-1- Intrinsic factors:

Drug-associated resistance acquisition is associated with the concept of mutation;, however, some intrinsic feature may also contribute to the poor efficacy of treatments. *Mtb* for example, with its thick cell wall replete with hydrophobic mycolic acids, prevents or slows down diffusion of antibiotics.[23, 24, 188, 189] Moreover, a complex interaction between antibiotic action and cell envelope maintenance was suggested that emphasizes the need to understand the genetic basis of antibiotic resistance in *Mtb*.[185,186] Thanks to new visualisation techniques and technology such as electron microscopy, cryo-electron tomography it was shown that mycobacteria cell wall is composed of two asymmetric inner and outer layers surrounding a membrane. The outer layer is made of long and short lipids, such as phosphatidylinositol mannosides (PIMs), lipomannan (LM), and lipoarabinomannan (LAM), which are freely interspersed with cell-wall proteins. These proteins are the soluble part of the cell wall and are considered to be the signalling and effector molecules of mycobacteria as they are known to interact with the host immune system. The inner layer is composed of peptidoglycan (PG), arabinogalactan (AG), and

mycolic acids (MA) covalently bonded together to form the insoluble MA-AG-PG complex (Figure II).[27]

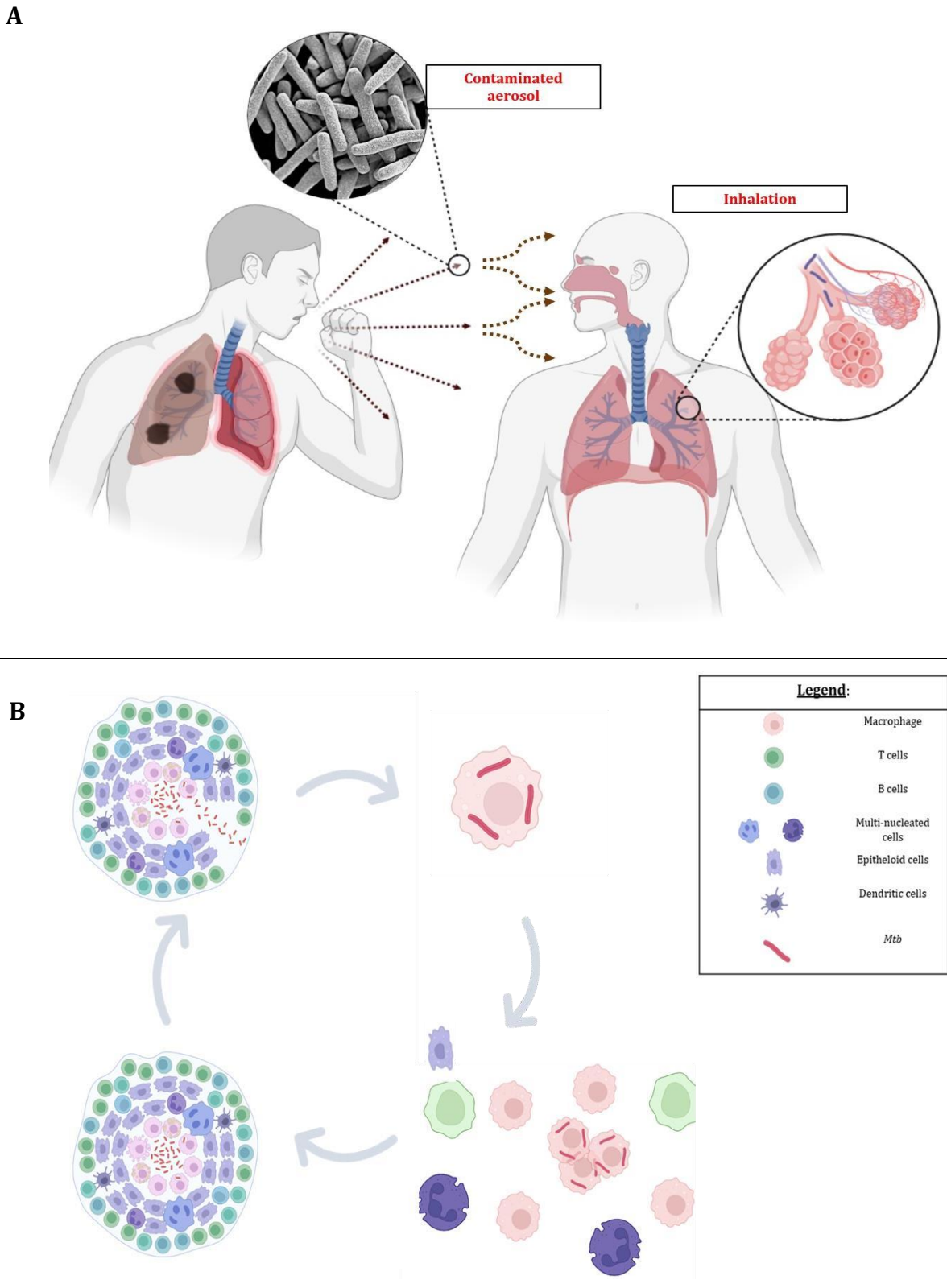


Figure I- *Mtb* infection and progression of granuloma formation. A-Transmission through aerosol. B-Infection, formation and collapse of granulomas. Adapted from Marakalala et al.[20] Created with BioRender.

There are other intrinsic resistance mechanisms that are common to a great number of bacteria, such as efflux pump which are membrane proteins able to expel drugs from the bacilli, among nutrients, waste, and other molecules. Another way for bacteria to escape drugs is to structurally modify them to reduce affinity for the target or to inactivate them. They can also produce lure that imitates the target but that have no role in metabolism. Besides these array of tools bacteria naturally possesses to overcome antibiotic treatment; they can also evolve to adapt to threat they never met before. Because of all these features, *Mtb* is naturally resistant to most antibiotics which severely restricts the treatments available.

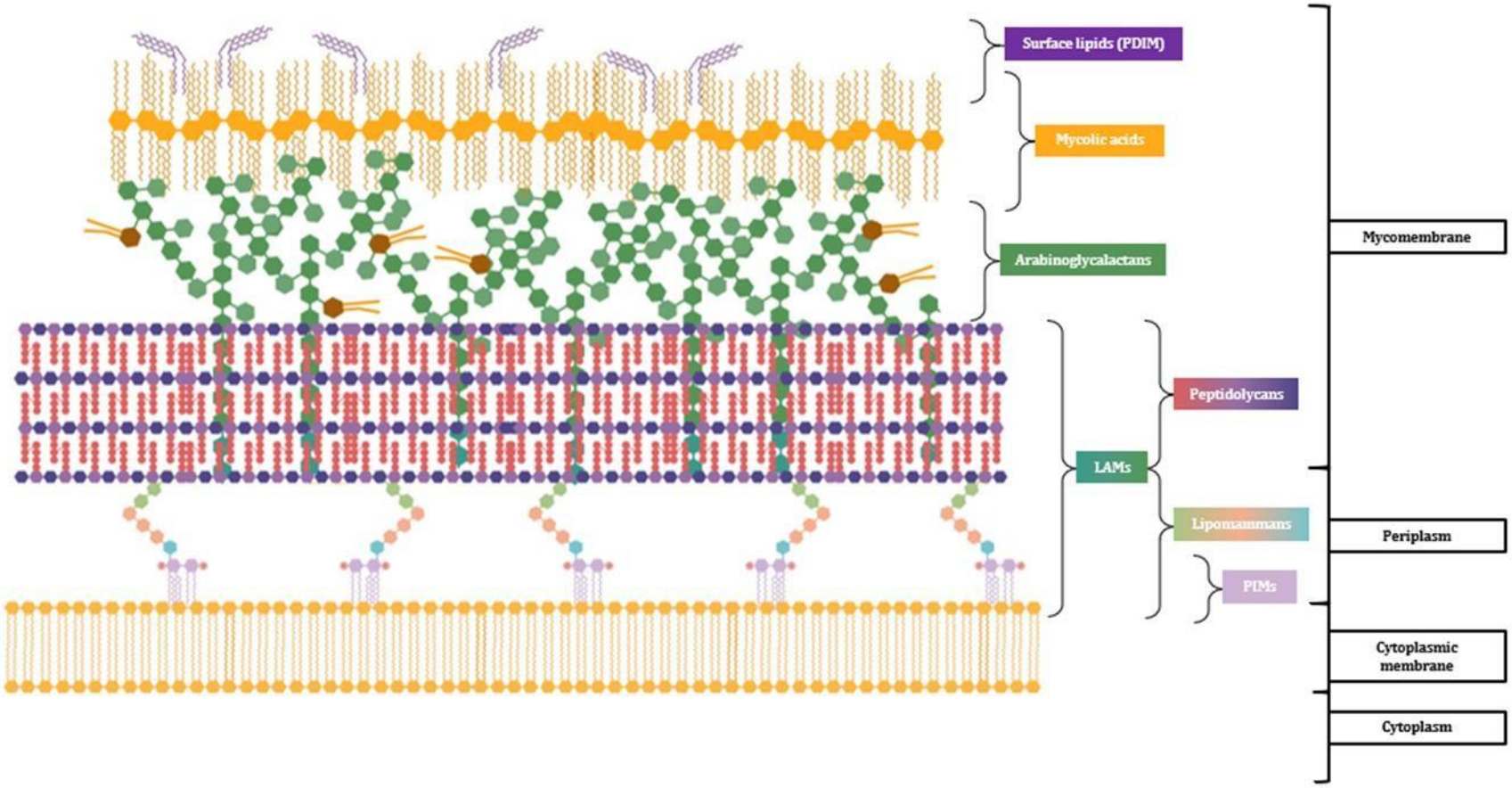


Figure II- Schematic representation of the mycobacterial cell wall: Mycolic acids with surface lipids form the outer layer bound to the inner layer via lipoarabinomannan. The combination of peptidoglycans and arabinogalactan forms the inne r layer. Created with BioRender.

1-3-2- External factors:

Furthermore, just like any living organism, *Mtb* can evolve and adapt to its environment by gaining new traits. These new traits will be passed from generation to generation, provided they offer a survival advantage without incurring too great a fitness cost. For *Mtb* and other pathogenic bacteria, that often means acquisition of resistance to antibiotics to which they were previously susceptible.[27] Concerningly, drug-resistant mutants can regain fitness via compensatory mutations. By doing so, they are increasing their transmissibility and virulence while lowering the associated fitness cost which will lead inevitably to the emergence of a totally drug resistant (TDR) phenotype.

In *Mtb*, the acquired resistance occurs through chromosomal mutations as no evidence exists for horizontal transfer of resistance genes.[27] The prolonged exposure to the drugs because of the length of the TB regimen contributes to the evolution of resistant mutants (Table I). For example, the essentiality of the mycobacterial cell wall makes it an attractive target for existing and new TB drugs; these include isoniazid and ethambutol, two of the frontline anti-TB agents. Rifampicin is another mainstay anti-TB drug, inhibiting RNA synthesis by the DNA-dependent RNA polymerase.[13,25,26] These drugs were developed almost 60 years ago and there is an urgent need to develop new ones, so that resistance does not become predominant in *Mtb* populations. Exposure to antibiotics is also altered when patients forget or refuse to take their treatments or if they take under-dosed counterfeit and sub-optimal drugs.[28]

Mtb strains resistant to the first-line drugs (Table I) are defined as Multi Drug-Resistant TB (MDR-TB), while those with additional resistance to a fluoroquinolone and any of the second-line drugs are known as Extensively Drug-Resistant TB (XDR-TB). Strains resistant to all first and second-line drugs are named Totally Drug-Resistant TB (TDR-TB) strains, though this form of resistance is not defined yet by WHO. The drug-susceptible TB treatment is 6 months long, whereas MDR-TB and XDR-TB necessitate a much longer course of therapy with a combination of second and even third-line medications. The resulting increased costs, as well as the severe side effects for the patient, are the main drivers of patient non-adherence and/or the use of counterfeit or suboptimal drugs which in turns favours the emergence of drug-resistant *Mtb* strains. In an attempt to end this vicious circle, an effort is made toward the development of new anti-TB treatment and to

identify new potential targets. The enzymes within mycobacterial replication machinery and DNA repair mechanism represent two of those targets.

1-4- Drugs targeting mycobacterial DNA replication and repair components:

Among the clinically used anti-mycobacterial drugs, only the fluoroquinolones target the *Mtb* replication machinery, specifically by inhibiting the type II topoisomerase comprising DNA gyrase A (GyrA) and B (GyrB) subunits. These include moxifloxacin and levofloxacin, while novobiocin – a coumarin class antibiotic – targets GyrB (Table I). Nonetheless, DNA gyrase represents a promising target as a number of exploitable ligand binding pockets exists in the tetrameric complex.[44] For now, the fluroquinolones are used as second-line anti-TB treatment to shorten the treatment of drug-susceptible *Mtb.*[45]

Quite recently, two potential antimicrobial treatments that target other DNA replication machinery were identified as potential anti-TB treatments. Produced by *Streptomyces* sp., the natural product griselimycin (Grs) was shown to be highly active against *Mtb* by inhibiting the *dnaN*-encoded β-sliding clamp of the mycobacterial replisome.[46,47] Nargenicin A1, which was first isolated from *Nocardia* species, is another bactericidal genotoxin[48] that was first shown to be effective against methicillin-resistant gram-positive bacteria such as *S. aureus.*[49] Recent work has shown that Nargenicin inhibits the essential replicative DNA polymerase DnaE1 in *Mtb* by preventing the incorporation of the incoming nucleotide during DNA replication and, therefore, should be considered as a candidate for future TB drug-discovery efforts.[50]

1-5- DNA replication and repair in *Mtb*:

The development of molecular and genetic tools might enable the identification of new targets that could possibly be used to fight against *Mtb* strains that are resistant to the classic anti-TB drugs. The aim of this work was to identify targets towards developing screening assays for new replication or mutagenesis inhibitors. Before getting to that stage, some questions had to be investigated first. Here, the work focused on characterising the exact role of the mycobacterial DnaQ homologues, DnaQ (Rv3711c;

Msm MSMEG_6275) and DnaQ-UvrC (Rv2191; *Msm* MSMEG_4259) in genome maintenance, including under conditions of genotoxic stress.

	Drug	Target	Associated gene	Gene function
First line	Isoniazid	Mycolic acid biosynthesis	*katG*	Catalase-peroxidase
			inhA	Enoyl ACP reductase
			ndh	NADH dehydrogenase II
			ahpC	Alkyl hydroperoxidase
	Rifampicin	Transcription	*rpoB*	β-subunit of RNA polymerase
	Pyrazinamide	Trans-translation	*pncA*	Pyrazinamidase
			rpsA	S1 ribosomal protein
	Ethambutol	Arabinogalactan synthesis	*embCAB*	Arabinosyltransferase
			embR	*embCAB* transcription regulator
	Streptomycin	Translation	*rpsL*	S12 ribosomal protein
			rrs	16S rRNA
			gidB	16S rRNA methyltransferase
Second line	Kanamycin	Translation	*rrs*	16S rRNA
			eis	Aceryltransferase
	Ethionamide	Mycolic acid biosynthesis	*ethA*	Flavin monooxygenase
			inhA	Enoyl ACP reductase
			ethR	*ethA* transcription repressor
			ndh	NADH dehydrogenase II
			mshA	Glycosyltransferase
	Fluoroquinolones	DNA gyrase	*gyrA*	DNA gyrase subunit A
			gyrB	DNA gyrase subunit B

Table I- List of first- and second-line drugs used for TB treatment: Targets and associated genes and function are specified.[27,28]

Most of the knowledge on bacterial replication comes from the *E. coli* model. Replication depends on a complex of three proteins, the DnaE1 high fidelity polymerase that adds the nucleotide to make the newly synthesized strand, the DnaQ proofreading exonuclease that removes any misincoporated nucleotide to ensure replication fidelity, and the DnaN β sliding-clamp that holds everything together (Figure III). Though informative, growing evidence supports the need to consider the mycobacterial replication machinery as a distinct system. Indeed, mycobacteria exhibit a basic replication model that lacks several components of the replisome found in *E. coli* or that have apparently different functions.[29,30] More details are provided in Chapter IV. For example, in the *E. coli* model, DnaQ functions as the proofreading exonuclease, responsible for removing misincorporated nucleotides. Deletion of *dnaQ* in *E. coli* causes a mutator phenotype, characterized by significantly elevated mutation rates. In contrast, previous work in *Msm* and *Mtb* indicated that loss of *dnaQ* did not affect the mutation rate *in vivo*[30–32]; instead, mycobacteria appear to rely on the PHP domain of DnaE1α polymerase for proofreading function.[31] This observation was surprising, especially given biochemical evidence that the mycobacterial DnaQ is able to interact with the replisome, and affect the replication processivity, by binding to the β-clamp.

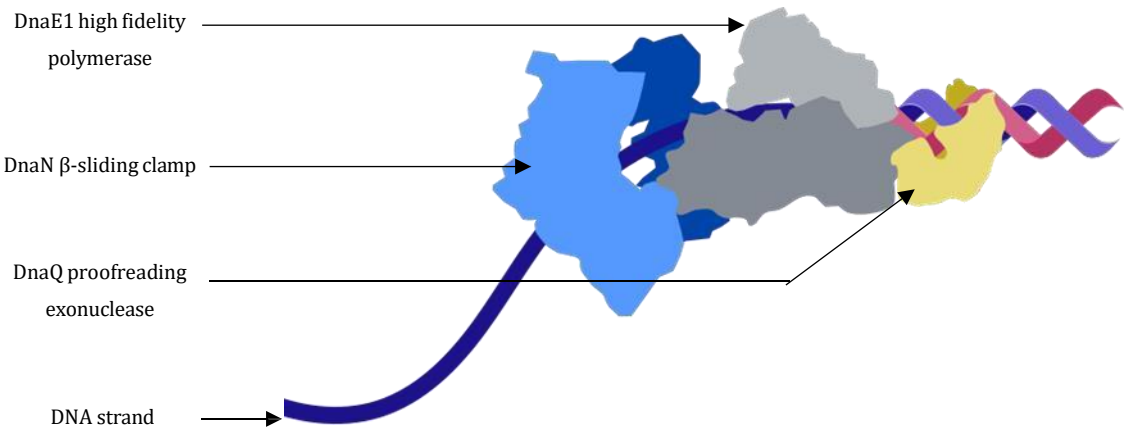
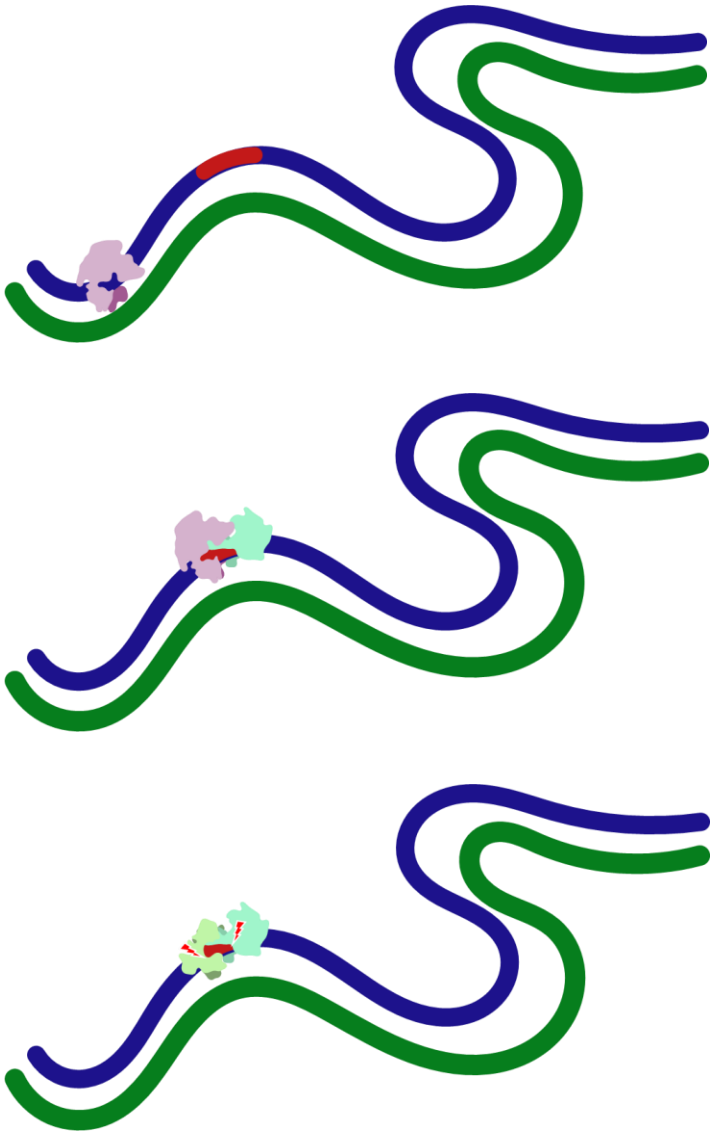


Figure III- Core of the bacterial replication model

The absence of a mutator phenotype in any of the mycobacterial *dnaQ* deletion mutants[30,31] highlights the need for more research about the factors which determine relative contribution of both PHP domain and DnaQ-mediated exonuclease function in the mycobacterial proofreading activity. This is reinforced by the fact that the *dnaQ* gene is highly polymorphic and has been identified as a target of independent mutation associated with drug resistance in *Mtb* and is under strong selective pressure.[34,35] As for many DNA repair and replication components, studies of DnaQ, so far, have failed to elucidate discernible phenotypes. This might be due to the presence of multiple copies of the studied gene and/or presence of fusions of other homologs that have functional redundancy. Moreover, the inadequacy of the microbiological assays employed has to be considered. Thus, additional research to resolve Δ*dnaQ* phenotypes would contribute to our understanding of this gene function.

Another feature of mycobacteria is the presence of a putative DnaQ-UvrC (Rv2191; *Msm* MSMEG_4259) hybrid protein. This was first observed by Kesavan *et al,*[36] who suggested a potential role in a DNA damage repair mechanism in *Mtb* bacilli contained in granulomas during latent or reactivated TB in rabbits. This mechanism is called Nucleotide Excision Repair (NER) and is one of a few repair systems found in most bacteria. While NER results in incision of the damaged DNA flanking regions, Base Excision Repair (BER) results in damaged single-base removal. NER can also recognise a broader range of DNA damage due to low specificity exonuclease. Another kind of repair mechanism requires dsDNA breaks to be active. Those recombination repair mechanisms are Homologous Recombination (HR) and Non-Homologous End-Joining (NHEJ).[122,193] In the vast majority of bacterial organisms NER depends mainly on three proteins named UvrA, UvrB and UvrC. UvrA is responsible for the damage detection while UvrC is responsible for the damaged DNA incision within the DNA strand. On the other hand, UvrB is a hub protein that allows the transition between damage detection and incision. It really is the key factor of NER as UvrC is not able directly bind UvrA or damages DNA by itself (Figure IV).



Step 1: UvrA scans DNA for damage detection

Step 2: UvrB binds UvrA/damaged DNA to for a pre-incision complex

Step 3: UvrC replace UvrA in the complex to perform the 3'-5' incision around damaged DNA.

Figure IV- Steps of the bacterial NER model

DnaQ-UvrC (Rv2191; *Msm* MSMEG_4259) is a natural hybrid protein that possesses an N-terminal that is similar to DnaQ (MSMEG_6275) and a C-terminal similar to UvrC (MSMEG_3078). The very existence of this protein is intriguing given the fact that mycobacteria encode separate canonical DnaQ and UvrC proteins. Although UvrC and DnaQ-UvrC are predicted to have a role in nucleotide excision repair (NER), they are not triggered by the same pathway: [37,38] Expression of *dnaQ-uvrC* is triggered by the PafBC-dependent alternate DNA damage response, whereas *uvrC* appears constitutively expressed.[39] Interestingly, PafBC-dependent alternate DNA damage response do not exist in *E. coli*. Moreover, in this organism, uvrC expression is triggered by the classical

RecA/LexA SOS response. More detail are provided in Chapter V. Owing to its sequence homology with DnaQ, the DnaQ-UvrC protein was briefly studied by Ford *et al*[40] but for its potential role in replication fidelity. Ditse and colleagues later reported mitomycin C (MMC)-induced DNA damage sensitivity in an *Msm dnaQ-uvrC* deletion mutant.[41] That intriguing observation, combined with the sequence homology, suggested the possibility that DnaQ-UvrC might contain dual proofreading and DNA repair functions, two key aspects of genome maintenance. Consequently, this work focused on the potential role of DnaQ-UvrC role under different conditions of genotoxic stress.

As mentioned earlier, mycobacteria exhibit numerous features that separate them from other bacteria. Because of this specificity, *E. coli* cannot be used as a model organism. Thus, a specific mycobacterial model is needed. Although it has limitations when studying pathogenicity,[42] the fast growing, non-pathogenic mycobacterium, *Msm*,[43] encodes thousands of conserved mycobacterial gene orthologs and has the same cell architecture and physiology.[190] For this reason, it was adopted in this work owing to its utility in previous studies elucidating fundamental aspects of mycobacterial DNA metabolism.

Thus, the hypothesises we tested were the following:

-DnaQ has a role in mycobacteria outside of replication fidelity and is involved in the rising of drug resistance.

- DnaQ-UvrC is part of canonical NER either as replacement of UvrC or operating in an alternate mechanism to NER in mycobacteria.

1-6- Aims and Objectives:

1-6-1- Bioinformatic investigation of *Msm* DnaQ-UvrC (MSMEG 4259) and DnaQ (MSMEG 6275) proteins to infer potential function(s) and domains:

-Identify conserved protein domain composition.

-Identify putative protein homologs in other organisms.

-Determine whether these proteins are similar in *Msm* and *Mtb*.

1-6-2- Construction of *Msm dnaQ-uvrC* (MSMEG 4259) and *dnaQ* (MSMEG 6275) gene deletion and site-directed mutants for phenotyping and structure-function analyses:

-Inactivate full-length gene by gene deletion, and target protein domains for site-directed mutagenesis to determine relative contribution to any observed phenotypes.

-Identify phenotypes or any other potential sensitivity to antibiotic treatments under standard laboratory conditions.

1-6-3- Investigate the contribution of mycobacterial UvrC to NER relative to DnaQ-UvrC:

-Generate unmarked *Msm uvrC* (MSMEG_3075) knock-out and identify phenotypes in DNA-damaging conditions.

-Confirm potential results with inducible CRISPR-interference (iCRISPRi).

1-6-4- Investigate any potential role of DnaQ in evolution of antibiotic resistance or mutagenesis in *Msm* during standard growth and following DNA damage:

-Comparison between wild type and *Msm* mutant strains in UV-induced mutation frequency assays.

-Adaptation of antibiotic evolution assay.

CHAPTER II:

Materials and Methods:

2-1- Bacterial strains and culturing conditions:

Mycobacterium smegmatis (*Msm*) and *Escherichia coli* strains (Table II) were used in this study. The plasmids and oligonucleotides used in this study are detailed below (Table V and VI). Information about strains and plasmids is provided in relevant chapters.

Strain	Description	Source
***Escherichia coli* DH5α**	*supE44 ΔlacU169* (*F80 lacZΔM15*) *hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	Lab stocks
***Mycobacterium smegmatis* mc²155**	High frequency transformation mutant of *Msm* ATCC 706	Snapper *et al.,* 1990[43]

Table II- General bacterial strains used in this study.

E. coli strains used for cloning experiments were grown in Luria Bertani (LB) broth supplemented with the appropriate antibiotic, shaking overnight at 37°C in an IncoShake incubator (Labotec). For growth on solid medium, cells were streaked on LB agar plates and incubated at 37°C, or 30°C in the case of knockout constructs. *M. smegmatis* strains (Table III) were grown in Middlebrook 7H9 broth 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) with the appropriate antibiotics described in Table IV. The antibiotic working concentrations are provided in relevant chapters; however, for antibiotic selection, they were used at the following standard concentrations for *Msm*: 50 µg/ml Hygromycin (Hyg) and 20 µg/ml Kanamycin (Kan). For *E. coli*: 200 µg/ml Hygromycin (Hyg), 100 µg/mL Ampicillin (Amp) and 50 µg/ml Kanamycin (Kan).

Strain	Description	Source
***Msm* mc²155**	Wild-type	Snapper *et al.,* 1990[43]
Δ*dnaE2*	*dnaE2* knockout mutant of mc²155	Saber Anoosheh, MMRU
Δ*dnaQ* (MSMEG_6725)	*dnaQ* knockout mutant of mc²155	Saber Anoosheh, MMRU
Δ*dnaQ-uvrC* (MSMEG_4259)	*dnaQ-uvrC* knockout mutant of mc²155	Saber Anoosheh, MMRU
Δ*dnaQ* + Δ*dnaQ-uvrC*	*dnaQ* and *dnaQ-uvrC* double knockout mutant of mc²155	Saber Anoosheh, MMRU
Δ*dnaQ*::*eGFP-dnaQ*	*dnaQ* knockout mutant complemented with *dnaQ* fused to eGFP	This study
Δ*dnaQ-uvrC*::*eGFP-dnaQ-uvrC*	*dnaQ-uvrC* knockout mutant complemented with *dnaQ-uvrC* fused to eGFP	This study
Δ*dnaQ-uvrC*::*dnaQ-uvrC* Δβ-clamp BD	*dnaQ-uvrC* knockout mutant complemented with *dnaQ-uvrC* with defective β-clamp Binding Motif	This study
Δ*dnaQ-uvrC*::dnaQ-uvrC Δuvr BD	*dnaQ-uvrC* knockout mutant complemented with *dnaQ-uvrC* with defective UVR Binding Motif	This study
Δ*dnaQ-uvrC*::*dnaQ-uvrC* Δβ-clamp BD/Δuvr BD	*dnaQ-uvrC* knockout mutant complemented with *dnaQ-uvrC* with defective β-clamp and UVR Binding Motif	This study
Δ*uvrC*	*uvrC* knockout mutant of mc²155	This study
Δ*uvrB*	*uvrB* knockout mutant of mc²155	Irene Gobe, MMRU
DnaN-mCherry	mc²155with *dnaN* fused to mCherry	Michael Reiche, MMRU
DnaN-mCherry + eGFP-DnaQ	mc²155with *dnaN* fused to mCherry with *dnaQ* fused to eGFP	This study
DnaN-mCherry + eGFP-DnaQ-UvrC	mc²155with *dnaN* fused to mCherry with *dnaQ-uvrC* fused to eGFP	This study
tet-dCas9::*uvrC*	mc²155with CRISPRi gene silencing plasmid targeting *uvrC*	This study
tet-dCas9::*uvrB*	mc²155with CRISPRi gene silencing plasmid targeting *uvrB*	This study
Δ*recA*	*recA* knockout mutant of mc²155	Krupa Naraan, MMRU
LexA^{Ind-}	Non-cleavable LexA mutant of mc²155	Krupa Naraan, MMRU

Table III- Mutant *Msm* strains used in this study- BD= Binding domain

Drug	*Description*	*Source*
Mitomycin C	Potent DNA crosslinker	[51]
Moxifloxacin	Inhibits the unwind activity reaction catalysed by GyrA	[52]
Ofloxacin		
Novobiocin	Inhibits the ATPase reaction catalysed by GyrB	
Griselimycin	Blocks binding of other proteins to the DNA polymerase β-sliding clamp subunit, DnaN	[46]
Nargenicin	Inhibits the DNA polymerase IIIα subunit, DnaE1	[53]
4-Nitroquinoline-1-oxide	Induces lethal DNA lesions	[54]
Nitrofurazone		
Kanamycin	Interference with 16S rRNA	[55]
Hygromycin	Inhibits 80S ribosome and blocks protein synthesis	[56]
Ampicillin	Interference in cell wall bioynthesis	[57]

Table IV- Antibiotics used in this study.

Vector	***Details***	***Source***	***Features***
pMCpAINT	Mycobacterial integrative cloning vector	Lab Stock	KanR
pMCpAINT-eGFP-dnaQ	pMCpAINT vector containing *dnaQ* fused to eGFP	This Study	KanR
pMCpAINT-egfp-dnaQ-uvrC	pMCpAINT vector containing *dnaQ-uvrC* fused to eGFP	This Study	KanR
pMV306	Mycobacterial integrative cloning vector	Lab Stock	HygR
pMV306-β^{clamp}	pMV306 vector containing with *dnaQ-uvrC* with defective β-clamp Binding Motif	This Study	HygR
pMV306-UVR	pMV306 vector containing with *dnaQ-uvrC* with defective UVR Binding Motif	This Study	HygR
pMV306-Interaction sites	pMV306 vector containing with *dnaQ-uvrC* with defective β-clamp and UVR Binding Motif	This Study	HygR
pGOAL19	Selection cassette vector	[58]	HygR, Lac Z, sacB
p2NIL	Gene manipulation vector	[58]	KanR
pUVRC	p2NIL:Δ*uvrC* delivery vector containing pGOAL17 suicide selection cassette	This Study	KanR, Lac Z, sacB
pDNAQ	p2NIL:Δ*dnaQ* delivery vector containing pGOAL17 suicide selection cassette	[40]	KanR, Lac Z, sacB
pDNAQC	p2NIL:Δ*dnaQ-uvrC* delivery vector containing pGOAL17 suicide selection cassette	[41]	KanR, Lac Z, sacB
pCRISPRi	$_{p}$CRISPRi containing inactivated Cas9 allele from *Streptococcus thermophilus* without sgRNA	[59]	Kan
pCRISPR-uvrB	$_{p}$CRISPRi containing inactivated Cas9 allele from *Streptococcus thermophilus* with *uvrB* targeting sgRNA	This Study	Kan
pCRISPR-uvrC	$_{p}$CRISPRi containing inactivated Cas9 allele from *Streptococcus thermophilus* with *uvrC* targeting sgRNA	This Study	Kan

Table V- Plasmids used in this study.

ID	Sequence	Details	PCR Size (bp)	Restriction Site	PCR Pair	ID	Sequence	Details	PCR Size (base pair)	Restriction Site	PCR Pair
SA101	ATGGTACCG CCTCAAGTTT TCGTCCTC	Forward 1st segment *dnaQ-uvrC*	747	Kpn I	-	SA152	GTT GTT CGC ACC CTT ATA CGT CGG CGT GGT C	*dnaQ-uvrC* Upstream Reverse Exo III	1242	-	SA101
SA102	CCAGCTCCAT GTTCTGCACT TGCTCCGCGC C	Reverse 1st segment *dnaQ-uvrC*	-	-	-	SA153	TAT GAT GGT GCG TAT AAC GCG CGG G	*dnaQ-uvrC* Downstream Forward Exo III	1394	-	SA104
SA103	GGAGCAAGT GCAGAACAT GGAGCTGGA AAGA	Forward 2nd segment *dnaQ-uvrC*	1907	-	-	SA154	GAA AGC CGC CGG CCG GTT CGG CAG GCC GTC	*dnaQ-uvrC* Upstream Reverse GVY I	1404	-	SA101
SA104	CAGGTACCGC GACGCGGTG GTGCTCAA	Reverse 2nd segment *dnaQ-uvr*C	-	Kpn I	-	SA155	CCG GCG GCT TTC CTG TTC C	*dnaQ-uvrC* Downstream Forward GVY I	1226	-	SA104
SA146	GATAAGCTT TCTTGACGG TGTAGGTGG	Forward. Upstream *Msm_uvrC*	1044	Hind III	SA98	SA156	CGC GGT CGC AGC GAA GAG CAC CT	*dnaQ-uvrC* Upstream Reverse GVY II	1449	-	SA101
SA147	TTGGTACCT AGCGCTTCCC CTCCCGTC	Reverse Downstream *Msm_uvrC*	879	KpnI	SA99	SA157	TTC GCT GCG ACC GCG GTC GAT CTG	*dnaQ-uvrC* Downstream Forward GVY II	1184	-	SA104
SA148	CCC TGC CGA AGA CTG CAC ATT GAC GAC GA	*dnaQ-uvrC* Upstream Reverse Exo I	801	-	SA101	SA164	GATAAGCTT TCTTGACGG TGTAGGTGG	Forward Upstream *Msm_uvrC*	1044	Hind III	SA98
SA149	AAT GTA CAG TCT TCT GCA GGG CGC GCC AAG G	*dnaQ-uvrC* Downstream Forward Exo I	1838	-	SA104	SA167	TTGGTACCT AGCGCTTCCC CTCCCGTC	Reverse Downstream *Msm_uvrC*	879	-	SA99
SA150	GTT GTA CGC CCC GTC CTA CCG CCA C	*dnaQ-uvrC* Upstream Reverse Exo II	1060	-	SA101	SA98	GTCCACCTGA TCGGGCACG GTTTCCAG	Rev. Upstream *Msm-_uvrC*	-	-	SA99
SA151	TAT GAT GGT TAT AAC ATC GGC TTC CTG	*dnaQ-uvrC* Downstream Forward Exo II	1577	-	SA104	SA99	ACCGTGCCCG ATCAGGTGG ACCGGGTAT GAGGGCCG	Forward Downstream *Msm_uvrC*	879bp	-	SA98

Table VI- Oligos used in this study

2-2- DNA manipulations:

2-2-1- *E. coli* small-scale plasmid extraction:

Plasmid DNA was purified using DNA Clean & Concentrator™ (Zymo Research). Over 10 mL of an overnight *E. coli* culture containing the recombinant plasmid was centrifuged at 16,100 × g for one minute. The supernatant was poured off and the pellet was resuspended in 2:1 volume of DNA Binding Buffer and was mixed briefly by vortexing. Then the mixture was transferred to a Zymo-Spin™ column in a collection tube and was centrifuged at 13000 rpm for one minute, the flow through was discarded. 200µL of DNA wash Buffer was added to the column before another centrifugation for one minute. The washing step was repeated twice, and the column was transferred to a clean 1.5 mL microcentrifuge tube. A variable volume of DNA Elution Buffer was added (between 6 to 20 µL), and the column was incubated at room temperature for one minute before being centrifuged again to elute DNA. The DNA concentration was then measured by a Nanodrop. The final solution was stored at -20°C.

2-2-2 : Purification of DNA from agarose gels and PCR reactions:

After restriction enzyme digestion and DNA separation by electrophoresis, the agarose gel was visualized on a UV trans-illuminator 2000 (Bio Rad). Visualization was performed under blue light to prevent the occurrence of mutations of the DNA sequence that may be caused by the use of UV light. The band of interest was cut out of the gel with a scalpel, weighed and purified using a Monarch® DNA Gel Extraction Kit. Four volumes of Gel Dissolving Buffer were added to the tube containing the slice of agarose before being incubated for 10 minutes at 55°C. The sample was transferred to a column in a collection tube and centrifuged for one minute at 13000 rpm. The flow through was discarded and 200 µL of DNA Wash Buffer was added to the column before being centrifuged for one minute. This step was repeated twice before the column was transferred to a clean 1.5 mL microcentrifuge tube. A variable volume of DNA Elution Buffer was added (between 6 to 20 µL), and the column was incubated at room temperature for one minute before being centrifuged again to elute DNA. The DNA concentration is then measured by a Nanodrop. The final solution was stored at -20°C.

2-2-3 Genomic DNA extraction from *Msm*:

50 mL of an overnight *Msm* culture was centrifuged at 4500 rpm for one minute. The supernatant is poured off and the pellet is resuspended in 500 µL of 1X TE Buffer (10 mM Tris-Cl; 0.1mM EDTA, pH = 8.0). 50 µL of a 10 mg/mL lysozyme solution was added and the solution was mixed. Then, the solution was incubated at 37°C overnight. 70µL of 10% SDS and 50 µL of a 10 mg/mL proteinase K were added to the solution. After a brief vortex, the solution was incubated for 2 hours at 37°C shaking at 400 rpm on a thermomixer. 80 µL of pre-heated 5 M NaCl solution were added to the mixture. Then 80 µL of a 10% CTAB solution (10% N-acetyl-N, N, N-trimethyl ammonium bromide and 40% NaCl) was added and the mixture was vortexed. The mixture was then incubated on a thermomixer at 65°C for 30 minutes while shaking at 400 rpm. The mixture was frozen at -80°C for 15 minutes and then was thawed on ice. After being re-incubated at 65°C for 15 minutes on a thermomixer at 400 rpm, 700 µL of a Chloroform: Isoamyl alcohol (24:1) solution was added to the sample. This was mixed by inversion and then centrifuged at 13000 rpm for 10 minutes at room temperature. Afterwards, the upper phase (aqueous phase) was transferred into a clean 1.5 mL tube along with 50 µL of 10 mg/mL of a RNase solution. After 2 hours of incubation, 700 µL of a Chloroform:Isoamyl alcohol (24:1) solution was added to the sample which was mixed by inversion and then centrifuged at 13000 rpm for 10 minutes at room temperature. Afterwards, the upper phase (aqueous phase) was transferred into a pre-frozen tube containing Isopropanol. After being mixed by inversion, the tube was kept at 4°C overnight. The solution was centrifuged at 13000 rpm for 20 minutes. The supernatant was then discarded, and the pellet was resuspended with 500 µL of ice-cold 70% ethanol and centrifuged at 13000 rpm for 10 minutes. The supernatant was poured off and the pellet was dried on a Speed Vac DNA concentrator (Gene Vac). The pellet was then resuspended into 50 µL ddH2O. The genomic DNA was then visualised on a 1% agarose gel.

2-2-4- DNA sequencing:

DNA was sequenced using Sanger sequencing at the Central Analytical Facilities (CAF) DNA Sequencing Unit at Stellenbosch University. Sequencing data were visualized

and analysed using the SnapGene software (http://www.snapgene.com/). Sequencing primers are indicated in Supplementary Table II.

2-3- DNA manipulation for cloning:

2-3-1- Digestion of DNA with restriction enzyme(s)

Enzymes were obtained from New England Biolabs (NEB) and used as per manufacturer's instructions. Up to 1 µg plasmid DNA and 1 µg PCR product was digested in a 20 µl reaction volume. Digested samples were incubated at the recommended temperature for 1h. DNA fragments were then separated and analyzed on agarose gels by electrophoresis.

2-3-2- DNA dephosphorylation

Following digestion from plasmid DNA, the vector was treated with Antarctic Phosphatase (NEB) to remove the 5'-phosphate group from the linearized vector DNA to prevent self-ligation. Dephosphorylation reactions were carried out at 37°C for 1h followed by inactivation at 80°C for 2 min. The DNA was purified using the DNA Clean & Concentrator™ as per manufacturer's instructions.

2-3-3- DNA ligation

DNA ligations were performed using T4 DNA Ligase (NEB), as per manufacturer's instructions. Blunt-end ligations were performed at 4°C and sticky-end ligations were performed at 16°C overnight.

2-4 Transformation of bacterial cells

2-4-1 Heat-shock transformation of E. coli cells

50 µL of DH5α *E. coli* cell were made chemically competent and stored at -80°C. Bacteria go through a series of centrifugation and resuspension in a CaCl2 buffer. During the entire protocol, bacteria are kept at low temperature (either on ice or at 4°C during

centrifugation). Before being stored at -80°C, bacteria are resuspended in a 15% v/v glycerol solution.

For experiment cultures were thawed on ice. Then they were added to 1 to 15 ng of DNA (ligated plasmid) in a pre-chilled 1.5mL tube. The solution was kept on ice for 30 minutes. Thereafter the solution is heat-shocked at 42°C for 30 seconds, then it is cooled down by placing on ice for 2 to 5 minutes. 500 µL of room temperature SOC broth is added and the culture is incubated at 37°C for one hour with shaking (110 rpm) to allow cells to recover. After that, cells are spun down at 13000 rpm for one minute. The cell pellet was then resuspended in 200 µL of LB broth and is plated onto LB agar plate with the appropriate antibiotic to ensure the selection of the transformed cells. The plates are incubated at 37°C overnight to allow antibiotic resistant transformants to grow.

2-4-2 Electroporation into *Msm*

Electroporation is performed using a Biorad Genepulser Xcell^{tm} electroporator unit with the following parameters: 1200V of constant current, 25µF of capacitance, 1000 Ω of resistance and a cuvette size of 1 mm. 50µL of cell-glycerol (10%) suspension is used in each electroporation with either 1, 2 or 4µg of DNA. Electro-competent cells were prepared by serial centrifugation and resuspension in a ice-cold 10% glycerol solution. Immediately post electroporation, the cell suspension is rescued into 500µL of 7H9 broth and incubated at 37°C for 4 hours to allow cells to recover. Thereafter, the cell suspension is centrifugated at 13000 rpm for 1 minute and the pellet is resuspended into 100µL of supernatant before being plated onto 7H10 plates containing the selection antibiotic. The cells are then allowed to grow for 3-5 days at 37°C to allow the plasmid to be integrated into the mycobacterial genome.

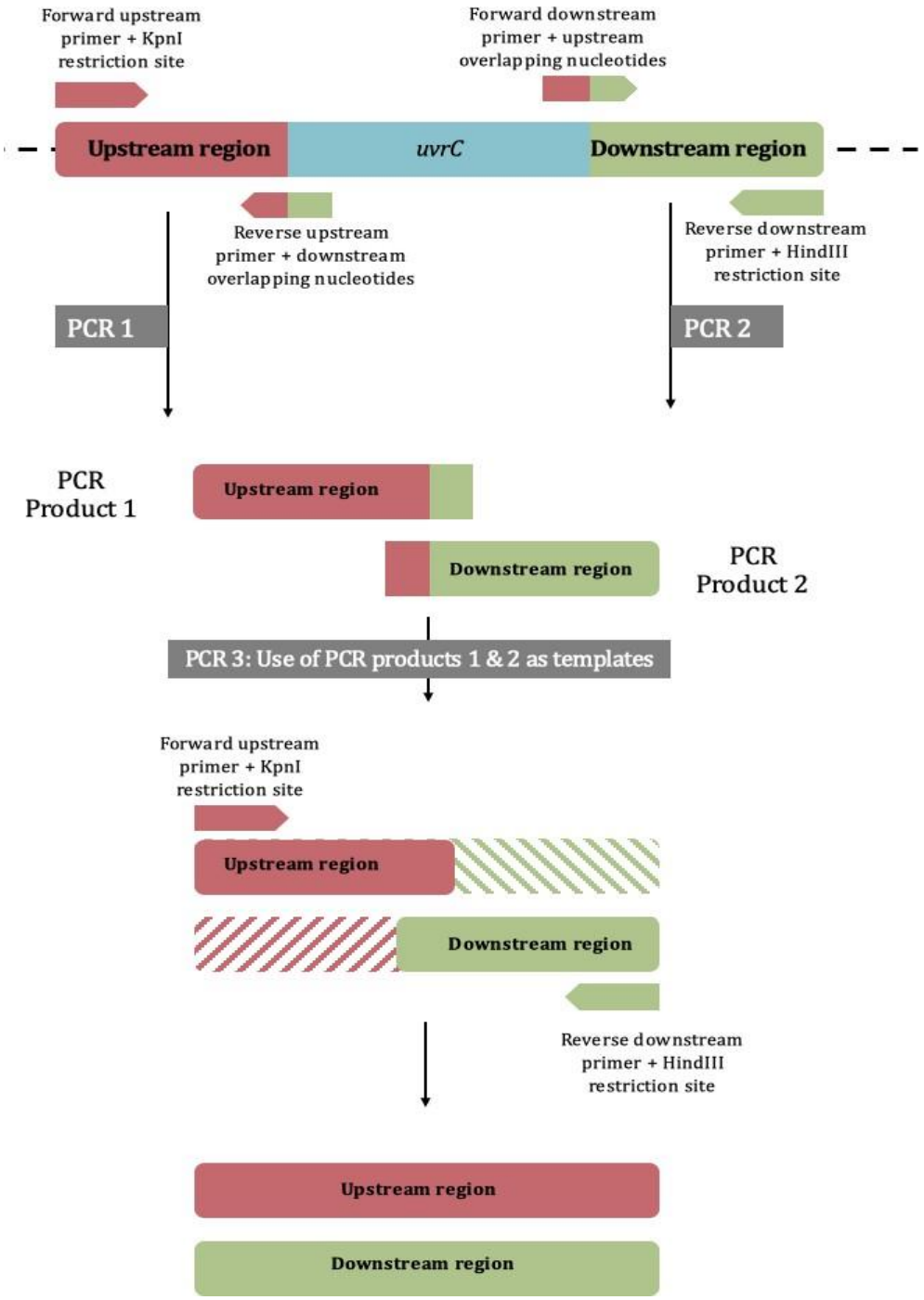


Figure V- Amplification of the targeted gene flanking sequences to clone them in the suicide vector: PCR 1 and 2 products were used as templates for PCR 3 to generate the final PCR product that was digested with *Hin*dIII and *Kpn*I restriction enzymes for cloning into the corresponding sites in the suicide vector.

2-5- Generation of unmarked *Msm* mutant strains by gene replacement:

2-5-1- Construction of suicide delivery vector:

As illustrated in Figure V, gene deletion was performed by cloning the targeted gene flanking sequences into the *Hin*dIII and *Kpn*I sites of p2NIL which contains a Kanamycin resistance gene. The selection cassette from pGOAL17 was then cloned into the *Pac*I site of p2NIL. This selection cassette contains *lacZ* and *sacB* genes which will allow the selection of the single and double cross-over recombinant strains.[58]

2-5-2- Isolation of mutant strains:

The suicide delivery vector containing the targeted gene flanking sequence (Figure V-VI) was then electroporated into wildtype mc^2155 *Msm* and the subsequent electroporated cells were plated on 7H10 OADC containing the selection antibiotic. Putative single cross-over recombinant strains, that were growing on the selection plates were selected to grow in liquid 7H9 OADC with no antibiotic to allow the second cross-over event to happen. After reaching OD_{600}~0.4 to 0.6, cells were plated on 7H10 plates containing 2% sucrose and X-gal to allow the detection of putative double cross-over recombinant mutants[53]. White sucrose-resistant colonies were screened by PCR using targeted gene flanking region primers to confirm the gene deletion. Afterwards, the genotype of the obtained unmarked mutant strain was ascertained by PCR and genome sequencing.

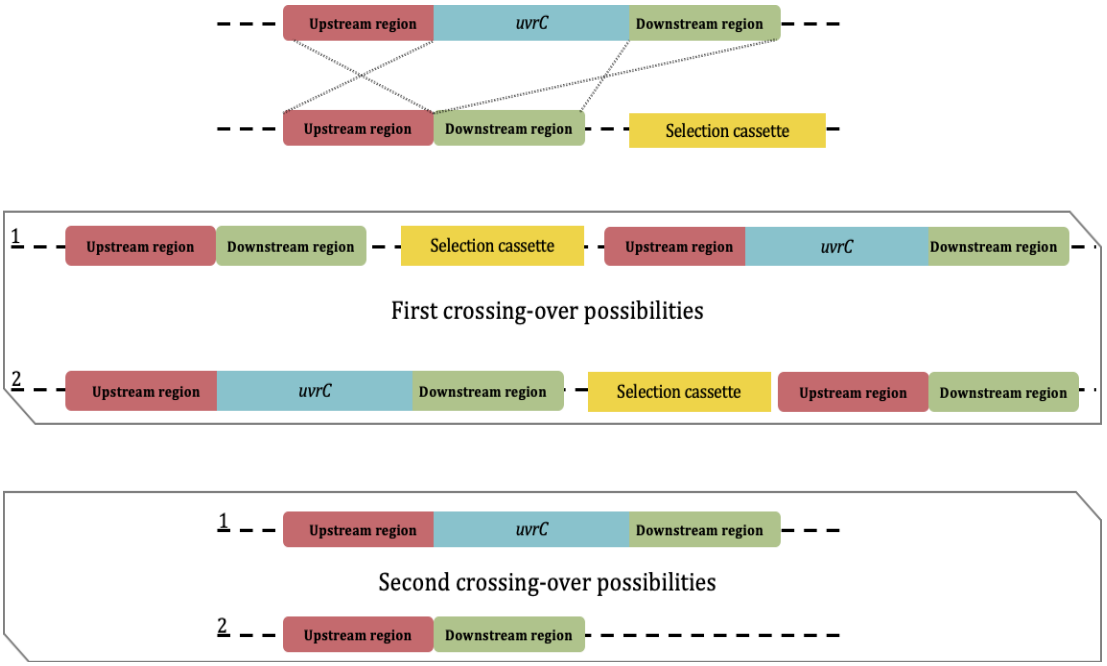


Figure VI- Experimental procedures involved in making an unmarked gene deletion process: The suicide vector contains the targeted gene deletion, homologous flanking regions along with pGOAL17 selection cassette. The first cross-over event will have two potential genome integration outcomes. Similarly, the second cross-over event will also have two potential outcomes, either the deletion of the targeted gene or the restoration of the wild type phenotype.

2-6- Sequencial targeted inactivation of DnaQ and DnaQ-UvrC domains by site-directed mutagenesis:

2-6-1- Generation of *dnaQ-uvrC*$^{\Delta\beta\text{-clamp}}$/ *dnaQ-uvrC* Δ*uvr* BD:

The sequence of the wildtype *Msm dnaQ-uvrC* gene was obtained from the EPFL mycobacterial database (http://mycobrowser.epfl.ch/smegmalist.html). The protocol used was common to the generation of all the mutated versions of *dnaQ-uvrC*. In the case of *dnaQ-uvrC*$^{\Delta\beta\text{-clamp}}$, the pairs of primers SA 101/ 102 and SA103/ 104 (Table VI) were designed to amplify the full length of the targeted gene along with its promotor located downstream of the gene. The upstream reverse and the downstream forward were designed to have their 5' sequence complementary to one an another (Figure V). These overlapping primer sequences also held the required mutation (Table VI). The upstream forward and downstream reverse primers contained an engineered *Kpn*I cutting site to allow the cloning of final PCR product into the integrative mycobacterial shuttle vector *pMV306* (Figure V). Thus, both final PCR product and the vector were digested using *Kpn*I

and ligated to generate pMV306 *attB::dnaQ-uvrC*$^{\Delta\beta\text{-}clamp}$. This construct was cloned into a strain carrying a deleted version of *dnaQ-uvrC*. A version of *dnaQ-uvrC* carrying mutated version of both β-clamp and UVR binding domains was engineered. The process was the same as described above, except that *dnaQ-uvrC* $\Delta^{uvr\ BD}$ was used as a template instead of the wildtype version of the gene. The final PCR product and the vector were digested with *Kpn*I and ligated to generate *dnaQ-uvrC*Δ^{IS}, which is a version of *dnaQ-uvrC* deprived of all its known interaction sites.

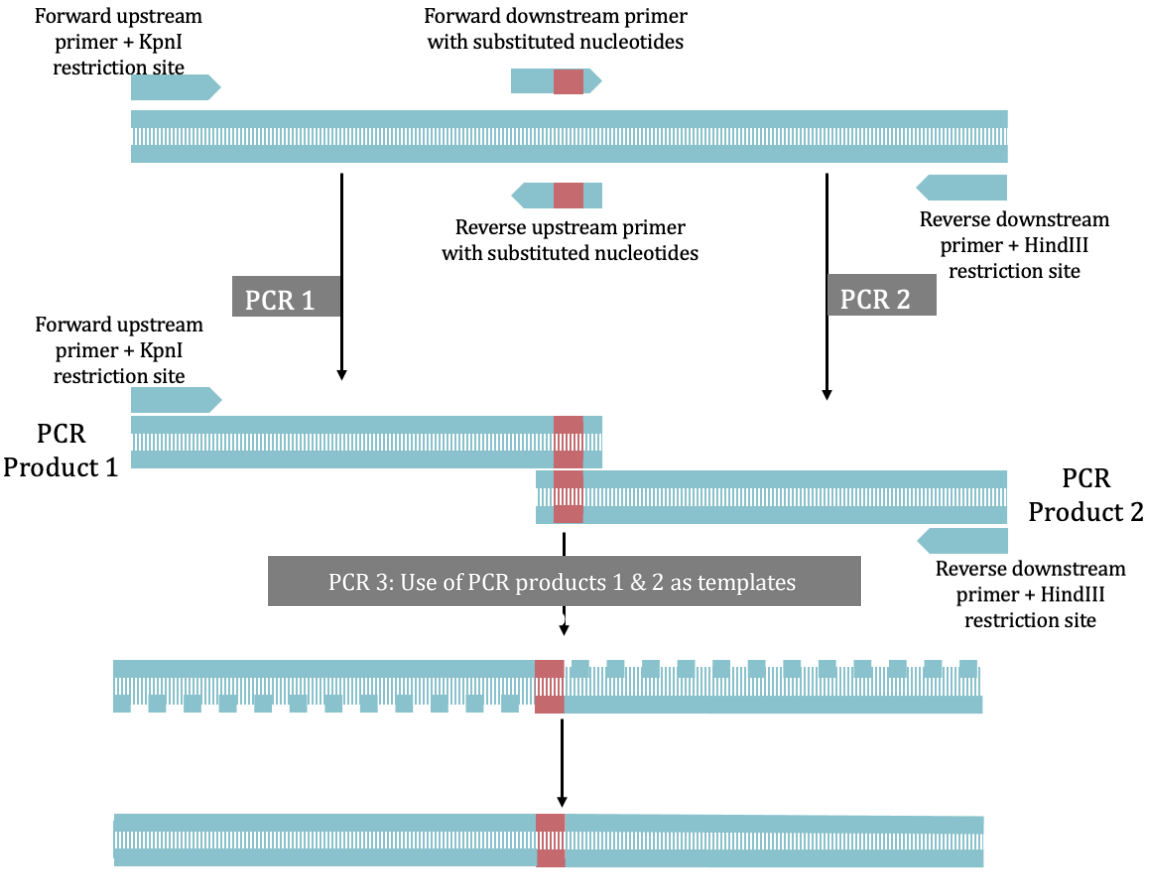


Figure VII- Strategy for sequential targeted inactivation of targeted domains. Products from PCR 1 and 2 were used as templates for PCR3 in order to generate the final PCR product carrying the inactivated targeted domain. PCR product 3 was digested with *Kpn*I before being cloned into the subsequent *Kpn*I sites of the pMV306 transformation vector described in Table V.

2-7- Bioinformatic Study:

2-7-1- Construction of the database:

Protein sequences were obtained from Ecole Polytechnique Fédérale de Lausanne (EPFL)(https://mycobrowser.epfl.ch).[60] They were then submitted to an All-*versus*-all BLAST using HMMer tool. This tool searches for sequence homologs and aligns them. Three runs were completed to identify homologs of UvrC, DnaQ and DnaQ-UvrC in the reference proteomes database (Supplementary Table IV). Because if the similarity between each protein, some homologs could be found in more than one database.

2-7-2- Protein clustering:

The clustering was performed using CLANS (CLuster ANalysis of Sequences), based on full-length sequence protein homology using *Msm* protein sequence as a query. CLANS is a Java program for visualizing the relationship between proteins based on their all against-all pairwise sequence similarities. The program implements a version of the Fruchterman-Reingold force directed graph layout algorithm to present the sequence similarities in a 2D or 3D graph.[61] The E-values of the BLAST High-scoring segment pairs (HSPs) were used to calculate the attractive forces between each sequence pair. The lower (better) the E-value, the higher the attractive force. In addition, each sequence repulses every other sequence with a certain force (inversely proportional to their distance in space). Clustering is achieved by iteratively moving sequences according to the force vector resulting from all pairwise interactions (attraction and repulsion).

The CLANS files were generated using the Max Planck Institute Bioinformatic toolkit (https://toolkit.tuebingen.mpg.de). The scoring matrix BLOSUM62[62] was used along with the extraction of blast HSPs up to E values of 1e-30.

2-7-3- Protein 3D structure:

Protein 3D structures were downloaded on Alphafold (https://alphafold.ebi.ac.uk); however, none of the proteins studied has been crystallized so the resulting 3D structures are only predicted. The confidence in 3D structure varies from domain to domain. Indeed, domains that are widespread in bacteria are more likely

to have already been crystallized hence a strong per-residue confidence score.[63,64] The figures were then generated using PyMol an open-source molecular visualization system.[65] This software used Python coding language; a comprehensive user guide is available on https://pymolwiki.org/ website.

2-8- Phenotypic characterisation assays:

2-8-1- MIC Assays:

Mycobacterial cultures were grown until mid-log phase (OD600 = 0.6-1.0). Then each culture was transferred to fresh 7H9 tube at a concentration of 100 cells per microliter. The assays took place on 96 well plates (Cell Star®). All wells from column 1 to 10 contained 50 µL of 7H9 with different concentration of antibiotic achieved through serial dilutions from column 1 to 10. Column 11 and 12 contained respectively the positive (no drug) and negative control (no cells). Then 50 µL of the 100 cells/µL solution were put into each well except for the negative control. After 48 hours of incubation at 37°C, 10 µL of 0.01% resazurin were added to each well and further incubated for 8 hours for the observation of the colour change. Results were read using a plate reader (FLUOstar OPTIMA, BMG Labtech).

The goal of these experiments was to challenge *Msm* strains with several drugs targeting different pathways of the DNA metabolism to highlight potential differential phenotypes of either Δ*dnaQ* or Δ*dnaQ-uvrC*. So, the drugs used (Table IV) targeted several components of the DNA metabolism either by blocking the interaction sites of the β-clamp (DnaN), preventing the DNA gyrase to unwind the DNA (GyrA/B), blocking the activity of the DNA polymerase (DnaE1) or inducing DNA damages in order to trigger DNA repair systems.

2-8-2- Spotting Assays:

Mycobacterial cultures were grown until mid-log phase (OD_{600nm} = 0.6-1.0). Then each culture was transferred to a fresh media at a concentration of 10 000 cells per microliter. Log10-fold dilutions (10^{-1} – 10^{-6}) of each culture were made and plated on a 7H10 OADC solid media containing the tested antibiotic and incubated for 48 to 72 hours

to allow visible colonies to form. These assays were used to confirm any results obtained with antibiotic MIC assays.

2-8-3- UV-induced mutation frequency:

UV-induced mutation frequency assays were performed as an adaptation of the Luria-Delbruck fluctuation assay as described by Boshoff *et al*[58]. Briefly, mycobacterial cultures were grown until mid-log phase (OD_{600} nm= 0.6-1.0). Log10-fold dilutions (10^{-1} to 10^{-7}) of each culture were plated on 7H10 OADC solid media for CFU enumeration. 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 to check bacteria survival and 1 ml was plated on 7H10 OADC plates containing 200 µg/ml rifampicin for determination of Rif-resistant mutants. The mutation frequency was calculated using the software developed in the MMRU by Z. Martin and J. Kent,[66] optimized from the Ma-Sandri-Sarkar (MSS) method as described by Roche and Foster.[67]

2-8-4- Temporal UV-induced mutation frequency:

To determine damage tolerance of the strains over time, UV-induced mutagenesis assays was performed as an adaptation of the assay described by Boshoff et al.[68] This assay performed exactly as section 2-8-3, except that plating occurred at several occasion not only at a single time point. Plating occurred after 8, 12, 16, 20 and 24 hours of incubation post UV exposure.

2-8-5- Antibiotic evolution assay:

Assay was carried out as described in Ragheb et al.[69] Mycobacterial cultures were grown until mid-log phase (OD_{600}cm = 0.6-1.0) and then inoculated into a 96-well plate at 100 cells per microliter. All wells from column 1 to 10 contained 50µL of 7H9 with different concentration of antibiotic achieved through serial dilutions from column 1 to

10. Column 11 and 12 contained the positive (no drug) and negative control (no cells), respectively, as shown in Figure VIII. Then 50 µL of the 100 cells/µL solution were put into each well except for the negative control. Cells were then grown in a 37°C incubator while shaking. until the no drug control reached OD = 1.5-2. Cultures that grew at least 50% relative to the no drug control at the highest concentration of drugs were passaged to a new 96-well plate to a total of 6 passages.

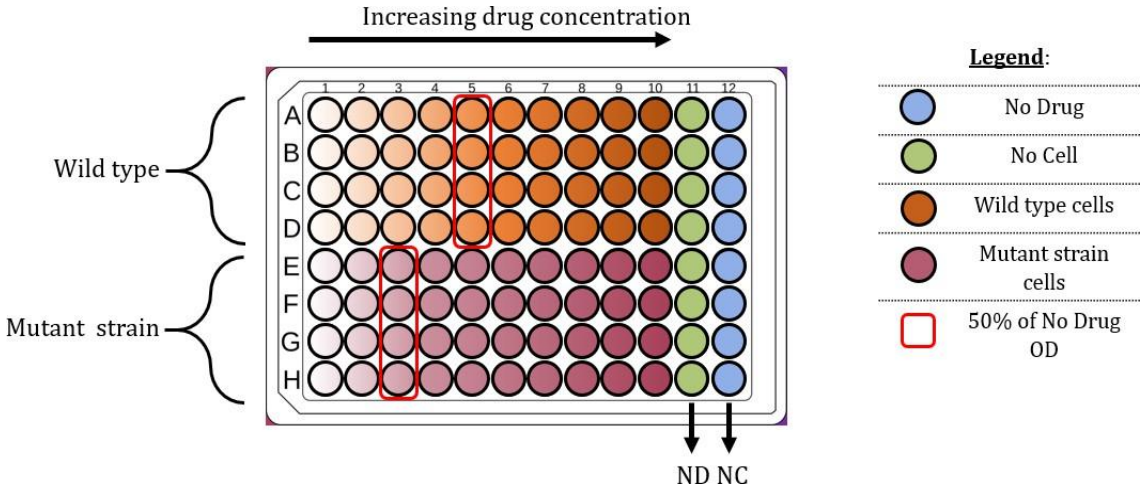


Figure VIII- Antibiotic evolution assay description: Overnight *Msm* strains were diluted back to 100 cells/µL in each well. Drug concentration will be described in Section 5.2.9. ND wells reached OD=1.5/2 (=1.6x10^{8} cells/µL), approximately after 3 days of incubation.

2-9- Generation of CRISPR interference strains:

In this work, only two genes – *uvrC* (MSMEG_3078) and *uvrB* (MSMEG_3816) – were targeted by CRISPRi to confirm the results obtained by gene knock-out strains. Indeed, genotypes all the knock-out strains in this study could be checked by sequencing at the exception of *uvrC* and *uvrB* knock-out strains due to time constraint. Thus, it was thought that repeating experiments with CRISPRi strains would give more confidence in the results obtained. As described by Wong et al,[59] CRISPR interference (CRISPRi) is a means to control sequence specific gene expression via a RNA guide. The associated pJR962 plasmid encodes a catalytically inactive *Sth*1 dCas9 that expression is tetracycline dependent. A golden gate handle is also present that is used to clone a sgRNA along with

a dCas9 handle (Supplementary Figure III) that allows the guide to interact with the dCas9 to be directed to its targeted location (Figure IX).

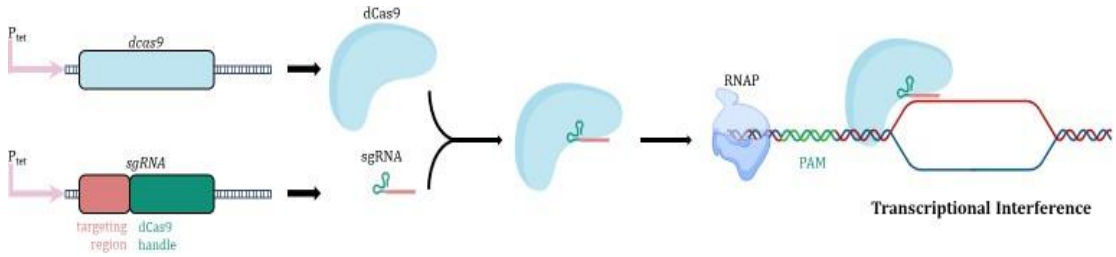


Figure IX- Gene silencing using CRISPR interference: Adapted from Peters et al 2016.[63] The dCas9, the handle and the sgRNA are expressed by the pJR962 plasmid and are then, assembled together. The sgRNA binds to its target further preventing the targeted gene transcription. RNAP=RNA polymerase, PAM= protospacer adjacent motif. Created with Bio Render.

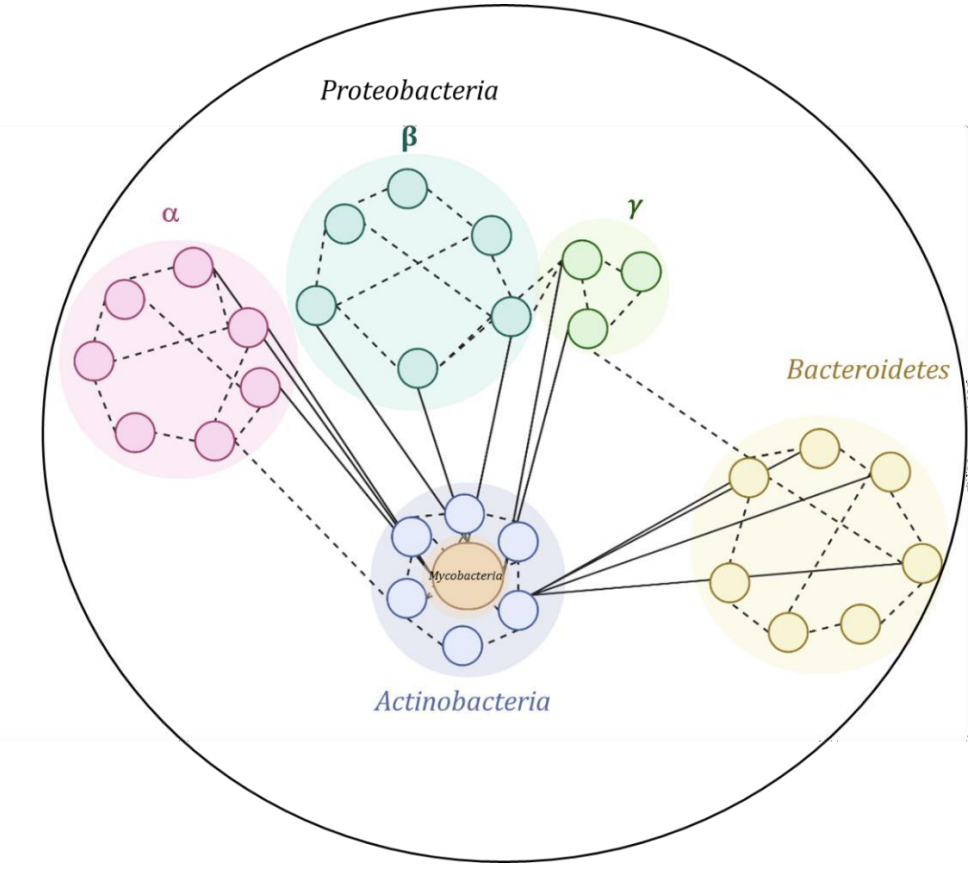
For its targeting purpose, the sgRNA requires a protospacer adjacent motif (PAM). This PAM is a 2-6 nucleotide long specific sequence located downstream of the target DNA sequence and is mandatory for the Cas9 cleavage reaction. There can be many different PAM sequences for a single target with different efficiency. Indeed, the more similar to the consensus sequence NNAGAAW, the more efficient the interference.[70] A list of PAM sequences per gene is available in the MMRU with over 15 PAMs per gene with a score that reflects their efficiency to silent their target (Supplementary Table III).[71]

For *uvrC* (MSMEG_3078) and *uvrB* (MSMEG_3816), two oligonucleotides (Supplementary Table III) were synthesized at 50 nmol and annealed by LGC Biosciences. Upon delivery, oligonucleotides were resuspended in water to a final concentration of 100 µM. Cloning was performed as previously described[62] but adapted for scale. Briefly, the plasmid pJR962 (Supplementary Figure III) was digested overnight with *Bsm*BI-FastDigest (Thermo) and gel purified. Ligations were performed with 1U T4 Ligase (NEB) and incubated at room temperate overnight. Ligation reactions were transformed into 5 µl High-Efficiency DH5α cells (NEB) with heat-shock, rescued in TY Broth, and plated on LB

agar prepared in 6-well plates. Following overnight growth at 37°C, single colonies were picked into 800 µl of LB broth, in 2 ml deep 96-well plates, and grown overnight at 37°C with vigorous shaking. All *E. coli* cultures were supplemented with Kanamycin (Roche) at a final concentration of 100 µg/ml. Plasmids were extracted with a Zyppy-96 Plasmid Miniprep kit (Zymo) according to manufacturer instructions, quantified via Nanodrop and sent for sequencing. Subsequent electroporation into *Msm* were performed as described in section 2-4-2. The reason for targeting only uvrC and uvrB by CRISPRi was because uvrC and uvrB knock-out strains could not be sequenced to ascertain their genotype. Therefore, using CRISPRi gave confidence that results obtained by gene knock-out were reliable.

CHAPTER III:

Unusual interaction domains highlight different functions for DnaQ homologs in mycobacteria:



Created with BioRender

3-1- Introduction:

3-1-1- Investigating protein domain organisation:

A core objective of this work was to determine the functions of the putative *Mtb* DnaQ (Rv3711c; *Msm* MSMEG_6275) and DnaQ-UvrC (Rv2191; *Msm* MSMEG_4259) homologs. The paucity of published literature on these mycobacterial genes, and the constraints of the COVID-19 pandemic (which prevented laboratory work), provided the motivation and opportunity to conduct a thorough investigation into the use of bioinformatic tools for protein/structure function analysis. As mentioned in Chapter I, DnaQ (Rv3711c; *Msm* MSMEG_6275) does not appear to be effective as a proofreading subunit of the mycobacterial replisome *in vivo* with proofreading shown to be performed by the PHP domain of the DNA polymerase IIIα-polymerase.[31] Although deletion of *dnaQ* impaired the processivity of the polymerase in biochemical assays,[33] the reason for its inability to perform an exonuclease activity remained unclear. Therefore, this work aimed to determine the potential for functional redundancy and/or existence of a mycobacterium-specific protein domain structure/ architecture that prevented this activity from being carried out. Following the successful *in silico* modelling of DnaQ, a similar workflow was used to study DnaQ-UvrC (Rv2191; *Msm* MSMEG_4259). The workflow is described below.

The bio-informatic study was performed using information and tools available online and included both experimentally verified observations and bioinformatic models and predictions. The annotated *Mtb* H37Rv and *Msm* mc^{2}155 genomes are available on the Ecole Polytechnique Federale de Lausanne (EPFL) website (http://mycobrowser.epfl.ch).[60] The database is updated regularly with *in silico*, experimental, and other information including functional annotations, 3D structures, presence of orthologues in other mycobacterial species, etc.[60] These updates are gathered from published literature, and other bioinformatic websites – many of which were used in this work.

Among these websites, InterPro is a central data resource regarding protein signatures. Indeed, it harbours over 189 million sequences that are classified in about 292,000 proteomes comprising multiple organisms including bacteria. This data is gathered from 13 other protein databases and is constantly updated, especially from genomics and

proteomics studies. Although such a mass of data might contain overlapping information, it is a powerful tool to gather knowledge on a specific set of proteins.[72–76] When available, the database gives access to extensive information about a specific protein in a specific organism, from function, expression, cellular location, interactions, structure, and family domain and, last but not least, known similar proteins (Figure X). Owing to the lack of experimental data for the proteins of interest in this study, this last aspect was especially useful. The strength of InterPro is to display a graphical summary of all the relevant data for a protein structure from detected protein domain to the position of key residues. It also provides a list of homologous superfamily entries where similar protein data can be found, as well as the GO terms associated with the protein of interest (Figure X).

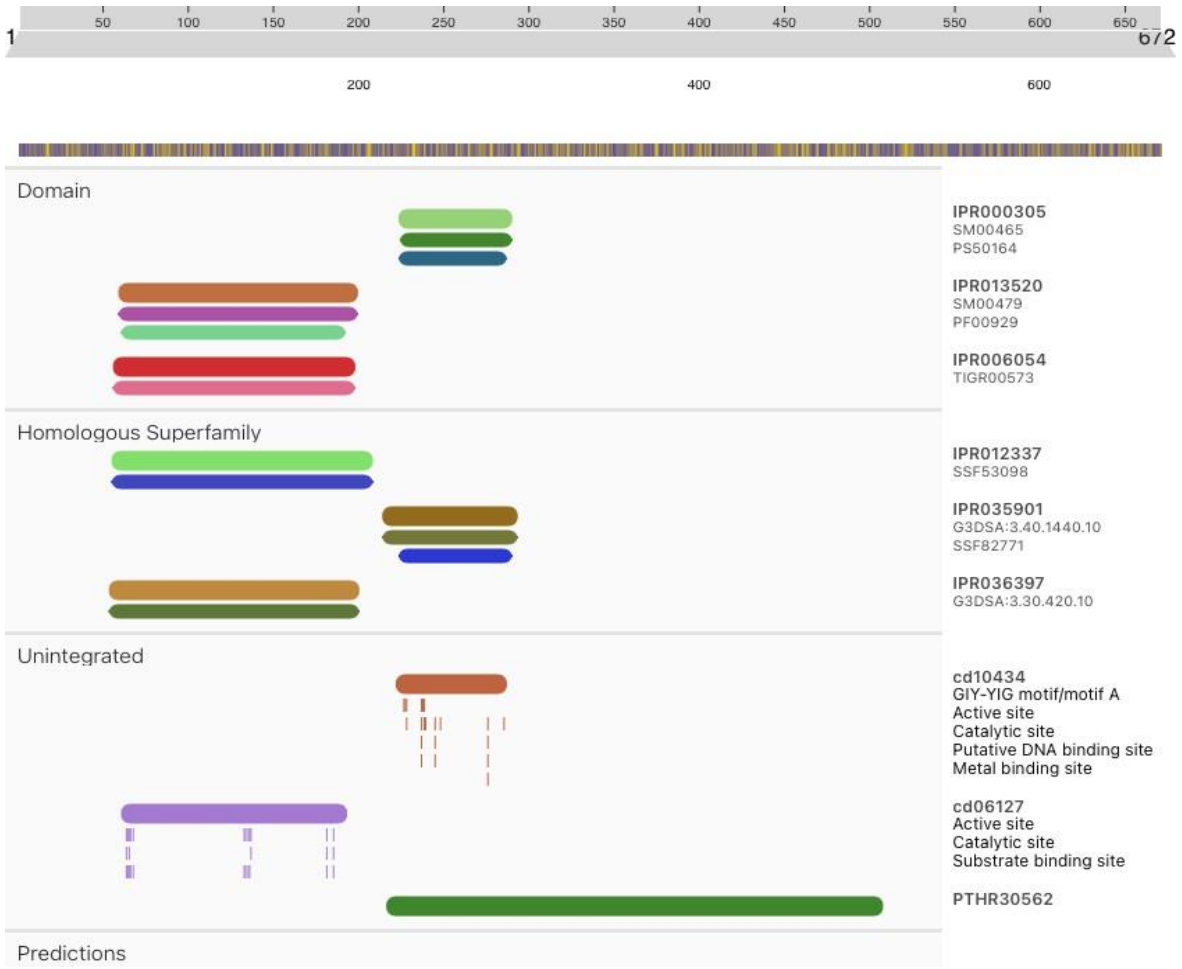


Figure X- Screenshot of Interpro graphical summary- Protein sequence is divided in several signatures that can be either domain, homologous superfamily, site, ... Each entry is linked to the corresponding database that gives information related to the domain, sequence or site clicked on. Several entries can be found for a single signature depending on the referencing available on Interpro.

3-1-2- Predict protein 3D structure:

Computational approaches were used to address more aspects of the proteins than just the identification of domains and/or residues. Predicting the three-dimensional (3D) structure of a protein solely based on its amino acids sequence has been an extremely difficult problem to solve. The advent of machine learning and artificial intelligence (AI)-based prediction tools – most notably, AlphaFold[64] – therefore represents a major enabling advance.[56] Solving 3D structures (Figure XI) is a powerful addition to predict protein/protein interactions and even their evolution. Since its launch, AlphaFold has been able to generate a database of more than 4000,000 predicted structures and it keeps expanding.[63,64]

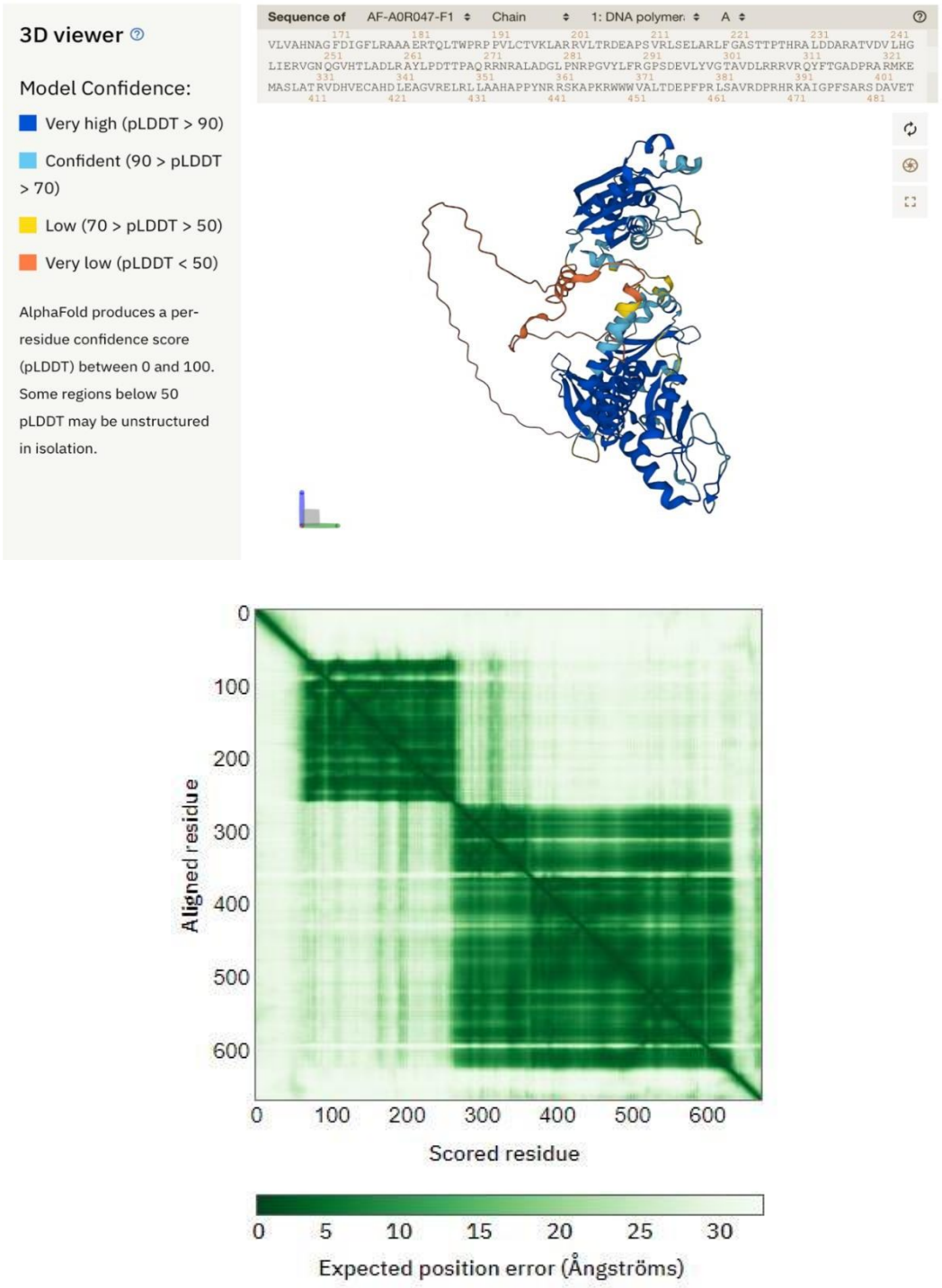


Figure XI- Screenshot of Alphafold entry of MSMEG_ 4259. A- Protein structure prediction- Protein sequence is indicated at the top and editable protein structure is displayed below. Confidence of the model per residues is displayed in a shade of color indicated on the left pannel B - Predicted aligned error. The expected position error score is displayed on the bottom, the closer to 0 Ångström, the greater the confidence. In such proteins that structure has not been solved by crystallography, the IA predicts the position of each residues relative to other residues of the protein based on information gathered in Interpro database. In MSMEG_4259, two domains can be clearly identified from residue 1 to 400 and the other from 400 to 700 which will be further explained below, in section 3-2-1.

The last version of AlphaFold is the most accurate predictive tool ever released.[63] Indeed, it has a median backbone accuracy of 0.96 Å. Furthermore, it can provide programmatic access to and interactive visualization of predicted atomic coordinates, per-residue and pairwise model-confidence estimates and predicted aligned errors. Overall, this gives confidence in AlphaFold's ability to generate reliable protein 3D structures (Figure XI). However, it does not allow to select or highlight multiple parts of interest in a protein or to compare the 3D structure of two or more proteins. For such visualization needs, PyMOL is better adapted.

PyMOL[65] is a molecular graphic software that allows visualization of proteins and smaller molecules even to nucleic acid. Its main utility is to edit molecules, visualize them in different representations such as cartoons, ribbons, dots and more, and even measure bond lengths between residues (Figure XII). Like a growing number of research-targeting software, including a few in this work, PyMOL is written in Python.

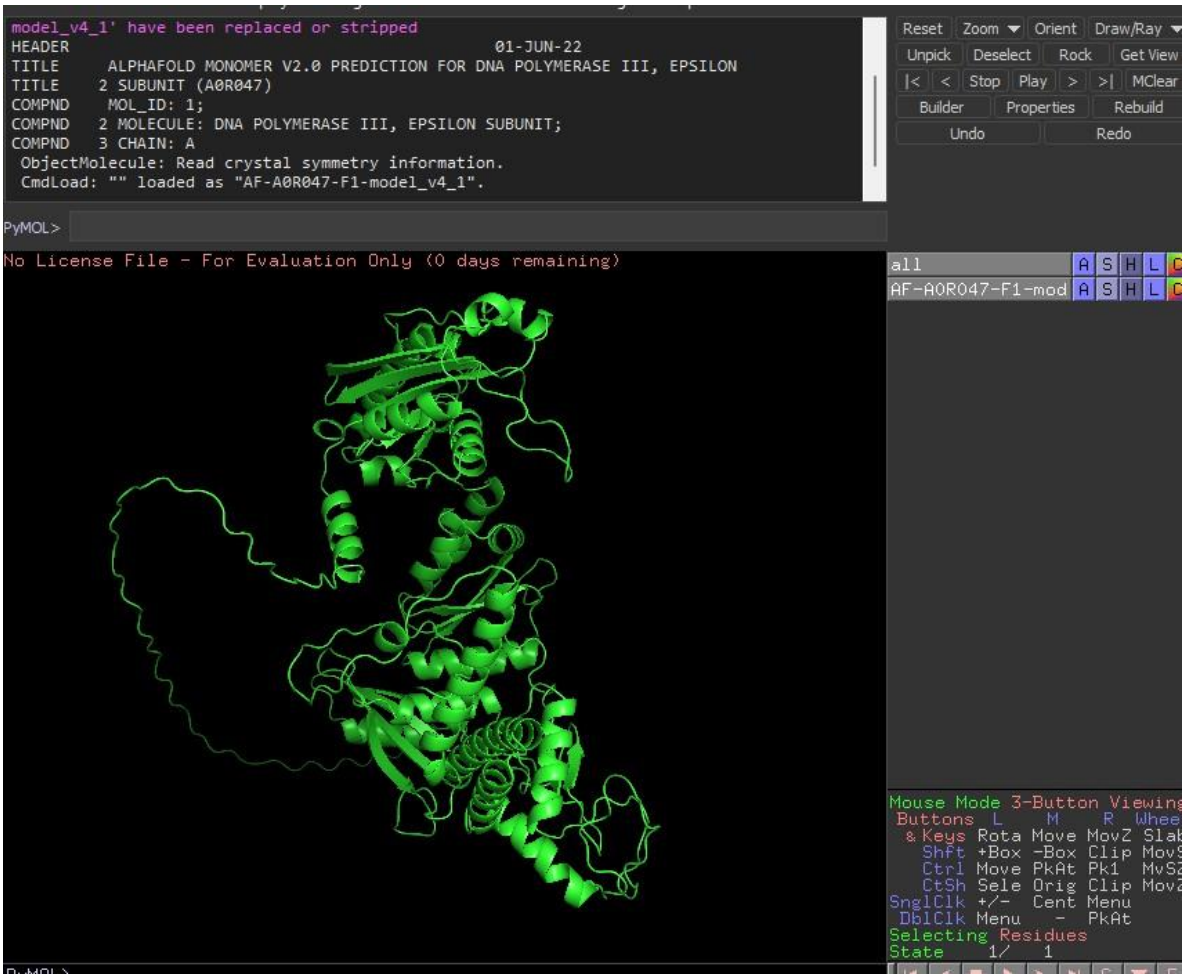


Figure XII- Screenshot of PyMol software displaying MSMEG_4259. Pymol can be used both by clicking or by using the command line. At the top, the command line and the history of command can be found. At the top right part, camera properties can be edited. At the top right part, camera properties can be edited. At the bottom right part, summary of keyboard shortcuts can be found.

3-1-3- Protein sequence homology to help determine function:

Deducing a function of a previously unknown protein based on sequence homology is a standard molecular biology investigation approach.[77] This can be performed by studying sequence similarities, multiple alignments, or phylogenetic analysis. However, the last two can be computationally demanding when studying a large number of proteins. On the other hand, sequence similarities analysis performed by BLAST or PSI-BLAST[77–79] tools are efficient and fast even when working with huge datasets and even allow detection of distantly related homologous proteins. Unfortunately, they are plagued by false positive matches and problems arising from amino acid composition bias causing, in many cases, the best BLAST hits not to be the closest sequence relatives.[80] The CLANS tool

uses a version of the Fruchterman-Reingold graph-layout algorithm[77,78,80,81] to combine the quality of sequence similarities, multiple alignments and phylogenetic analyses. From unaligned FASTA format sequences, the programme performs an all-against-all BLAST search and displays the result as either a 2D or a 3D graph. Similar to Pymol, it required Python coding but, recently, an online toolkit allowing to precompile CLAN files was released[82,83] which facilitates the use of CLANS for users with low skills in Python programming.

In this chapter, DnaQ (*Mtb* Rv3711c; *Msm* MSMEG_6275) and DnaQ-UvrC (*Mtb* Rv2191; *Msm* MSMEG_4259) protein structures were first compared to the *E. coli* model, followed by 3D structure model comparison to determine protein differences that may explain their functional differences. Then, a protein sequence homology search was performed which consisted of the detection and comparison of DnaQ and DnaQ-UvrC homologous protein sequences in over a thousand different organisms. The goal was to detect a homologous protein in another organism which function would already been annotated and thus, help elucidate the function of mycobacterial DnaQ homologs.

3-2- Results:

3-2-1- Structure of DnaQ and UvrC in the model organism:

The functions of the putative *M. tuberculosis* DnaQ (Rv3711c; *Msm* MSMEG_6275) and DnaQ-UvrC (Rv2191; *Msm* MSMEG_4259) homologs are unknown. To gain insight into their potential roles, we first compared each of these proteins with the better characterized *E. coli* model of DNA replication.[84–86]

In *E. coli*, DnaQ is the epsilon subunit of the replisome, and an active member of the replication complex (described in detail in Chapter IV). The DnaQ structure is indicated in Figure XIII and XVI; it is 243 amino acids in length, comprising a classic 3'–5' proofreading exonuclease with DEDDh triad.[87] Three exonuclease active sites are located in the proofreading exonuclease domain, followed immediately by the β clamp-binding motif. The other identifiable site is at the very C-terminus of the protein and is responsible for the interaction of DnaQ with the DNA PolIIIα subunit.[40]

Protein	DnaQ-like exonuclease active sites																GIYYIG endonuclease active site								
	Exo 1					Exo 2					Exo 3						In-site			Off-site					
Consensus	D	+	E	T	T	G	H	N	A	G	F	D	H	+	A	L	+	D	Y	Y	G	+	H	R	Y	E	N
E. Coli *dnaQ*	D	T	E	T	T	G	H	N	A	A	F	D	H	G	A	L	L	D									
MSMEG_6275	D	V	E	T	T	G	H	N	A	G	F	D	H	D	A	L	D	D									
Rv_3711c	D	V	E	T	S	G	H	N	V	A	F	D	H	D	A	F	D	D									
MSMEG_4259	D	L	E	T	T	G	H	N	A	G	F	D	H	R	A	L	D	D	Y	Y	G	T	H	R	Y	E	N
Rv_2191	D	L	E	T	T	G	H	N	A	G	F	D	H	R	A	L	D	D	Y	Y	G	T	H	R	Y	E	N
E. Coli *uvrC*																			Y	Y	G	K	H	R	Y	E	N
MSMEG_3078																			Y	Y	G	K	H	R	Y	E	N
Rv_1420																			Y	Y	G	K	H	R	Y	E	N

Na

Chemical range of amino acids:
- Nucleophile
- Base
- Acid

Table VII- Multiple sequence alignment of the nuclease domains of the studied DnaQ and UvrC homologs. Sequences were aligned using the Clustal ω tool on the MPI Bioinformatic Toolkit (https://toolkit.tuebingen.mpg.de/tools/clustalo). Bold residues represent the nucleases active sites. The chemical range of the residues is indicated with shades. Numeration of the active site is arbitrary.

UvrC, which will be further described in Chapter V, is a 610 amino acid protein that is composed of two nucleases: the N-terminus contains a GIYYIY endonuclease and the C-terminus a RuvA2-like endonuclease. The GIYYIY endonuclease is named after the conserved residues found in the DNA repair nuclease, UvrC. The RuvA2-like endonuclease shares structural homology with RNaseH with an unusual DDH triad.[88] These two nucleases encapsulate a mid-protein domain that is arbitrarily named UvrC-mid. It contains a single interaction site, UVR Binding motif, that is involved in UvrABC protein interaction during Nucleotide Excision Repair (further described in Chapter V).

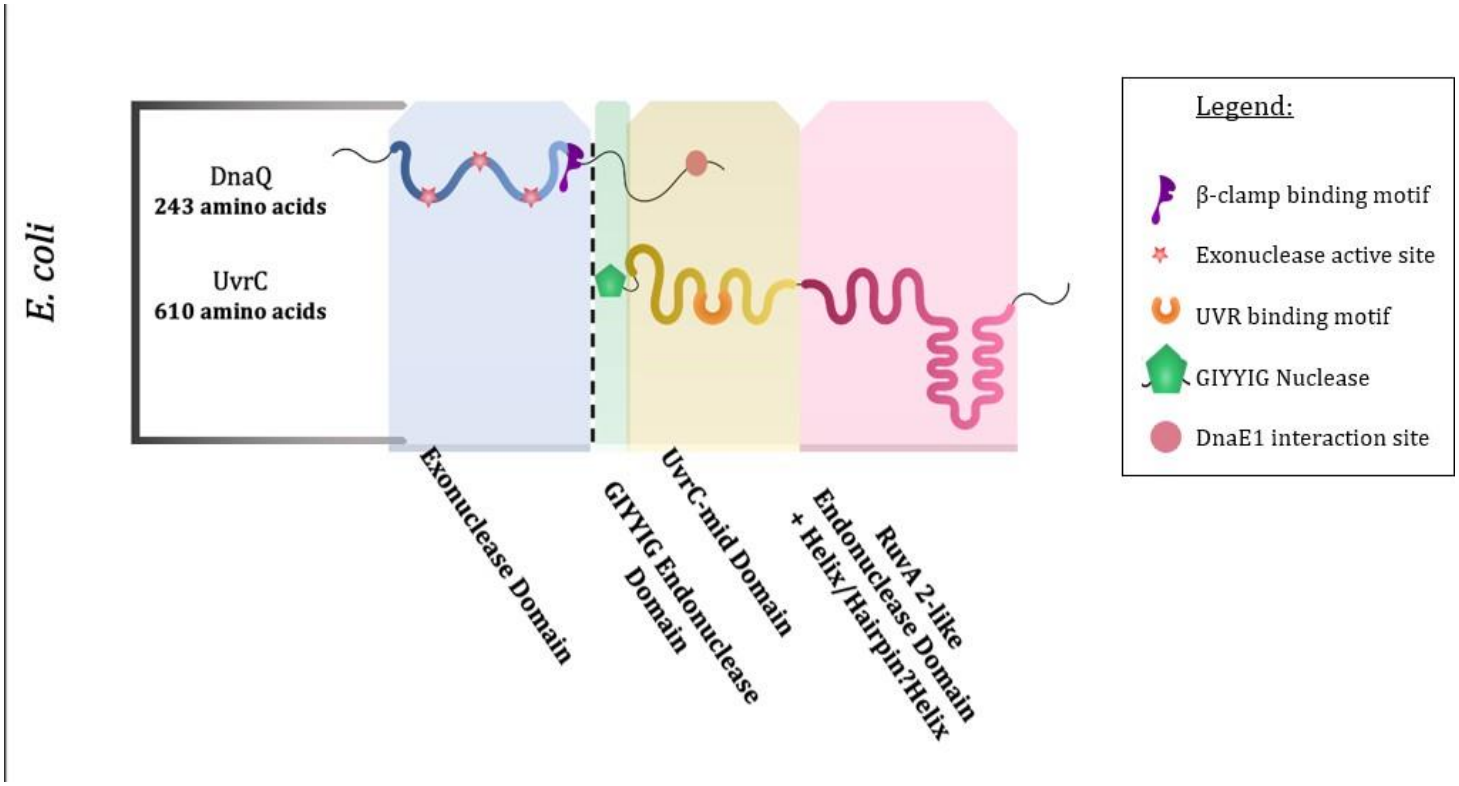


Figure XIII- Domain maps of the *E. coli* DnaQ and UvrC proteins. DnaQ is composed of an exonuclease domain in blue among which three active site are designated by red stars. A β-clamp binding motif in purple follows and a DnaE1 interaction site is located at the C-terminus in pink. UvrC has two nucleases located at opposite extremities of the protein. At the N-terminus, there is a GIYYIG exonuclease (indicated in green) and a RuvA2-like endonuclease at the C-terminus (light purple). The arbitrarily named UvrC-mid domain is indicated in yellow. This domain contains a UVR Binding motif in orange. As the GIYYIG exonuclease, UvrC-mid, and RuvA2-like endonuclease domains are found in different proteins studied in this project, the domains are shaded consistently to facilitate their identification throughout.

3-2-2- Mycobacterial DnaQ homologs exhibit well conserved DnaQ-like 3' – 5' proofreading exonuclease:

All the mycobacterial DnaQ homologs exhibit the same DnaQ-like 3'–5' proofreading exonuclease as *E. coli* at their N-termini. This exonuclease, which is two metal ion-dependent and contains a DEDD (aspartate/glutamate/aspartate/aspartate) motif, usually operates during replication by performing a phosphodiesterase reaction on the two-last added nucleic acids on the new DNA strand, resulting in the generation of 5'-phosphate and 3'-OH groups.[89,90] The reaction is catalysed by two Mg^{2+} metal ions and can be divided into three steps: first, a nucleophilic attack on the 3' bridging P-O bond, followed by the formation of a highly negatively charged P-O_5 intermediate, then the scissile bond is broken and the reaction completed (Figure XIV). Usually, the concerted activity of three kinds of amino acids is required - a nucleophile, a base, and an acid. In the case of DnaQ-like 3'–5' proofreading exonucleases, the nucleophile role is not supported by an amino acid but rather by surrounding water which can be deprotonated by the histidine. Together with the DEDD residues, this forms a DEDDh motif that most members of the DnaQ-like 3′ exonuclease superfamily possess.[89,90]

A multiple sequence alignment was performed for all DnaQ homologs to identify and compare the DnaQ-like exonuclease active sites (Table VII). This analysis revealed that the main catalytic residues are conserved, forming the mandatory DEDDh motif. However, some minor differences could be identified such as a missing nucleophilic amino acid in mycobacterial proteins (threonine) that may affect their exonuclease activity.

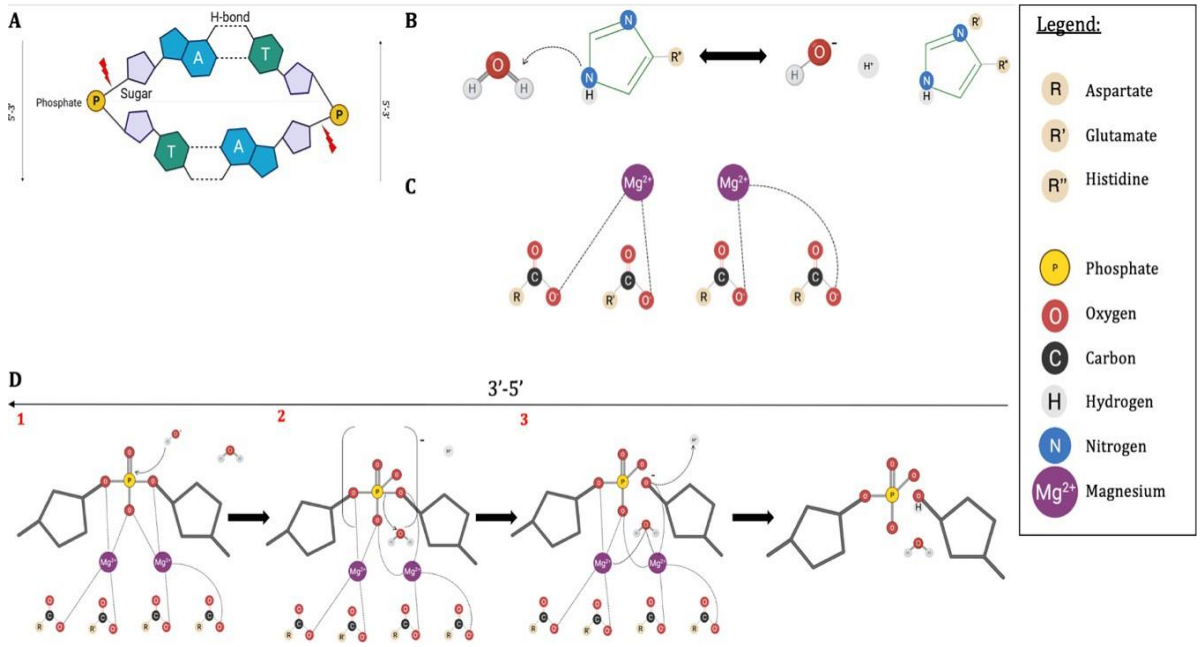


Figure XIV- A: Interaction between nucleotides in dsDNA. Red lightning shows the location of the 3'5' excision. -B: Deprotonation of surrounding water by the Histidine. -C: Interactions of the DEDD motif and the two metal ions. -D: Step of the phosphodiesterase reaction. 1 = Nucleophilic attack. 2= Generation of a negatively charged P-O_5. 3= Breakage of the P-O scissile bond and end of the reaction. Created with BioRender.

3-2-3- An unusual natural DnaQ-UvrC fusion protein found in mycobacteria:

Nucleases are enzymes able to excise nucleic acids from a DNA strand by cleaving phosphodiester bonds. They are necessary for proofreading by allowing removal of a 3' mis-incorporated nucleotide (exonuclease) or to remove damaged nucleic acids within a DNA strand (endonuclease). Both *Mtb* Rv2191and *Msm* MSMEG_4259 contain a second, GIYYIY endonuclease in their C-terminal regions. The GIYYIY endonuclease is named after the conserved residues found in the DNA repair nuclease, UvrC, which is involved in cellular processes such as DNA repair and recombination and is highly conserved across organisms (Table VII). Even though the existence of this domain was previously established, its activity has never been evaluated. However, it is predicted to cleave dsDNA to produce 5'P and 3'OH groups owing to a conserved catalytic triad in which the histidine is responsible for deprotonation of surrounding water that carries out the nucleophilic attack. Unlike 3'–5' exonucleases, most of the catalytic residues are found outside of the GIYYIG motif, and the reaction is dependent on a one-metal-ion mechanism.[89] And, whereas UvrC exhibits a second RuvA2-like endonuclease at its C-

terminal, an equivalent nuclease is not found in either *Msm* MSMEG_4259 nor *Mtb* Rv2191. Nuclease active sites contain a catalytic triad that is required for activity: these amino acids act in a coordinated manner to disrupt the phosphodiester bond between two nucleotides (Figure XIV). In order to proceed with the reaction, they need to be brought close to each together to form an effective active site. This is usually done during the folding that forms the 3D structure (Figure XV and Table VIII).

3-2-4- Protein folding allows catalytic triad to be active:

To gain additional insight, the 3D structures of Rv3711c and Rv2191, for which crystallography data are available, were studied; in contrast, the corresponding *Msm* structure could only be predicted owing to the unavailability of solved structures. As for *E. coli* DnaQ, in *Mtb* Rv3711c and Rv2191(respectively, mycobacterial DnaQ and DnaQ-UvrC) (Figure XV and Table VIII), the characteristic DEDDh residues are close enough to each other to perform the nuclease reaction and cleavage of phosphodiester bonds. Therefore, the reason for the apparent absence of exonuclease activity in the mycobacterial DnaQ homologs – as inferred from microbiological assays[33] – remains uncertain.

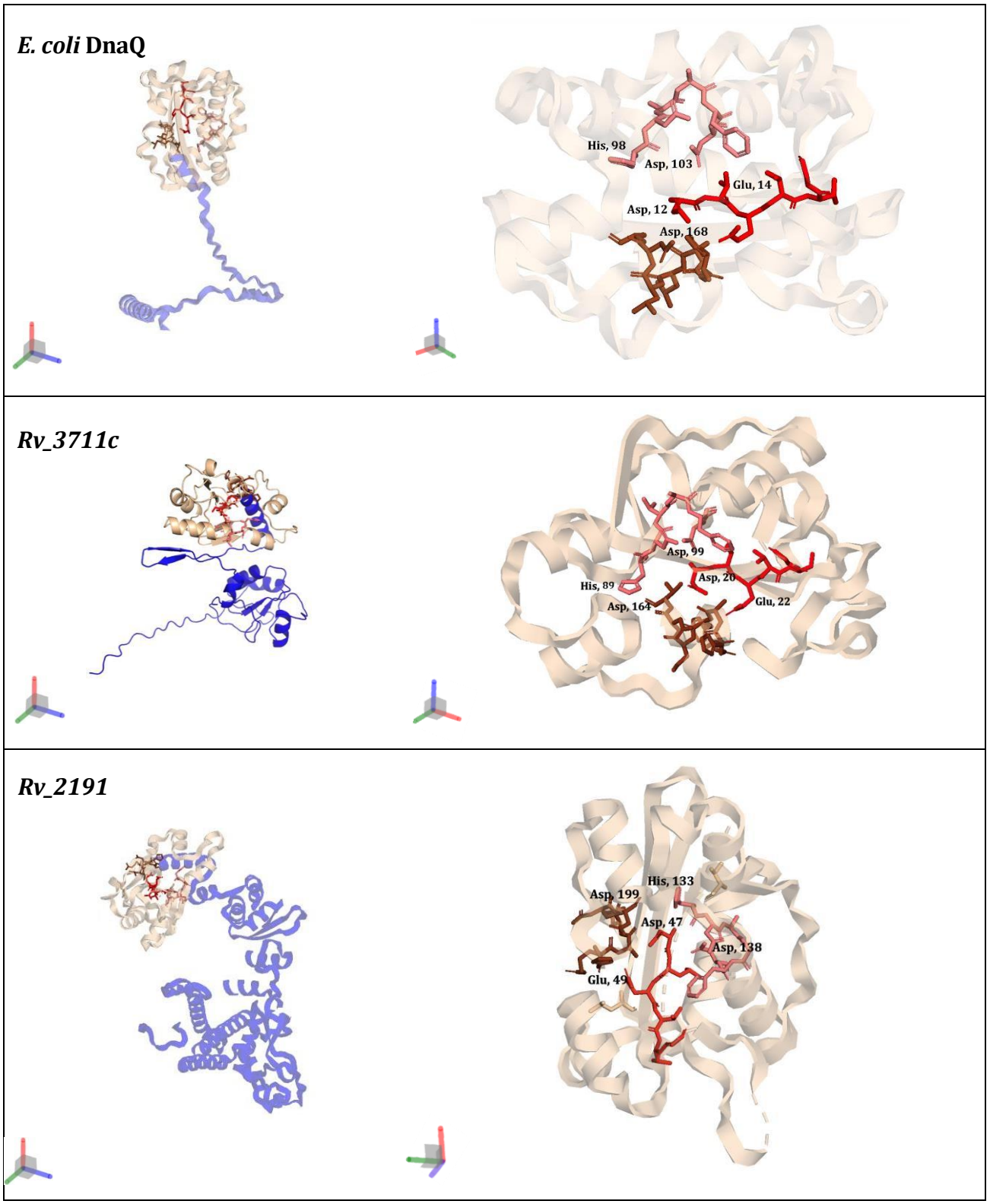


Figure XV- Highlight of the catalytic active site. On the left: full view of the protein. On the right: Zoom of the active site. 3D structures were downloaded from Alphafold (https://alphafold.ebi.ac.uk/entry/P9WQK7. They were then colored and labelled using PyMol software(https://pymol.org). The domains of the proteins are identified as following: Exonuclease domain , Exo I active site , Exo II active site , Exo III active site , Rest of the protein .

Protein	**Exonuclease catalytic residues**				
Rv_2191		Asp, 47	Glu, 49	His, 133	Asp, 138	Asp, 199
	Asp, 47					
	Glu, 49	6.3				
	His, 133	12.4	13.7			
	Asp, 138	7.3	10.1	9.2		
	Asp, 199	4.6	6.1	8.8	10.2	
Rv_3711c		Asp, 20	Glu, 22	His, 89	Asp, 99	Asp, 164
	Asp, 20					
	Glu, 22	6.3				
	His, 89	17.2	21.6			
	Asp, 99	8.1	12.9	25.3		
	Asp, 164	4.9	8.7	19.5	8.8	
***E. coli* DnaQ**		Asp, 12	Glu, 14	His, 98	Asp, 103	Asp, 168
	Asp, 12					
	Glu, 14	6.4				
	His, 98	9.5	12.5			
	Asp, 103	5.6	12.3	9.8		
	Asp, 168	6.5	7.3	11.3	11.5	

Table VIII- Average distance between residues in angström - Average distance was calculated by averaging distance between the atoms of each residues. Calculation were made using Pymol measurement function.

3-2-5- Mycobacterial DnaQ homologs exhibit unusual interaction sites:

Like *E. coli* DnaQ, *Mtb* Rv3711c and Rv2191, and their *Msm* MSMEG_6275 and MSMEG_4259 homologues, appear to possess the ability to access the replication fork. This capacity is inferred based on the presence, in all four proteins, of a β clamp-binding motif. Such a motif allows a protein to interact with the β sliding-clamp. This clamp itself acts as a bridge to allow all the components of the replication machinery to work together during replication and repair processes.[33,91,92] It will be further described in section 4-1-3.

Although the small size of the β clamp-binding motif makes it difficult to detect,[93] it appears well conserved among the *E. coli* and mycobacterial DnaQ homologues (Table IX). Even so, the functionality of the β clamp-binding motif in DnaQ-UvrC (Rv2191; *Msm* MSMEG_4259) requires experimental validation, although some confidence in the prediction that the sequence is a valid a β-clamp binding motif is derived from the observation that it is followed by a flexible linker, which may allow MSMEG_4259 to interact with the replisome complex. The main difference was found in the location of the β clamp-binding motif (Figure XVI). Indeed, in the mycobacterial DnaQ, it is located at the very C-terminal whereas in DnaQ-UvrC (Rv2191; *Msm* MSMEG_4259), it is at the N-terminal (Figure XVI). Notably, all mycobacterial DnaQ homologs are different from the *E. coli* model, and this may impact the way each mycobacterial DnaQ homolog interacts with the replisome complex and the DNA. This possibility is reinforced by the fact that the DnaE1 interaction site was not detected at all.

The UvrC-mid protein domains of the mycobacterial DnaQ-UvrC proteins (MSMEG_4259 and Rv2191) contain a UVR motif (Table IX) which, in NER, is thought to allow UvrB and UvrC proteins to interact with each other to incise DNA damage. [158] The type of the third residue of the motif differs in *E. coli* UvrC *versus* either *Msm* UvrC (MSMEG_3078) or *Mtb* UvrC (Rv1420) but is conserved in the mycobacterial DnaQ-UvrC (Rv2191; *Msm* MSMEG_4259) proteins. In addition, the mycobacterial proteins DnaQ (*Mtb* Rv3711c; *Msm* MSMEG_6275) are notable for containing a carboxy-terminal domain of the Breast Cancer Gene (BRCT) domain in their C-terminal regions, just before the β clamp-binding motif (Figure XVI). The BRCT domain is found in numerous proteins in both prokaryotes

and eukaryotes and, usually, are involved in DNA damage-checkpoint or DNA repair pathways,[94] though their precise function remains to be fully elucidated.

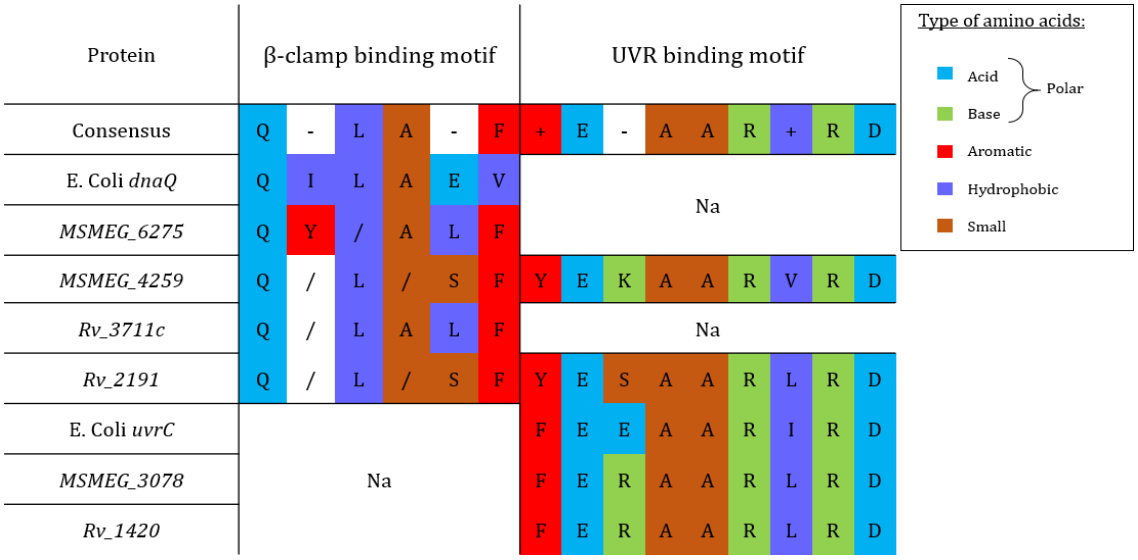
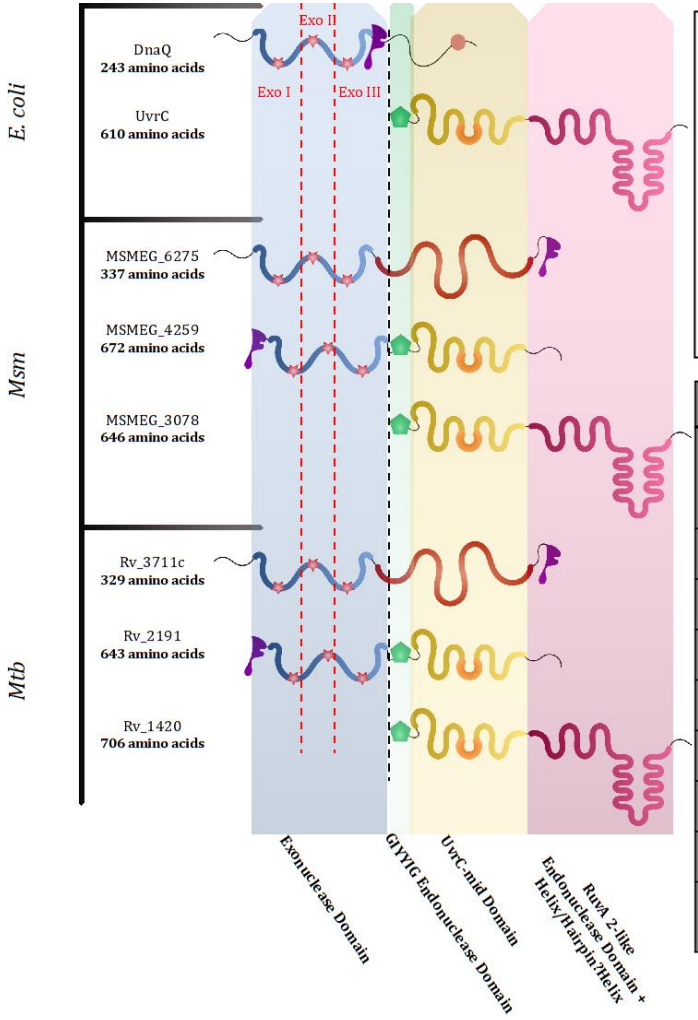
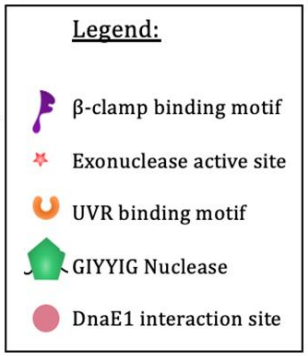


Table IX- Multiple sequence alignment of the β-clamp and UVR binding motif of the studied DnaQ and UvrC homologs. Sequences were aligned using the Clustalω tool on the MPI Bioinformatic Toolkit (https://toolkit.tuebingen.mpg.de/tools/clustalo). The type of amino acids is indicated with shades.

3-2-6- Mycobacterial DnaQ and DnaQ-UvrC exhibit similar 3D structures in *Msm* and *Mtb*:

This study used *Msm* as mycobacterial model organism, necessitating caution in extrapolating any observations to the pathogenic relative, *Mtb*. To ascertain the degree of homology between the *Msm* and *Mtb* proteins, the 3D structures of the identified domains were compared. Because the structures of the *Msm* proteins have not been solved experimentally, Alphafold was used as predictive tool. As shown in Figure XVII, the 3'–5' proofreading exonuclease, the GIYYIG endonuclease, the UvrC-mid and the BRCT domains were predicted to adopt very similar 3D structures in both organisms.





Protein names	Protein Domains						
	Exonuclease Domain			B-clamp binding motif	GIYYIG Endonuclease Domain	UVR Binding motif	RuvA domain
	Exo I	Exo II	Exo III				
DnaQ	12-17 D*ET*G	98-103 HN**FD	162-168 H****D	169-174 QILAEV	Na	Na	Na
UvrC	Na	Na	Na	Na	19-34 GVYYVG	204-239 DQVL...VRRV	538-649 AVIG...FALG
MSMEG_6275	28-33 D*ET*G	107-112 HN**FD	166-172 H****D	333-337 QYALF	Na	Na	Na
MSMEG_4259	80-85 D*ET*G	167-172 HN**FD	227-233 H****D	64-70 QLSLDME	251-266 GVYYVG	463-499 ADSA...VAID	Na
MSMEG_3078	Na	Na	Na	Na	45-60 GVYYVG	235-268 RLAK...ALKR	519-638 DELG...DALR
Rv_3711c	20-25 D*ET*G	99-104 HN**FD	158-164 H****D	325-329 QLALF	Na	Na	Na
Rv_2191	42-59 D*ET*G	132-137 HN**FD	194-199 H****D	16-23 QLSFADLA	285-300 GVYYVG	434-469 NAAL...IEAL	Na
Rv_1420	Na	Na	Na	Na	19-34 GVYYVG	208-243 DRFA...PRKS	521-606 PPDS...KIFW

Figure XVI- Map and composition of the domain identified. *E. coli* = *Escherichia coli*; *Msm* = *Mycobacterium smegmatis*; *Mtb* = *Mycobacterium tuberculosis*. Length of the protein is indicated below the name of the corresponding protein. * = rand om unconserved amino acids.

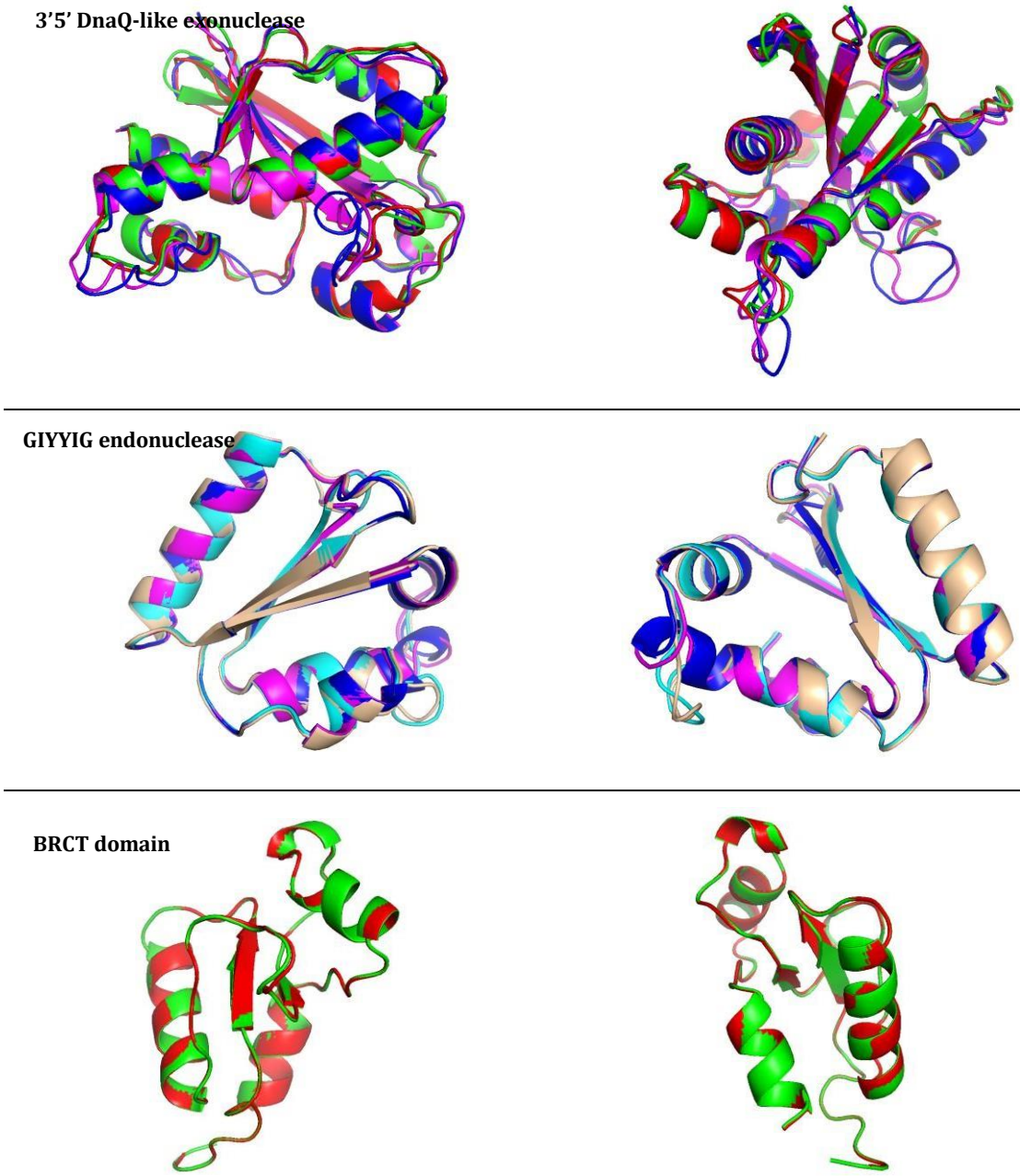


Figure XVII- Alignment of domains 3D structures. Top panel: 3’5’ DnaQ-like exonuclease. Middle panel: GIYIG endonuclease. Bottom panel: BRCT domain. 3D structures were downloaded from Alphafold (https://alphafold.ebi.ac.uk/entry/P9WQK7. They were then colored and labelled using PyMol software(https://pymol.org). The signification of the colours is indicated in the legend. The organisms and associated proteins are identified as following: DnaQ = MSMEG_6275 ■, Rv_3711c ■. DnaQ-UvrC = MSMEG_4259 ■, Rv2191 ■.UvrC = MSMEG_3078 ■,Rv1420 ■.

3-2-7- Mycobacterial-like DnaQ cluster is easily separated from *E. coli*:

Over 3500 DnaQ homologs were identified in 945 bacterial genomes, and they were all submitted to an all-against-all BLAST+ similarities and subsequently clustered based on full protein sequences. Given the propensity for error in online annotations, the results provided by the clustering must be cautiously interpreted. However, the goal was to try to identify homologies of mycobacterial DnaQ (MSMEG_6275/ Rv3711c), and DnaQ-UvrC (MSMEG_4259/ Rv2191) proteins in other organisms, and to use literature about their roles and functions to help predict the functions of the mycobacterial DnaQ and DnaQ-UvrC proteins.

The identified homologs which returned an HSP (High-scoring Segment Pair) below e^{-30} were considered a good hit for homology matches.[95] About five main clusters can be identified in Figure XVIII. On the bottom right, three DnaQ clusters can be observed. The mycobacterial-like DnaQ cluster was easily separated from *E. coli* based on the presence of the BRCT domain and the unusual β clamp-binding motif location. The last DnaQ cluster contained proteins with a truncated 3'-5' DnaQ-like exonuclease and no BRCT domains.

The DnaQ-UvrC-like cluster was very close to the UvrC-like cluster, while some proteins were equidistant from both so could not be included as a part of a cluster. The last cluster, close to UvrC'-like and DnaQ-UvrC-like, contained PolC proteins owing to the presence of an integrated nuclease that is similar to that found in UvrC.

The mycobacterial DnaQ cluster was relatively small (only 68 sequences in 59 genomes) and heterogeneous in terms of architecture (Figure XIX). Only 34 genomes of the 59 exhibited both a β clamp-binding motif and a BRCT domain. Interestingly, they were all part of the Actinobacteria class. The other closest groups had no β clamp-binding motif and showed variations in BRCT domain (either truncated or two BRCT domains). They mostly belonged to *Streptomyces* sp. The most divergent clusters had either a slightly different 3'–5' proofreading exonuclease or a different nuclease, TerB. As a comparison, over 350 genomes exhibited an *E. coli*-like DnaQ, and almost all of these were proteobacteria.

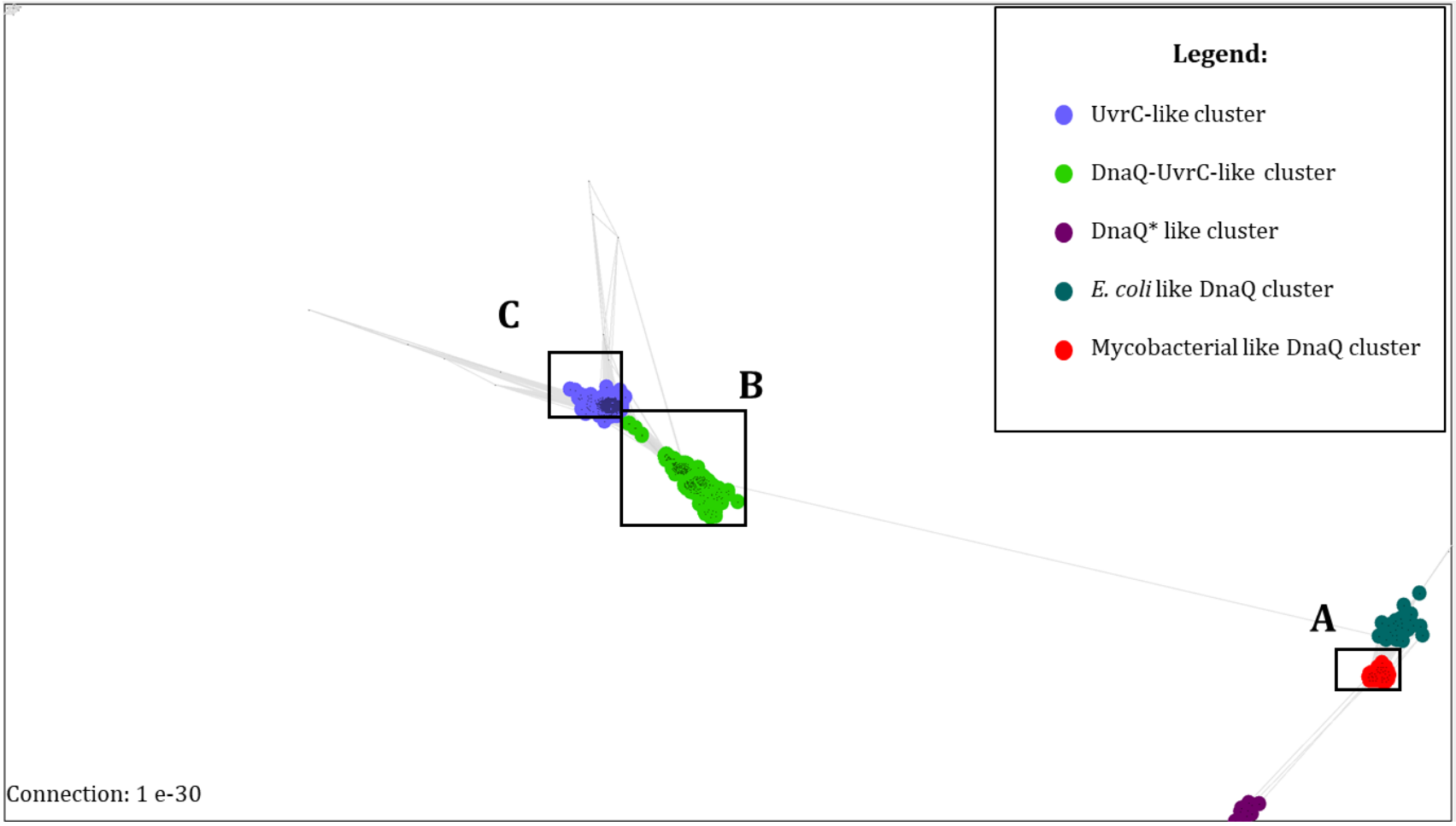


Figure XVIII- Cluster map of DnaQ homologs. A-Protein cluster used for subsequent of DnaQ -like homologs clustering. B- Protein cluster used for subsequent of DnaQ-UvrC -like homologs clustering. C- Protein cluster used for subsequent of UvrC -like homologs clustering. Protein sequences were clustered based on their full protein sequences using CLANS program on the MPI Bioinformatics toolkit (https://toolkit.tuebingen.mpg.de/tools/clans). Clusters of interest are indicated by black squares and color shades described in the legend. Geninfo identifier numbers of protein are in the Supplementary Table IV. All against all blast+ files can be found at https://drive.google.com/drive/folders/1CltC3y3rykfHKPrsX-b2bBaKh2FJKTH9?usp=drive_link.

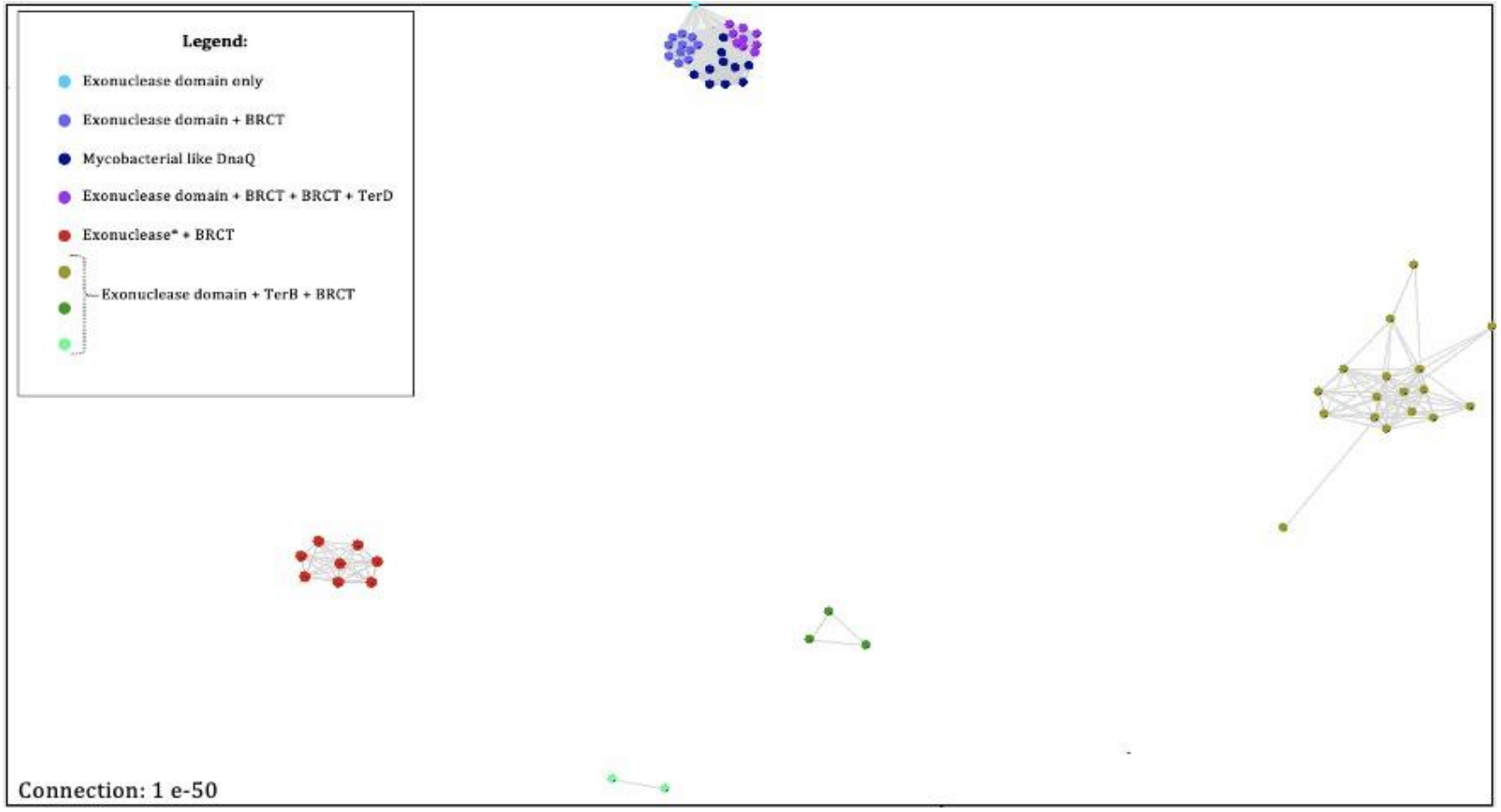


Figure XIX- Cluster map of DnaQ -like homologs. Protein sequences were clustered based on their full protein sequences using CLANS program on the MPI Bioinfor matics toolkit (https://toolkit.tuebingen.mpg.de/tools/clans). Clusters of interest are indicated by black squares and colo ur shades described in the legend. Geninfo identifier numbers of protein are in the Supplementary Table IV. All against all blast+ files can be found at https://drive.google.com/drive/folders/1CltC3y3rykfHKPrsX-b2bBaKh2FJKTH9?usp=drive_link.

3-2-8- DnaQ-UvrC is specific to Mycobacteria:

The DnaQ-UvrC-like cluster consists of 184 different proteins that are diverse in terms of architecture. The mycobacterial-like DnaQ-UvrC was identified as a part of a 68 protein cluster characterized by the presence of an exonuclease domain, a GIY-YIG nuclease domain, a β clamp domain and a UVR binding motif (Figure XX). This cluster only appeared in Actinobacterial genomes. The largest cluster represented a hundred genomes and exhibited shorter C-terminal, with either missing β-clamp domain or UVR binding motifs and they were dispersed across the bacterial organisms. The furthest cluster was the most different as it contained no parts of UvrC.

Only 17 of the genomes studied exhibited both mycobacterial-like DnaQ and DnaQ-UvrC and there were almost all narrowed to the *Mycobacterium* genus; on the other hand, the highly conserved UvrC protein could be detected in 96% of the genomes studied.

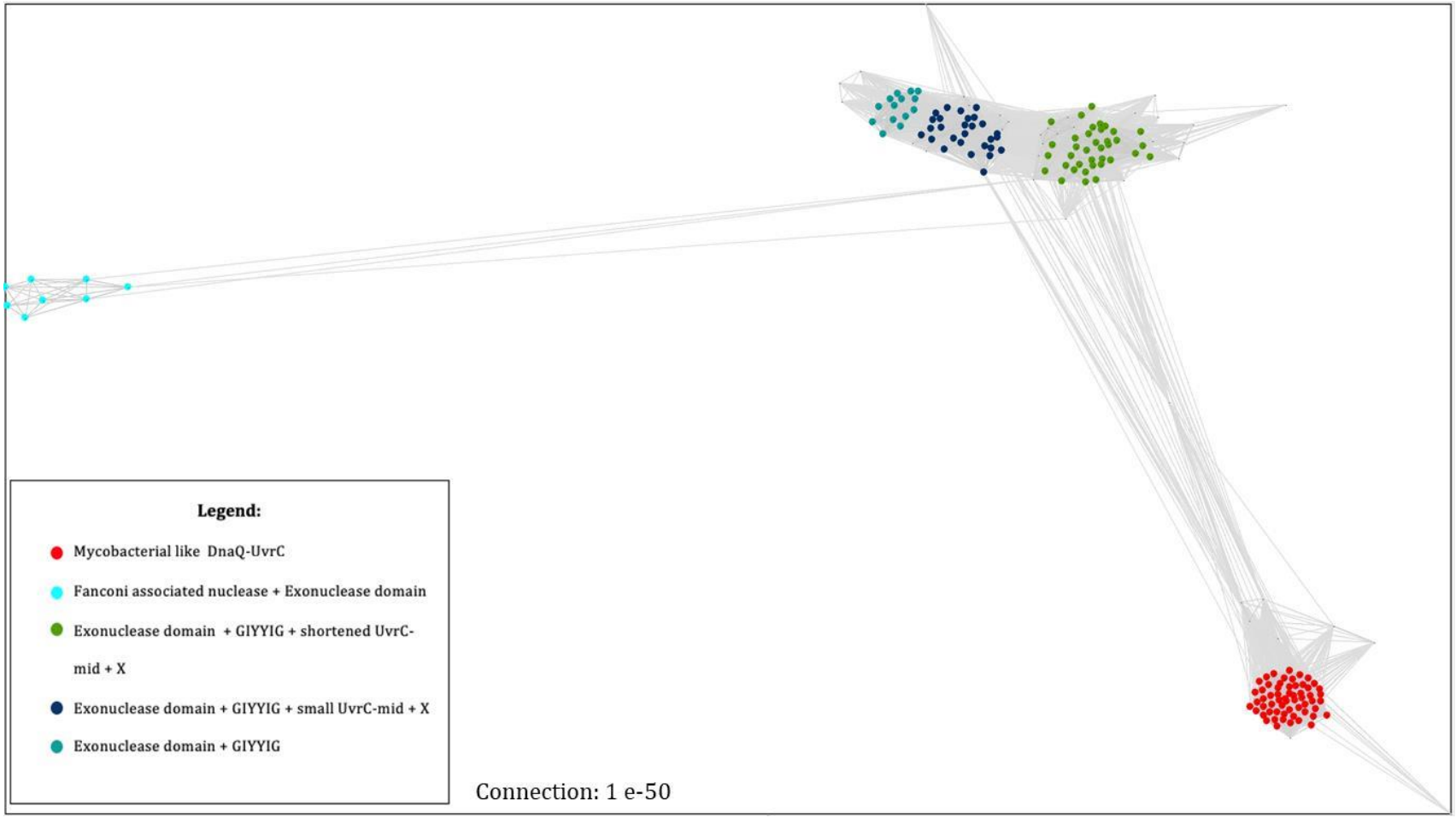


Figure XX- Cluster map of DnaQ-UvrC-like homologs. Protein sequences were clustered based on their full protein sequences using CLANS program on the MPI Bioinformatics toolkit (https://toolkit.tuebingen.mpg.de/tools/clans). Clusters of interest are indicated by black squares and color shades described in the legend. Geninfo identifier numbers of protein are in the Supplementary Table IV. All against all blast+ files can be found at https://drive.google.com/drive/folders/1CltC3y3rykfHKPrsX-b2bBaKh2FJKTH9?usp=drive_link.

3-3- Discussion:

3-3-1- Mycobacterial DnaQ homologs exhibit functional nucleases:

A DnaQ-like 3'–5' proofreading exonuclease domain was identified in both *Msm* (MSMEG_6275 and MSMEG 4259) and *Mtb* (Rv_3711c and Rv_2191). When compared to their *E. coli* counterpart, they are rather well conserved and thus, expected to be operational *in vivo*. However, it was demonstrated at multiple times that the DnaQ-like 3' – 5' proofreading exonuclease had no or low functional proofreading role *in vitro*.[31,41] The other GIYYIG endonuclease – found in MSMEG_6275 and Rv_3711c – is rather well conserved too and thus, also predicted to be functional *in vivo,* although, this has never been studied neither *in vivo* nor *in vitro*.

3-3-2- Mycobacteria displays different interaction sites compared to the model organism:

Mycobacterial DnaQ homologs all contain a β clamp-binding motif; however, its location differs from the model organism: while it is located following the exonuclease domain in *E. coli* DnaQ, the motif is located at the very N-terminal for MSMEG_6275 and Rv_3711c, and at the very C-terminal for MSMEG 4259 and Rv_2191. This unusual localization of a β-clamp binding motif combined with the absence of a DnaE1 interaction site might affect the way mycobacterial DnaQ homologs access DNA. Indeed, these differences might affect their ability to reach the DNA and thus, have an impact on their ability to perform the exonuclease activity. Moreover, the presence of a BRCT domain that does not exist in *E. coli* DnaQ, should allow MSMEG_6275 and Rv_3711c to form different interactions and consequently, they are likely to have different functions in the cell. Interestingly, BRCT domains are known to form homodimers with complementary BRCT domains. In mycobacteria, such domains can be found in helicases.[94]

3-3-3- MSMEG 6275 and Rv 3711c are potential alternatives to UvrC in Nucleotide Excision Repair:

Both MSMEG 4259 and Rv2191 possess in their respective UvrC-mid domain, the required binding motif to interact with the replisome and to be a part of the NER. The existence of this domain has been reported before in UvrB and UvrC and they are thought to interact through this motif. Indeed, Sohi et al[96] showed that a UVR motif dimer forms a

stable complex. Since MSMEG_4259 and Rv2191 exhibit both interaction sites, they are likely to be able to bind either -or both- the β-clamp or UvrB. Moreover, they possess both an exo- and an endonuclease. Thus, they may be involved in a proofreading pathway thanks to their exonuclease and β-clamp binding site, but it may also be involved in the NER pathway thanks to their endonuclease and UVR motif.

This raises a natural question: is UvrC even necessary in mycobacteria? Or are the different UvrC homologs part of different specific DNA-damage repair pathways? The fact that MSMEG_4259 and Rv_2191exhibit two kind of nucleases whereas UvrC proteins have two endonucleases seem to point the second option and will be explored in Chapter V.

3-3-4- Mycobacterial DnaQ and DnaQ-UvrC key domains exhibits similar 3D structure in *Msm* and *Mtb*:

Msm is non-pathogenic model of *Mtb*, yet, this model has a limit as some aspects of *Msm* and *Mtb* are very different from one another.[42] For example, *Msm* generation time is about 4 hours whereas *Mtb* is about 24 hours. Thus, it is important to make sure that the work performed in the model organism can be carried over in *Mtb*. The sequences and the protein domains identified in *Msm* (Figure XVI) were all conserved in *Mtb*, moreover, it was noticed that the exonuclease domain of DnaQ and DnaQ-UvrC, the UvrC part (GIYYIG and UvrC-mid domains) of DnaQ-UvrC and UvrC, the BRCT domain of DnaQ and the RuvA_2-like domains of UvrC had very similar 3D structures in both organism giving assurance that the worked performed in this study could be transposed in *Mtb* (Figure XVII).

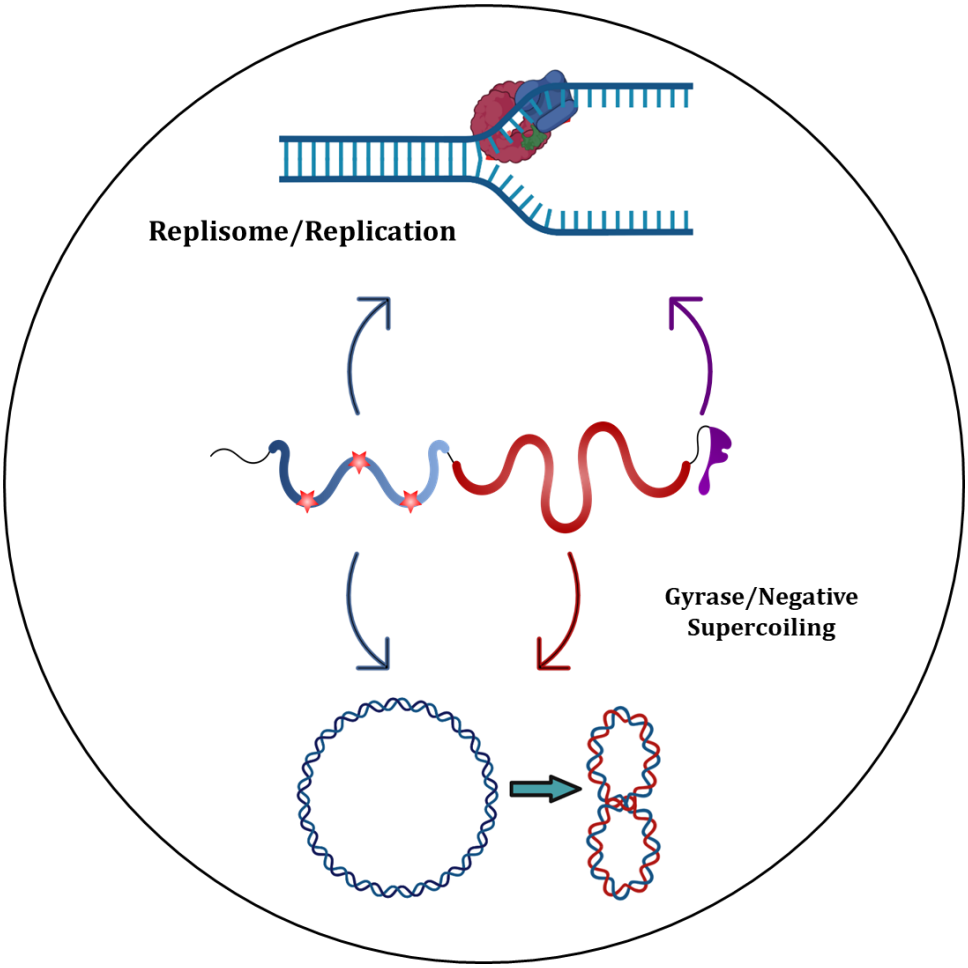
3-3-5- Mycobacterial DnaQ and DnaQ-UvrC are not widespread among bacteria:

It was first hoped that a homolog of either DnaQ (MSMEG_6275, Rv_3711c) or DnaQ-UvrC (MSMEG_4259 and Rv_2191) could be identified outside of Mycobacteria with a known function that could help elucidate their role. Interestingly, only Actinobacterial genomes display proteins with similar domains as they were previously identified in *Msm* and *Mtb*. Furthermore, organisms displaying at the same time both DnaQ and DnaQ-UvrC were all narrowed down to the *Mycobacterium* genus, on the other hand, a conserved UvrC

protein could be detected in 96% of the genomes studied (Figure XVIII to XX). What functions, that are almost exclusive to Mycobacteria, may DnaQ and DnaQ-UvrC could be involved in?

CHAPTER IV:

DnaQ is involved in gyrase pathway rather than replication fidelity:



Created with BioRender

4-1- Introduction:

4-1-1- Bacterial DNA Replication:

DNA replication is a process that occurs in all living organisms.[96] In bacteria, the action of a complex, multi-protein DNA replication machine – or replisome[84] – ensures the high-fidelity production of two identical copies of the original DNA molecule that are separated between the mother and (new) daughter cells, thereby enabling the survival and perpetuation of the species. Because of its critical importance, DNA replication is a highly regulated, multistep process.[26] For obligate pathogens such as *Mtb* which encounter hostile environmental conditions during host infection, tight regulation of the cellular machinery is essential to maintain genomic integrity[97] while allowing scope for adaptation.[98]

As for many other aspects of macromolecular synthesis, there are many components of the DNA replication machinery in mycobacteria which remain poorly understood. Therefore, current knowledge relies heavily on inferred (even assumed) analogy to better study model bacterial systems. In *E. coli*, DNA replication is performed by a multiprotein replisome that is responsible for numerous steps including double-stranded (ds) DNA unwinding, RNA primer synthesis, clamp loading, and DNA synthesis that enable copying of the single, circular bacterial chromosome. The replisome components are mostly conserved across bacteria, and can be divided into three broad complexes: (i) the helicase/primase complex – a DnaB helicase unwinds the dsDNA and a DnaG primase is loaded by a multi-subunit helicase loader onto the DNA lagging strand to synthesize the short RNA primer to initiate DNA replication; (ii) the core complex – which is made up of the DNA polymerase alpha subunit (Pol IIIα), the exonuclease ε, and the small subunit, θ. Two core complexes synthesize the new DNA on both leading and lagging strand templates. The core complex binds to the β_2 sliding-clamp that holds the core together and prevents it from dissociating from the template DNA,[29] as well as playing an essential role in replication processivity (discussed later); and (iii) the clamp-loader complex – which comprises three copies of τ subunits (encoded by *dnaX*) which form a heteropentamer wih δ *and* δ' subunits (encoded by *holA* and *holB* respectively). The $\tau_3\delta_1$ δ'_1 pentamer has an ATP-dependent activity that is responsible for the loading of the β_2 sliding clamp onto the β clamp onto newly synthesized RNA primers and guides DNA

single-strand binding (SSB) proteins onto the unwound DNA strands.[84–86,99,100] The, χ and ψ subunits serve as a bridge between the clamp-loader complex and the ssDNA.

4-1-2- Mycobacterial DNA replication:

Because mycobacteria lack many of the replisome's components compared to *E. coli*, it contributed to the definition of a simple (or minimal) bacterial replication module. Indeed, while *E. coli* requires the participation of 13 different proteins, mycobacteria only need 10(Figure XXI) comprising the replication initiator protein (*dnaA*), the helicase (*dnaB*), the primase (*dnaG*), Pol lllα (*dnaE1*), the *dnaN*-encoded β_2 sliding-clamp, the ε proofreading exonuclease, τ, δ_1, δ'_1, SSB, DNA ligase and DNA Pol I, which is required to remove the RNA primers.[29,41]

Replication of the 4.4-Mb *Mtb* chromosome is initiated by a DnaA-dependent primosome[101] which assembles at the origin of replication, *oriC*. The replication complex disassembles at the termination sequence, *ter*, located at the opposite side of the circular chromosome once replication is completed. The *oriC* region contains multiple binding sites that are recognised by DnaA which can oligomerise quickly owing to its intrinsic ATPase activity.[101] Thereafter, the hexameric DnaB replicative helicase and the DnaG primase are loaded, forming with DnaA the pre-replication complex, leading to DNA strand separation, early unwinding of the parental chromosomal DNA, and short RNA primer synthesis for lagging-strand replication.[29] Next, the mycobacterial replisome is loaded onto each of the two replication forks for template-directed DNA synthesis. The mycobacterial replisome comprises a trimeric Pol III replicase core made up of the PolIIIα polymerase (DnaE1), the 3'-5' exonuclease subunit, ε (DnaQ), and the β_2 sliding clamp (comprising two DnaN subunits). The interaction between the replisome and helicase is mediated by the clamp loader complex, comprising *holA*- and *holB*- encoded subunits.

While chromosomal mutations enable adaptation of the bacillus to host infection and allow for antibiotic resistance, uncontrolled mutagenesis could lead to error catastrophe and death.[102,103] There is consequently a competition between high-fidelity DNA maintenance and mutagenesis. This balance between genetic integrity and adaptation via mutagenesis, which will be discussed through this work, must be carefully regulated.

Mtb is not a natural mutator;[40,104] instead, the mycobacterial mutation rate compares favourably with other bacterial systems such as *E. coli*. Current knowledge suggests that proofreading (exonuclease) function is the main mechanism to maintain replication fidelity in *Mtb*, and this function is ensured by the replisome.[31] *Mtb* lacks a canonical mismatch repair (MMR) system,[105] however the recent identification of an archaeal mismatch specific endonuclease suggests the possibility of non-canonical MMR, mediated by NucS.[106] Moreover, multiple lines of evidence[105,107–109] suggest that alternative repair mechanism such as NER[109] also contribute to genome fidelity.

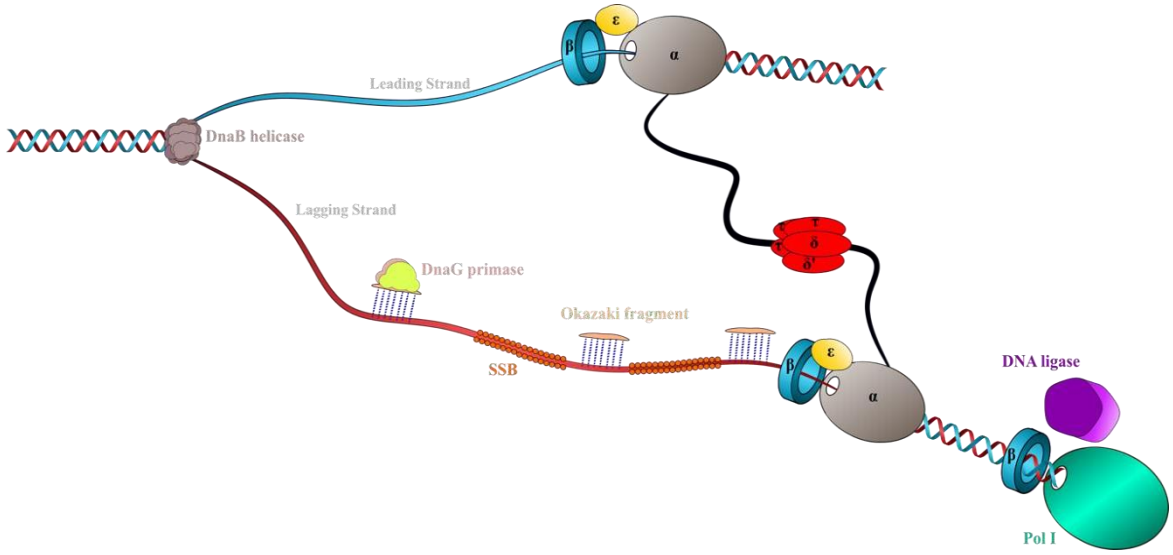


Figure XXI- The mycobacterial replisome: Schematic representation of the model replisome consisting of the PolIII core polymerase, the homodimeric β2-sliding clamp, the τ3δ1δ1′ clamp-loader complex, DnaB helicase (red hexamer), DnaG primase (blue), PolI (pink) DNA ligase (purple), and SSB (orange). Recent biochemical evidence suggests that, in *M. tuberculosis*, the ε proofreader forms part of the core replicase together with the β2 and PolIIIα subunits.[30,85,86,110] As noted in the main text, the precise stoichiometry and architecture of the mycobacterial replisome remain to be established; similarly, it is not known whether the mycobacterial replisome functions as a di- or tri- polymerase system, nor whether DnaE2 is able to access the replisome under non-DNA-damaging conditions in the absence of ImuB and ImuA′ accessory factors. Adapted from Ditse et al.[41]

4-1-3- Replication fidelity:

In *E. coli*, replication fidelity is maintained by three principal mechanisms: nucleotide insertion fidelity, the removal of misincorporated nucleotides, and post-replicative MMR.[111] Most bacterial genomes encode only one high-fidelity replicative C-

family DNA polymerase.[29,110,112,113] This is also the case in *Mtb* which encodes DnaE1, an essential replicative polymerase comprising four highly conserved domains: (i) Polymerase and Histidinol Phosphatase (PHP) domain which contains proofreading activity; (ii) Palm domain, which carries one of the catalytic carboxylates residue for template elongation; (iii) Thumb domain, which binds the DNA backbone and, together with the (iv) Fingers domain forms an oligonucleotide binding (OB) domain and a β clamp-binding domain which is critical for protein-protein interactions within the replisome.[92,114,115]

The PHP domains of some bacterial replicases, including *Mtb*, contain active metal-dependent nucleases that play a role in proofreading. In other bacterial systems, such as *E. coli* DNA polymerase III, the PHP proofreading domain has lost several metal-coordinating residues and is likely to be catalytically inactive. The domain nevertheless fulfils a major role in the Pol III core as it modulates the stability and the activity of the polymerase.[92,114,116] The removal of misincorporated nucleotides in *E. coli* requires a separate 3-5' proofreading exonuclease, the *dnaQ*-encoded ε subunit. Previous work[117–119] showed that a defective ε subunit was not able to remove misincorporated nucleotides, elevating rates of mutagenesis.

Like *E. coli*, *Mtb* possesses a separate 3-5' exonuclease subunit, encoded by Rv3711c, in addition to its high-fidelity replicative DnaE1 polymerase. There have only been a few studies of mycobacterial DnaQ but all highlighted critical departures from the *E. coli* model. For example, deletion of *dnaQ* in *Mtb* and in *Msm* did not result in a significant increase in the mutation rate,[30,31] even though both mycobacterial orthologs have a functional exonuclease domain and exhibit functional exonuclease activity *in vitro.*[28] Therefore, prior to the elucidation of the role of the DnaE1 PHP domain in proofreading, it was hypothesized that replication fidelity in *Mtb* was mediated by an alternative exonuclease.[33,91,115] In addition to DnaQ/Rv3711c, *Mtb* encodes a hybrid protein, Rv2191, containing an N-terminal DnaQ exonuclease domain fused to a C-terminal UvrC-like endonuclease domain. Similar to Rv3711c, deletion of Rv2191 did not result in an increased mutation rate,[30,31,41] despite biochemical confirmation of functional exonuclease activity *in vitro*[28]. Since both *Msm* and *Mtb* possess separate homologs of canonical UvrC and DnaQ proteins, the existence of the additional "DnaQ-UvrC" fusion protein is intriguing and seems to be a specific feature of mycobacteria – as will be

discussed further in Chapter V. Interestingly, in the basic replication model encoded in *M. leprae*, in some ways considered a "minimal mycobacterial genome", these genes are not present as functional genes but rather as pseudogenes and, therefore, might be dispensable for replication fidelity.[120]

Rock and colleagues[66] showed that the PHP domain of *Mtb* DnaE1 possesses an active metal-dependent nuclease that has been lost in *E. coli.* Their results, together with the absence of a mutator phenotype in DnaQ-deficient mycobacterial mutants, suggest that replication fidelity in *Mtb* is mostly determined by DnaE1. Consistent with the interpretation that the PHP domain of DnaE1[25–27] is the major replicative exonuclease, targeted mutations of the PHP metal-coordinating residues resulted in severe growth defects and significantly elevated mutation rates. Subsequently, Gu et al.[33] were able to reconstitute a functional *Mtb* DNA Pol III holoenzyme *in vitro*. They also showed that, whereas the *in vivo* role of DnaQ remains cryptic, its catalytic domain is functional. Moreover, the combination of the two exonucleases appeared more active than the individual proteins *in vitro*, suggesting that the interaction between DnaE1 and DnaQ might be important for replication fidelity.

The same study also elucidated the critical role of the β clamp. As in the *E. coli* model, the *Mtb* DNA PolIII replicase is able to switch between polymerisation and proofreading modes. This transition is mediated by the β clamp, which can regulate interactions between the different subunits within the PolIII core.[92,112,115] The β clamp promotes polymerisation and reduces the exonuclease activity when the conditions are suited to DNA polymerisation. In contrast, when dNTPs are limiting, the β clamp promotes exonuclease activity. However, this regulation is not yet understood in *Mtb*.[120]

Studies in *E. coli* showed that the exonuclease subunit is attached to the polymerase subunit by a long, flexible tether[114], and that both subunits interact with the β clamp. This tether allows the polymerase core to undergo conformational changes that affect the polymerase activity from polymerisation to proof-reading. It also allows β to regulate access of other polymerases to the replication fork.[121] Unlike the *E. coli* model, the mycobacterial ε subunit does not directly associate with α. The β clamp acts as a bridge between α and ε subunits, stabilizing the PolIII core, and is therefore necessary to maintain both replication fidelity and processivity. [33,91,115]

It has also been hypothesized that the β clamp can regulate access of other polymerases to the DNA. For example, when *Mtb* is challenged by antibiotic treatment or UV light, double-stranded DNA breaks (DSBs) can occur. Most DNA-dependent polymerases (such as DNA pol III) can copy and replicate the genome with a high degree of fidelity.[32] However, these polymerases are generally unable to synthesize DNA over damaged template strands, especially DSBs.[102] Since DSBs would usually prove fatal to the cell, repair systems that allow the DNA to be repaired are critical to survival. For pathogens such as mycobacteria, mechanisms enabling survival under DNA damaging conditions might be fundamental to host parasitism.[98,122]

In this chapter, the roles of DnaQ (*Mtb* Rv3711c; *Msm* MSMEG_6275) and DnaQ-UvrC (*Mtb* Rv2191; *Msm* MSMEG_4259) in maintaining replication fidelity will be investigated. Although the role of DnaQ (*Mtb* Rv3711c; *Msm* MSMEG_6275) in mycobacterial replication has already been studied, the individual contribution of both DnaQ homologs has not been investigated, nor any potential synergy between them (Table III Section2-1). Similarly, the contributions of the proteins to maintenance of genome integrity following UV exposure has also not been investigated.

4-2- Results:

4-2-1- Deletion of MSMEG_6275 has no effect on the mutation frequency:

Mtb dnaQ is highly polymorphic and has been identified as a target of independent mutations associated with drug resistance and as a gene under strong selective pressure;[34,35] however, the function of *Mtb* DnaQ remains enigmatic. Separate studies[40,41] showed that DnaQ was dispensable for DNA damage-induced mutagenesis and damage tolerance in *Msm* (MSMEG_6275) and *Mtb* (Rv3711c), respectively; in contrast, the DnaQ-UvrC hybrid (Rv2191 / MSMEG_4259) was required for DNA damage survival.[30] To confirm these observations, an unmarked MSMEG_6275 deletion was generated (Table III Section 2-1) together with a double Δ*dnaQ* Δ*dnaQ-uvrC* knock-out. To confirm the deletion genotype, a PCR amplifying the flanking regions of genes was performed (Figure XXII). For further microscopic and proteomic analyses, fluorescence-tagged DnaQ-GFP and DnaQ-UvrC translational fusion strains were also generated. However, the planned localization and interaction experiments could not be carried out owing to time constraints consequent on the COVID-19 pandemic.

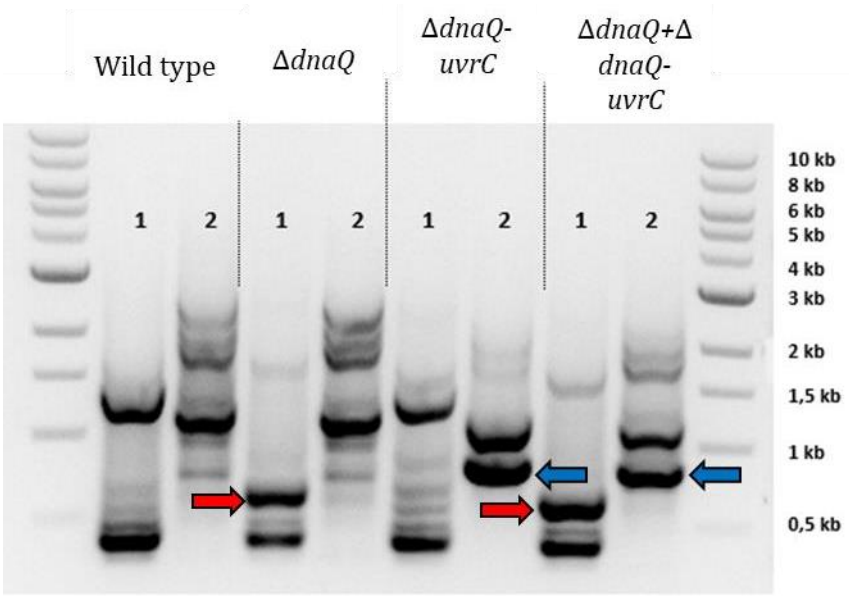


Figure XXII- PCR confirmation of *dnaQ*, *dnaQ-uvrC* and double *dnaQ dnaQ-uvrC* deletion mutants. Primers were used to amplify the 3' of the gene along with the downstream flanking sequence. Lanes 1 correspond to *dnaQ* amplification, intact gene should exhibit a 1.2kb band whereas the deleted version should exhibit a 0.6kb band (red arrow). Lane 2 corresponds to *dnaQ-uvrC* amplification, intact gene should exhibit a 1.1kb band whereas the deleted version should exhibit a 0.8 kb band (blue arrow).

For mutation rate determination, UV exposure was employed as described previously.[68,102,123,124] The mutation rate was estimated using an adaptation of the Luria-Delbrück fluctuation assay,[121] with the acquisition of rifampicin (RIF) resistance as the selective phenotype. The fluctuation assay set-up strictly requires that the bacteria are exposed to the mutagenic agent (UV or antibiotic) constantly during the experiment; therefore, the use of a once-off/transient perturbation such as UV treatment contravenes a core principle of the fluctuation assay. Since, in these experiments, the bacteria were exposed to UV at a single time point (as described in section 2.8.4.) they will be referred to as "mutation frequency" assays. For this purpose, we utilized 3 parallel cultures of each strain in each repeat of the assay, and the number of RIF^R mutants was enumerated by plating on RIF-containing solid medium. Another 3 parallel cultures of each strain were also plated on medium with no drug to verify the bacteria survival and ascertain that the observed Rif resistance was not due a difference in strains survival to UV.

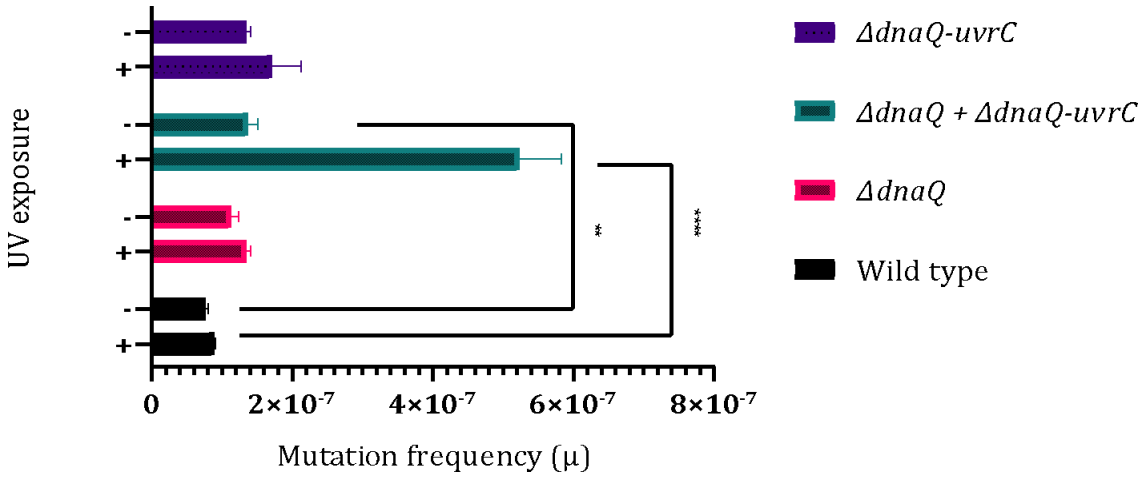


Figure XXIII- Mutation frequency assays. Mutation frequencies were calculated with the software developed in the MMRU by Z. Martin and J. Kent[59] based on the results of three independent experiments. A Sidak-s multiple comparisons test was used to compare means variation between strains in both conditions. Individual p-values of 95% confidence intervals are displayed as stars (<0.05= *; <0.005=**; <0.0005=***; <0.0001=****. Na = Not attributed.

The mutation frequencies were determined for each single-mutant strain relative to wild type. As shown in Figure XXIII and Table X, there was no significant difference in the mutation frequency without UV exposure. However, the double *dnaQ dnaQ-uvrC* knock-

out strain exhibited an increased mutation frequency before and after UV exposure (Figure XXIII and Table X), suggesting a potential combined activity of both DnaQ and DnaQ-UvrC in maintaining replication fidelity, especially following UV-induced DNA damage.[125] To confirm that the variation in mutation frequency data was not a consequence of bacterial death, a fraction of bacteria was plated on no drug plates (Table XI). Similar survival rates were observed for all strains, eliminating hypersusceptibility as a confounder in the calculation of mutation frequencies.

Strains comparison	wild type VS *ΔdnaQ*	wild type VS *ΔdnaQ +ΔdnaQ-uvrC*	wild type VS *ΔdnaQ-uvrC*	UV exposure
Mean fold-change	2.272	20.05 ****	3.963 **	+
	1.668	2.735 *	2.628	-

Table X- Mean comparison of frequency rates. The fold change in the mutation frequency was calculated by dividing the mutation frequency in the presence of mutant strains by the mutation rate of the wild-type, parental *Msm* mc²155. A Sidak-s multiple comparisons test was used to compare means variation between strains in both conditions. Individual p-values of 95% confidence intervals are displayed as stars (<0.05= *; <0.005=**; <0.0005=***; <0.0001=****. Na = Not attributed.

Strain colony counts	wild type			Δ*dnaQ*			Δ*dnaQ-uvrC*			Δ*dnaQ* + Δ*dnaQ-uvrC*		
Dilution	10^{-5}	10^{-6}	10^{-7}	10^{-5}	10^{-6}	10^{-7}	10^{-5}	10^{-6}	10^{-7}	10^{-5}	10^{-6}	10^{-7}
Pre UV-exposure	239	22	3	234	20	1	198	27	3	185	21	2
Dilution	Neat	10^{-1}	10^{-2}	Neat	10^{-1}	10^{-2}	Neat	10^{-1}	10^{-2}	Neat	10^{-1}	10^{-2}
Post UV-exposure	85	7	0	88	9	1	65	8	0	70	6	0

Table XI- *Msm* survival in the mutation frequency assay - Bacteria were diluted as indicated and plated on 7H10 plates without any drugs.

4-2-2- Replication fidelity relies mainly on DnaE1 proofreading activity:

Previous work from several teams[31,40] showed that replication fidelity in mycobacteria mostly depends on the PHP domain of the DnaE1 polymerase. To determine if DnaQ has any activity in proofreading, a set of strains was generated which carried a Tet-inducible PHP proofreading-defective DnaE1 (pJR650).[66] This mutated DnaE1 lacked a metal ion-coordinating residue, compromising PHP domain function.[31,116] Because *dnaE1* is an essential gene, the Tet-inducible DnaE1 mutant allele was introduced in a wildtype background.[66] In addition, the same mutant *dnaE1* construct was introduced in the *Msm* mutants Δ*dnaQ* and the Δ*dnaQ*Δ*dnaQ-uvrC* double knock-out. As Δ*dnaQ-uvrC* had almost no impact on mutation frequency following UV exposure and as it was further studied in Chapter V, it was no longer part of the Chapter IV experiments.

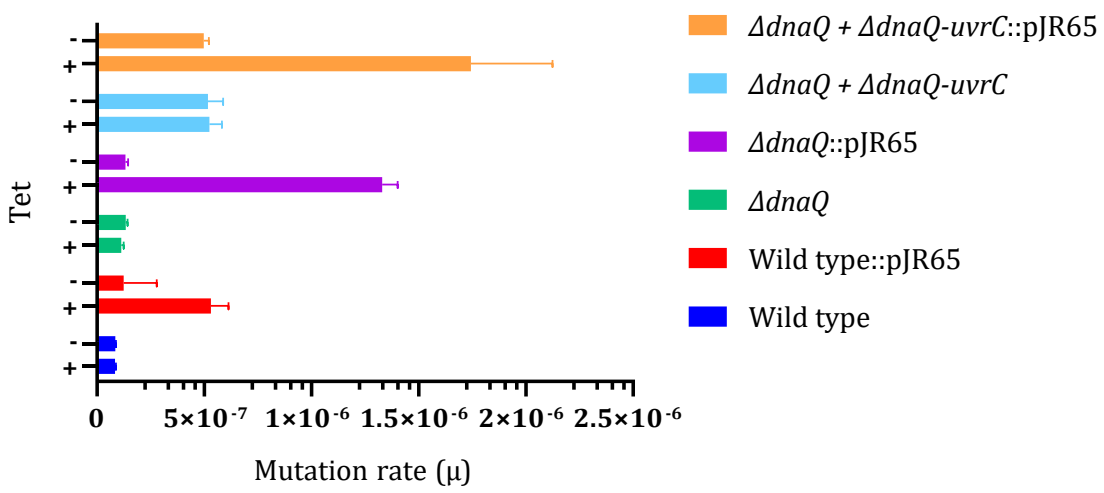


Figure XXIV- Spontaneous mutation frequency. Mutation rate was calculated with the software developed in the MMRU by Z. Martin and J. Kent [59] based on the results of three independent experiments. Experiment was carried out as an adaptation of the Luria Delbruck fluctuation assay as described in the materials and methods.

The spontaneous mutation frequency of all strains were measured in the absence and in the presence of tetracycline (Figure XXIV) Spontaneous mutation frequency implies that no extra-factor was used to induce mutagenesis. Three repeats of the experiment were carried out using 10 parallel cultures of each strain in each repeat. Mutation rates were determined by plating on RIF-containing plates. As it was studied in previous papers,[31,66] the wild type::pJR65 strain was used as a control in these experiments.

Without tetracycline, all strains exhibited similar mutation frequencies. Regarding Δ*dnaQ* and Δ*dnaQ+dnaQ-uvrC* strains, the results were similar to the previous experiment (Figure XXIV) before UV-exposure. Under tetracycline treatment, the mutant *dnaE1* (pJR65 plasmid) had a significant impact on the mutation frequency when compared to wild type (Table XII). This impact, however, was dependent on the genetic background of the strain. When compared to the wildtype, the mutated DnaE1 had an increase in the mutation frequency of about 6.3-fold in the wildtype background, about 15.8-fold in the Δ*dnaQ* background, and about 20-fold in the Δ*dnaQ*Δ*dnaQ-uvrC* background. Interestingly, Δ*dnaQ+dnaQ-uvrC* without the mutated DnaE1 also had an increase under tetracycline treatment. It is possible his might be caused by damage to DNA by tetracycline, but this requires further investigation.[144]

Strains	Wildtype	Wild type::*pJR65*	Δ*dnaQ*	Δ*dnaQ::pJR65*	Δ*dnaQ* + Δ*dnaQ-uvrC*	Δ*dnaQ* + Δ*dnaQ-uvrC :: pJR65*
Wildtype						
Wild type::*pJR65*	6.329 **					
Δ*dnaQ*	1.35	0.214 *				
Δ*dnaQ::pJR65*	15.865 ****	2.506 ****	11.69 ****			
Δ*dnaQ* + Δ*dnaQ-uvrC*	6.24 **	0.98	4.6 *	0.39 *		
Δ*dnaQ* + Δ*dnaQ-uvrC :: pJR65*	19.80 ****	3.25 ****	15.32 ****	1.31 ****	0.3 ****	

Table XII- Mean of fold change in the mutation frequency in the presence of tetracycline. The fold change in the mutation frequency was calculated by dividing the frequency of each strain on the left of the table to the one above. A Sidak-s multiple comparisons test was used to compare means variation between strains in both conditions. Individual p-values of 95% confidence intervals are displayed as stars (<0.05= *; <0.005=**; <0.0005=***; <0.0001=****. "empty" = Not attributed.

4-2-3- Knock-out of *dnaQ* does not result in hypersusceptibility to replisome-targeting drugs:

Identifying drug-hypersusceptibility phenotypes of mutant strains is a way to characterise gene function.[126] So far, no phenotype has been observed for deletion of *dnaQ*. However, most of the work performed to date measured the maintenance of replication fidelity after UV treatment. To address this gap, the susceptibility of the

unmarked *dnaQ* deletion mutant to replisome-targeting drugs was investigated (Table IV Section 2-1). The experiment was performed in three independent biological replicates and used a total of three strains per repeat. Along with the wild-type and Δ*dnaQ* knock-out strains, a complemented strain carrying a *dnaQ-eGFP* translational fusion was used. The drugs that were used targeted either the *dnaN*-encoded β clamp (Griselimycin), DnaE1 polymerase (Nargenicin), or Gyrase B (Novobiocin). As shown in Figure XXV and XXVII, none of these drugs incurred a significant sensitivity phenotype that might provide a clue about DnaQ function. Thus, it was not possible to determine if the complemented strain was able to restore the wild type phenotype. Table XIII confirms that there was no significant difference observed between any strains.

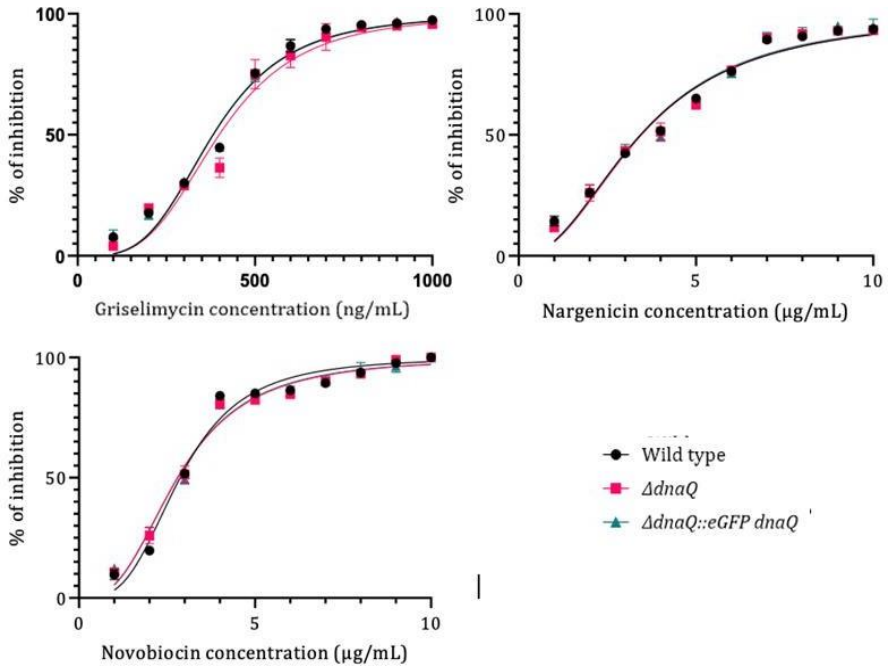


Figure XXV- DnaQ does not exhibit sensitivity to replisome targeting drugs. Data were normalized as a percentage of inhibition is measured relative to wild type data. Individual data are displayed along with fitting curves.

4-2-4- DnaQ activity linked to Gyrase A activity:

Msm and *Mtb* encode a single Type II topoisomerase comprising heterotetramers of GyrA and GyrB subunits.[127] Previously, a GyrB inhibitor (novobiocin) was used which

did not elicit a phenotype in the *dnaQ* mutant relative to wild type. To test the impact of inhibiting GyrA, a fourth-generation fluoroquinolone, moxifloxacin, was tested using the same experimental set-up as previously. As shown in Figure XXVI, the treatment resulted in a slight (around 10%) but significant decrease in the IC_{50} for the *dnaQ* knock-out (Table XIII). However small, this decrease in the IC_{50} may reflect a link between DnaQ and GyrA activity. As fluoroquinolones usually trigger a SOS response, the observed effect might be related to such a response. However, nargenicin and griselimycin are known SOS response inducers,[44,65] and they did not have any effect on the mutant strain (Figure XXVII). Thus, this observed effect appears not to be SOS-dependent but rather comes from targeting GyrA. The complemented strain seems able to restore the wild type phenotype under moxifloxacin treatment.

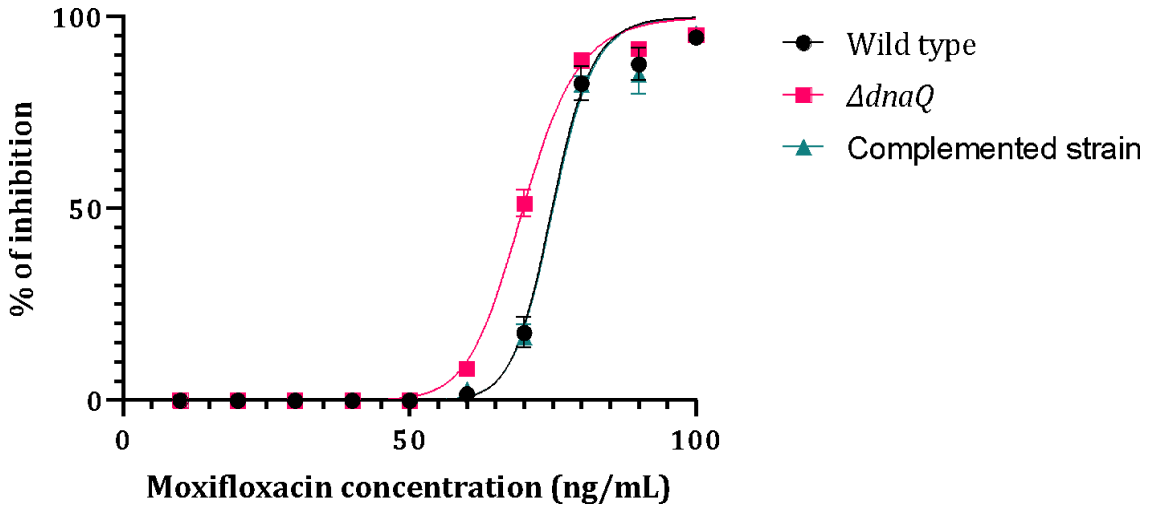


Figure XXVI- Δ*dnaQ* knock-out strain sensitive to Moxifloxacin. Data were normalized as a percentage of inhibition is measured relative to wild type data. Individual data are displayed along with fitting curves. Means were calculated out of three independent biological repeats. Statistical analysis used was a two-way ANOVA at each concentration. Δ*dnaQ* exhibited difference with the wild type at 60 ng/mL (pvalue < 0.0332 = *) and 70ng/mL (pvalue < 0.0002 = ***).

Drugs	Strains	Wild-type	*ΔdnaQ*	*ΔdnaQ::dnaQ-eGFP*
Griselimycin	**IC_{50} (µg/mL)**	382	399	381
	95% Likelihood Interval	[364.5-399.2]	[374.2-423.1]	[364.1-399]
Nargenicin	**IC_{50}**	3.434	3.464	3.439
	95% Likelihood Interval	[3.232-3.634]	[3.253-3.674]	[3.206-3.670]
Novobiocin	**IC_{50}**	2.871	2.805	2.799
	95% Likelihood Interval	[2.732-3.009]	[2.672-2.937]	[2.656-2.939]
Moxifloxacin	**IC_{50}**	74.94	69.98***	75.11
	95% Likelihood Interval	[74.13-75.75]	[69.36-70.43]	[74.20-76.02]

Table XIII- IC_{50} for MMC treatment and 95% likelihood intervals (L.I) of *Msm* Strains. Values shown represent IC_{50} (µg/mL) were generated with a non-linear regression of [inhibitor]VS normalized response and 95% likelihood intervals of experiments Individual p-values of 95% confidence intervals are displayed as stars (<0.05= *; <0.005=**; <0.0005=***; <0.0001=****. "empty" = Not attributed.

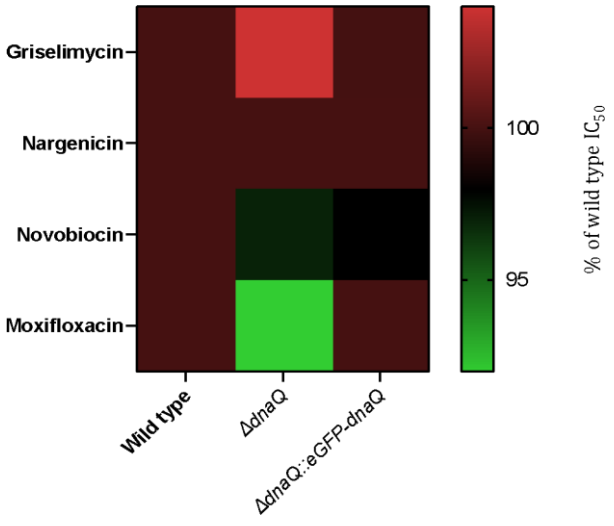


Figure XXVII- IC_{50} of the *Msm* strains for the tested drugs relative to the wild-type strain. The scale gives the correspondence between color and percentage of wild-type IC_{50} for Δ*dnaQ* and, complemented strain. IC_{50} value were calculated using the Nonlinear fitted curves tool using the [inhibitor] vs normalized response — Variable slopes — on Graph Pad Prism.

4-2-5- Deletion of *dnaQ* slows down antibiotic adaptation:

Genome maintenance counterbalances mutagenesis in all organisms including mycobacteria. From comparative genomics analyses, DnaQ was implicated in the emergence of drug resistance in *Mtb.*[34] To investigate this putative association, an antibiotic evolution assay was designed based on the previous work of Ragheb *et al*.[69] Mycobacterial cultures were inoculated into a 96-well plate at the same concentration and then exposed to serial dilutions of drugs. Only strains that grew at least 50% relative to the no drug control at the highest concentration of drugs were passaged to a new 96-well plate to a total of 6 passages. Growth is measured by OD at 600nm. Full assay details are provided in section 2-8-5.

Treatment with MMC generated similar results across strains. Regardless of individual strain sensitivities to the drugs tested, the IC_{50} values remained the same across passages (Figure XXVIII and Table XIV). When treated with Rifampicin, it was noticed that, around the fifth passage, the IC_{50} was increased for all strains except *dnaE2* KO strain. Interestingly, the Δ*dnaQ* strain exhibited a limited increase of the IC_{50} while not preventing it. In the case of moxifloxacin, the adaptation phenotype was present, too, for the wild type and Δ*dnaQ-uvrC*. Similarly to Rifampicin treatment, Δ*dnaQ* was attenuated yet still exhibited an adaptation phenotype.

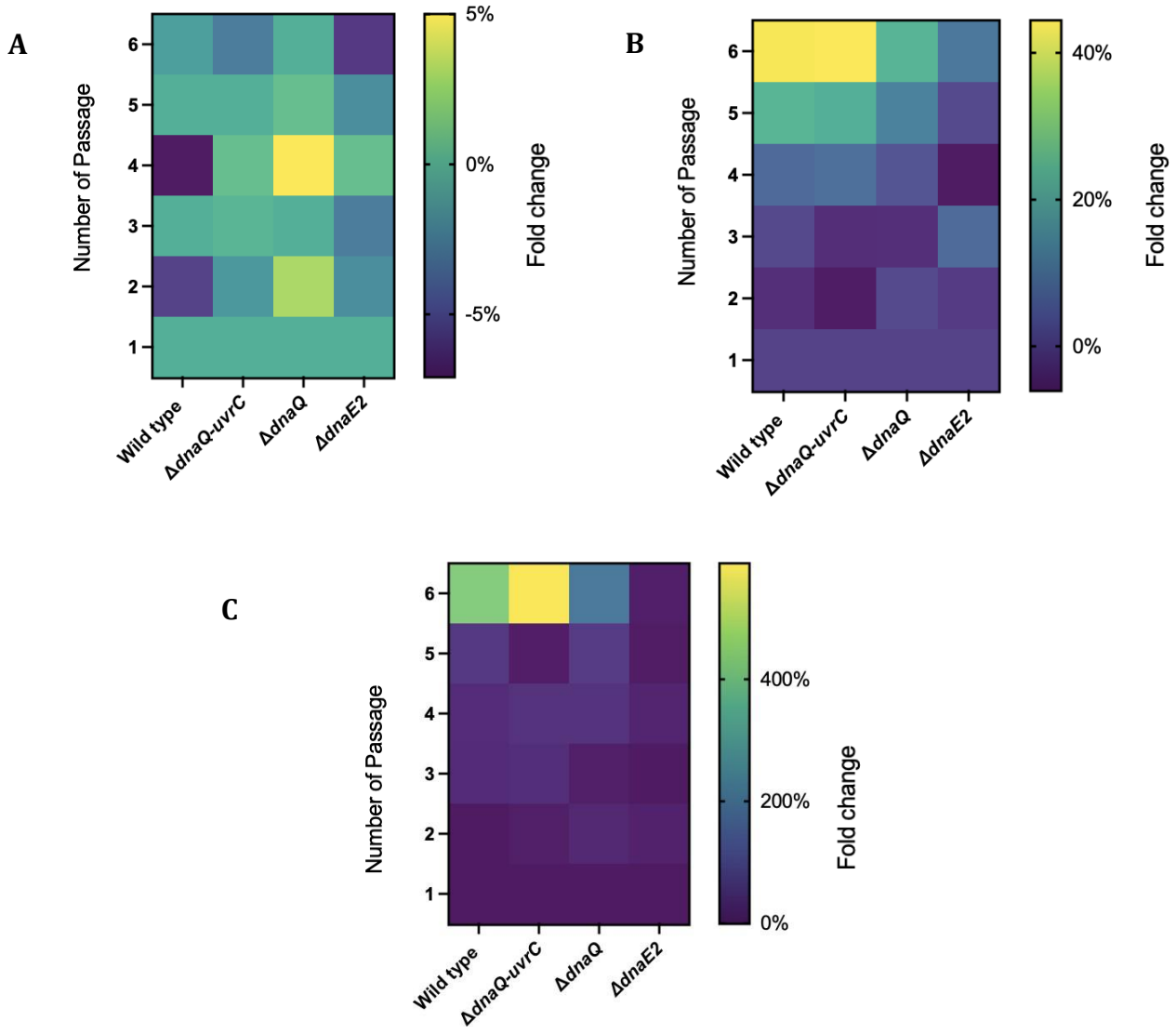


Figure XXVIII- Antibiotic evolution assay- A-Mitomycin C. B-Moxifloxacin. C-Rifampicin. Plot shows IC_{50} for strains at each passage. Statistical significance was determined using a two-tailed Mann-Whitney U test (∗p value < 0.10). Number of biological replicates for each strain of *Msm* is three.

Drugs		**Mitomycin C (ng/mL)**				**Rifampicin (µg/mL)**				**Moxifloxacin (ng/mL)**			
Strains		Wild type	Δ*dnaQ-uvrC*	Δ*dnaQ*	Δ*dnaE2*	Wild type	Δ*dnaQ-uvrC*	Δ*dnaQ*	Δ*dnaE2*	Wild type	Δ*dnaQ-uvrC*	Δ*dnaQ*	Δ*dnaE2*
Passage	**1**	155	75****	149	95****	11.3	10.6	12.2	10.1	70	72	65	82*
	2	148	82****	152	81****	11.2	11.3	15.6	11.5	68	68	66	81**
	3	157	83****	158	88****	14.6	14.2	13.5	10	71	73	63	88***
	4	144	78****	250	97****	14.8	15	17.2	11.8	75	78	67	77
	5	158	89****	152	83****	17.5	16.3	19.1	10.4*	88	89	73**	83
	6	154	73****	155	90****	50.1	51.5	23.7****	11.1****	101	104	81****	90*

Table XIV- Statistical analysis of antibiotic evolution assay relative to wild type - Means and mean differences were determined by a two-way ANOVA test of three independent biological repeats. A Dunett's multiple comparisons test was used to compare means variation between mutant strains and wild type at each concentration. Individual p-values of 95% confidence intervals are displayed as stars (<0.05= *; <0.005=**; <0.0005=***; <0.0001=****. Na = Not attributed.

4-3- Discussion:

4-3-1- Replication fidelity in *E. coli* and mycobacteria:

In the model organism, *E. coli*, and in most of prokaryotes organisms, the replication fidelity is dependent on three complementary mechanisms.[111] Unlike error-prone polymerase, such DnaE2, the Pol IIIα encoded by *dnaE1* has the inherent ability to add the right nucleotide at the right position.[34,97,112,126–128] This phenomenon is called intrinsic fidelity. But nothing is perfect in nature and, the polymerase is seconded by two other mechanism that help guaranteeing the replication fidelity. The proofreading ε-subunit, encoded by *dnaQ*, is able to remove misincorporated nucleotides as the replication goes.[106,129] Following replication, a MutS-MutL- MMR system acts to reduce the mutation rate. It is now established that mycobacteria exhibit important differences even in fundamental biological processes[29,32] when compared to the model organism. For the replication fidelity, it was numerously proven [30,31] that the proofreading activity is performed by DnaE1 PHP domain. Yet, *dnaQ* is a highly polymorphic gene which has been identified as a target of independent mutation associated with drug resistance in *Mtb*. It is also under a strong selective pressure.[34,35] Consequently, this chapter focused on characterizing the role of DnaQ under the prism of replication-targeting drugs.

4-3-2- Confirmation of the lack of contribution of mycobacterial DnaQ to the replication fidelity after UV exposure:

Before further analysis could be carried out, previous work about DnaQ had to be repeated. Deletion of *dnaQ* was never shown to increase the mutation frequency following UV-exposure. The data generated in this work is consistent with the literature and other previous experiment carried out in the MMRU. However, repeated failures to generate a phenotype in the adapted fluctuation assays might be due to the limitations of the assay. Indeed, such assays are able to detect mutation on the single *rpoB* locus and thereby neglect neutral or lethal mutations and mutations in noncoding areas of the genome, which likely occur more frequently, are ignored [107]. As mutations are not necessarily equally distributed on the genome [107,130], the effects of *dnaQ* may have effects that remain to be detected.

4-3-3- Potential synergistic activity of mycobacterial DnaQ homologs to maintain the genome integrity:

Ford et al did not observed an increase of the mutation rate for this double knock-out mutant [40] and showed that the replication fidelity relies mostly on the PHP domain of the DnaE1 polymerase. Those results gave new insight in mycobacterial replication fidelity and identify it as a prime target for anti-TB therapy. Repetition of this experiment at the start of this project brought to light unprecedented results. Indeed, deletion of both DnaQ homologs resulted in an increase of the mutation frequency in the mutation frequency assay. As this result is not consistent with the current literature and because of the inherent variability of the fluctuation assay, it must be carefully considered. If this interaction is real, its effect on the mutation rate is rather modest (less than a 10-fold increase) but quite reproducible in the assays performed. The effect was still noticeable in the DnaE1 mutant strain. Together with a PCR screening of the strains, that ascertained their genotype, it gave confidence in the validity of results obtained.

4-3-4- Conserved metal coordinating residues of the PHP domain of mycobacterial DnaE1 is mandatory to maintain replication fidelity:

The mutator phenotype of strain exhibiting a mutated PHP domain in DnaE1 has been repeatedly studied.[31,32,66] Only the mutant exhibiting the strongest increase in the mutation rate has been used in this study, hence the *dnaE1*-D228N mutant. The key results of those previous work were replicated here in the presence and absence of tetracycline in different *dnaQ*-related backgrounds. In the wild type background, the fold change in the mutation rate upon tetracycline induction was significantly higher than with the native DnaE1, which was consistent with the literature. This phenotype could also be observed in the various knock-out backgrounds used in this study. Interestingly the highest increase in the mutation rate was observed in the double *dnaQ dnaQ-uvrC* knock-out background by about a 200-fold. this result gives credit to the possibility of a synergistic activity of DnaQ homologs to in the replication fidelity. Regarding the WT::PJR65, similar mutation frequency data on Rif plate were obtained in other research work performed by Zela Martin or Rock et al.[31,66]

4-3-5- Mycobacterial DnaQ is involved in the Gyrase A pathway:

Previous work from Mduli et al identified DNA gyrases as an effective target to treat MDR-TB[52]. DNA gyrase is a type II A topoisomerase formed by hetero-tetramerisation of 2 GyrA subunits and 2 GyrB subunits, to form the active complex.[127] GyrA is responsible for the breakage-reunion of the active site while GyrB promotes ATP hydrolysis. Together, they are responsible for catalysis of the negative supercoiling of DNA by generating transient dsDNA breaks which is opposite to the action of the topoisomerase that unwinds dsDNA. Quinolones inhibit gyrases in such way that dsDNA breaks accumulate which prevents the replication from being carried out.[52,131–133] So far, every attempt to generate a Δ*dnaQ* phenotype failed. However, this study could not identify a strong phenotype for such a knock-out, a slight but statistically significant and reproducible sensitivity to moxifloxacin, could be observed. Even though this sensitivity is too small to mention a direct involvement of DnaQ in gyrase A activity and could be due to a lack of DNA repair or a completely different impairment, it might help direct further experiments to characterise DnaQ activity.

4-3-6- Mycobacterial DnaQ is required for antibiotic adaptation:

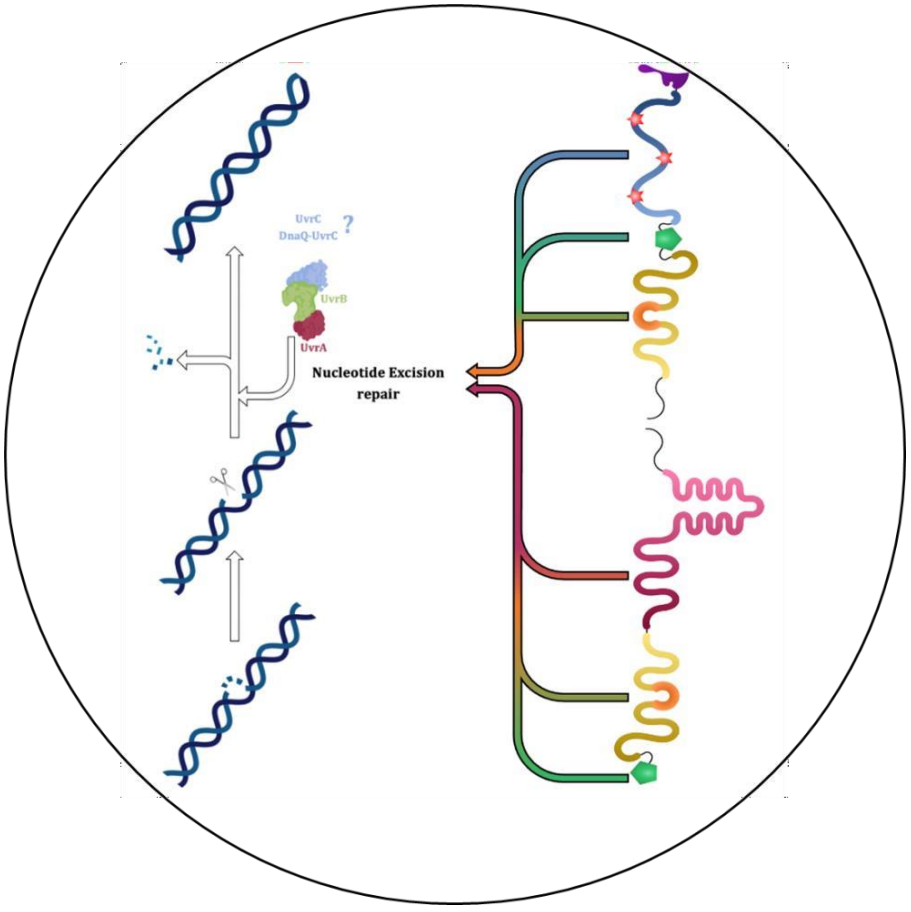
Antibiotic adaptation can occur through two phenomena, both part of the genome maintenance. Mutagenesis, which will be further described in Chapter V, is the main driver of antibiotic resistance. However, modification of the DNA sequence can also take place during the replication. Even though no molecular clear role has been highlighted for DnaQ in the replisome, it has been shown to be linked to the emergence of drug resistance in *Mtb*[148]. Here we provide new evidence of this assessment. Indeed, Δ*dnaQ* reduced the adaptation observed in the wild type and other "adaptative" strains. On the other hand, Δ*dnaE2,* almost completely inhibit antibiotic adaptation which is consistent with the current literature[102] that identified DnaE2 has the translesion (TLS) polymerase mainly responsible for mutagenesis.

4-3-7- Recommendations for future research:

This study failed to provide a robust *dnaQ* deletion phenotype under antibiotic treatment. Even though, the slight moxifloxacin sensitivity was significant, (and the first ever reported), it does not allow us to draw conclusions relative to DnaQ function. However, in the light of the Chapter 3 results and the presence of BRCT domain, an interaction between DnaQ and gyrases seems convincing. However, its exact role remains to be established. The Δ*dnaQ::eGFP-dnaQ* strain was designed for fluorescent microscopy purpose. Moreover, there is a Flag-C tag between the eGFP and the rest of the protein that could be used for protein pull-down experiments. The main questioning was about the ability of this strain to restore the wild type phenotype. Protein domain study performed in Chapter III showed that DnaQ protein had unstructured amino acids before the exonuclease domain. This coupled to the Flag-C tag should give the protein enough flexibility so the eGFP does not prevent DnaQ from performing its activity. Furthermore, this strain was shown to restore the wild type phenotype after moxifloxacin treatment giving confidence to it could be used as a proxy to study native DnaQ.

The spontaneous mutation frequency assay highlighted a synergy between DnaQ homologs. This synergy reported here by a genetic approach should be pursued from a biochemistry approach. Indeed, eGFP-fused DnaQ homologs were generated during this project, and they could be used for fluorescent microscopy purposes by tracking their localization relative to each other or known member of the replication.

CHAPTER V:
Opposed phenotypes between *dnaQ-uvrC* and *uvrC* in Mycobacterium smegmatis highlights a damage specificity in Nucleotide Excision Repair



Created with BioRender

5-1- Introduction:

5-1-1- DNA repair:

DNA damage can be sustained by any organism as a function of the conditions in which it lives. This is especially true for intracellular pathogens such as *Mtb* which must endure demanding conditions during an infection owing to the host's defence systems and, for TB patients, antibiotic treatments which can also damage DNA. [30,39] To survive and replicate, bacteria have evolved genetic systems which increase the expression of genes that encode DNA repair components. In most organisms, the DNA damage response (also known as the SOS response) is under the control of two key regulatory proteins, RecA and LexA.[39,51,121,134,135] LexA binds to a sequence named the SOS box, which is located in the promoter region of the genes it regulates, thus blocking their expression. [136,137] Upon detection of DNA damage, RecA, in the presence of single-stranded DNA,[138] induces an autocatalytic reaction that results in LexA protein cleavage. This results in derepression of the RecA/LexA-regulated DNA repair genes.[139]

5-1-2- Mycobacterial DNA damage response:

Mycobacteria, along with most organisms, harbour multiple DNA repair mechanisms that are triggered depending on the kind of damage sustained and/or the replicative state of the bacillus. Among the major DNA repair pathways are base excision repair (BER), nucleotide excision repair (NER), homologous recombination (HR), non-homologous end-joining (NHEJ), and single-strand annealing (SSA)[140]. In this work, NER is of paramount importance and is described below in section 5-1-3.

In mycobacteria such as *Mtb* and *Msm*, the majority of inducible DNA repair genes are triggered independently of RecA/LexA. As shown by Rand et al,[39] upon exposure to mitomycin C, a number of DNA repair genes were triggered in a *recA*-deficient strain, suggesting an alternative SOS response. Even *recA* expression itself is dependent on two promoters[141]; (i): the classical LexA/ RecA-dependent promoter; and (ii) an alternative RecA-independent promoter originally named RecA-NDp.[51,135,141–143] Interestingly, the vast majority of DNA repair genes are triggered by the alternate SOS response[34] such as *uvrA*, *uvrB* in NER, *xthA*, *nei* in BER, *ogt* in damage reversal and *recA, radA* in recombination. This suggests that this mechanism is of paramount importance in *Mtb* and *Msm*, just as was documented in *Xylella fatidiosa* and in *Acinetobacter*

calcoaceticus.[141,143,144] The corresponding genes in *E. coli* model are all regulated by the classical RecA/LexA system.[145–148]

Over time, this alternative SOS response has been better characterised. First, by identifying its conserved regulating promoter,[149] named RecA-NDp, and then its main activator, PafBC proteins. The genes *pafA*, *pafB* and *pafC* are organized in an operon in mycobacteria and in other actinobacteria. Together, PafB and PafC are able to form a heterodimer that acts as a transcriptional activator of most of DNA repair genes, as Muller et al have described.[37,38]

5-1-3- Nucleotide Excision repair:

Unlike other repair processes, NER can identify and fix a wide variety of DNA lesions such as cyclobutane pyrimidine dimers (Pyr<>Pyr), (6–4) photoproducts, benzo(a)pyrene-guanine adducts, and acetylaminofluorene-guanine (AAF-G) or cisplatin-d(GpG) diadducts. Because of the poor specificity of exonuclease, NER can also act as a backup of the BER system for the removal of oxidative and alkylation damage. The UvrABC proteins are the major components of the NER process and, while the mechanism is identical throughout all kingdoms of life, the proteins utilised may change. However, the outcome of NER is invariably the same: damage is removed in a form of a 12 to 13 nucleotides long oligomer and the subsequent gap is filled by DNA polymerase I without generating any mutations.[150–153]

Originally, NER was discovered as an *E. coli* system that removed UV-induced cyclobutene thymine dimers and comprised a 3-protein complex, UvrABC. NER is initiated by an ATP-dependent two-step process that recognise a variety of DNA damages[154]. Even though the exact recognition process remains to be elucidated, UvrA seems to be the initiator of NER by sensing and binding DNA damage without the assistance of the other components. [154,155] UvrA is a 104 kb DNA-binding ATPase that acts as the sensor of damaged DNA.[156]Indeed, it was shown that UvrA dimerizes through an ATP-dependent reaction to form a complex with damaged DNA The UvrA/damaged-DNA complex is unstable and requires ATP hydrolysis to form a stable complex that unwinds the DNA duplex that will be recognized and bound by UvrB to form a damage recognition complex.[154,155] Even though the 3D structure of UvrA has been visualised with cryoelectron microscopy (cryoEM),[155] the mechanism underlying the genome scanning and the damage detection

remain to be understood. Not only does UvrA reject native DNA (by virtue of its shape) from inappropriate repair,[157] it also selects damaged DNA for further processing by allowing interaction with UvrB.

UvrB is the central component of NER as it is able to interact with all the other components of the system: UvrA, UvrB, UvrD, Helicase II, DNA polymerase and DNA .[154,158,159] However, critically, UvrB has a low affinity for DNA and, therefore, is not able to bind any DNA damage that has not been detected by UvrA.[154,157] Yet, previous work on UvrB highlighted that UvrB can scan dsDNA to check the presence and type of damage thanks to its helicase domains.[159,160] Its structure is similar to that of a helicase but it has a single-strand DNA translocase activity and rather limited helicase activity. This structural analysis[159] allowed the authors to speculate that UvrB is able to move along the DNA strand in an ATP dependent motion, until it meets a UvrA/damaged DNA complex.[183,184] Indeed, a mutation in UvrB that abolishes helicase and translocase activity shows specific defects in the later steps of NER.[157] The way UvrA, damaged DNA and UvrB interact remains unclear, but it is thought that a UvrB dimer binds a UvrA dimer to form the pre-incision complex. Symmetry of the complex could allow each UvrB subunit to be loaded onto opposite strands of the DNA duplex.[154] However, once loaded on the damaged-DNA/UvrA complex, UvrB induces a conformational change to form a stable pre-incision complex that initiates the second phase of the repair. [154,157–159]

The second phase starts with the concomitant unloading of UvrA and the loading of UvrC. Because all UvrABC proteins interact through the same domain, the UVR binding domain, they can only bind one protein at a time. Therefore, even though UvrB is able to recognize and bind both UvrA and UvrC, no complex involving the three proteins has ever been observed[161]. This might be part of a regulation system to ensure that the damage recognition occurs prior to the incision. UvrB loading onto DNA by UvrA, as well as each subsequent step of NER, are ATP-dependent – which allows a time delay between damage recognition and the incision.[162]

UvrC is not as extensively studied as the other components and it was only recently shown that UvrC allows both 3’ and 5’ incision reactions thanks to two separate catalytic sites. [150,151,154,155,157,163] The 3’ incision is performed by the GIYYIG domain of UvrC which is located at its N-terminal. At the C-terminal is a RuvA2-like endonuclease domain which is responsible for the 5’ incision. This endonuclease is located right next to a helix-hairpin-

helix (HhH) domain that mediates UvrC exonuclease activity because it allows a tight contact with DNA required for both incisions.[184,195] Once both incisions have been carried out, the helicase UvrD1 excises the DNA oligomer prior to the repair synthesis and ligation steps, performed by DNA polymerase I and by DNA ligase I which joins the newly synthesised sequence to the rest of the DNA strand.[29,156,164,165] As polymerase I possesses a 3'-5' exonuclease activity,[155] NER is a high-fidelity DNA repair system. Unlike BER, NER will result in the incision of the flanking regions of damaged DNA whereas BER results in single-base removal. The kind of damage that NER deals with is broader than BER due to the low specificity of the exonucleases involved in the process. [29,156,164,165]

Mycobacteria exhibit differences with the model organism. Even though UvrABC exonucleases and the UvrD helicase are conserved in mycobacteria, they have an unusual feature as they encode two DNA helicase II proteins, UvrD1 and UvrD2.[166] The gene *uvrD1* is closely homologous to the *E. coli uvrD* although it is not essential in the model organism. On the other hand, *uvrD2* seems to be closer to *prcA* which is essential in *B. subtilis* and *S. aureus*. So, it is possible that *uvrD1* and *uvrD2* have overlapping functions and are of equal importance as, they show the same level of conservation in clinical strains.[167]

5-1-4- NER sub-pathways:

NER is divided into two sub-pathways; (i): Global genomic repair (GGR) identifies defects everywhere throughout the genome; whereas (ii): Transcription-coupled repair (TCR) occurs when RNA polymerases are inhibited by lesions on the template DNA strand. GGR can be considered as described previously, whereas TCR requires additional factors. The mutation frequency decrease (Mfd) factor, also known as transcription coupled repair factor (TCRF), is the first protein involved in TCR in *E. coli.* [168] Mfd is a translocase that attaches to the arrested RNAP and pushes it forward from behind; at the same time, Mfd recruits UvrA via its UvrB homology domain, initiating GGR.[164,169] However, *mfd* inactivation has no effect on UV sensitivity or survivability,[170,171] indicating that an alternate Mfd-independent route exists.[168,172] UvrD has a prominent role in this response, according to Epstein et al.[173] UvrD attaches to RNAP and travels with it along the DNA template. When the RNAP pauses, UvrD can actively pull the polymerase back, unwinding the duplex DNA "behind" the transcription bubble using its helicase activity.[173] This would be in addition to UvrD's action in dissociating the oligonucleotide produced by incision

during NER. Although well understood, gaps in the knowledge about these sub-pathways of NER remain. Indeed, the factors influencing the rate of repair in TCR versus GGR remain to be characterized. Given the vast amount of undamaged DNA that must be scanned for relatively rare lesions,[153] the search for DNA damage by the pre-incision complex must be the rate determinant, whereas 1-dimensional scanning by the elongation complex may lead to more rapid encounters with lesions in transcribed strands than 3D scanning by the UvrAB pre-incision complex.[164] Another factor to consider is which mode of TCR functions on a certain DNA sequence and for a specific sort of lesion: numerous lines of evidence suggest that *Mfd* only interacts with the polymerase after the elongation complex is created.[172] Replication is usually carried out by DNA-dependent polymerases such as DNA polymerase III. Those are often unable to synthesize DNA over damaged template strands.[110] Bacteria and almost all other organisms have evolved specialist translesion synthesis (TLS) polymerases capable of completing DNA synthesis across these lesions.[174] A secondary consequence of these TLS polymerases is that, as many are error-prone, they generate genetic diversity by causing inadvertent mutations at a high frequency within the repaired DNA. This may be highly mutagenic, but its ultimate purpose is to offer the cell an opportunity to survive by allowing DNA and cell replication to continue.[102,103]

5-1-5- Mutagenic DNA repair pathways:

Some DNA damage cannot be properly repaired, blocking the replication machinery and resulting in cell death. Specialized TLS polymerases are able to bypass such damage and resume DNA replication over them. This mechanism can be error-prone and can result in mutagenesis.[175] In the model organism *E. coli*, mutagenesis is performed by multiprotein complex called the mutasome that was shown to be responsible for the emergence of drug-resistance *in vivo*.[123] There are several TLS polymerases in *E. coli* that form the mutasome: the B-family Pol II and the Y-family Pol IV and Pol V (encoded by *dinB* and *umuDC* respectively). As these genes do not have a counterpart in mycobacteria, it was long thought that TLS relied on two Pol IV polymerase homologs,[174,176] DinB1 and DinB2. However, previous work has shown that, in mycobacteria, the mutasome is composed of TLS polymerase DnaE2 along with two accessory protein ImuA' and ImuB[158] which are encoded on a split mutagenesis cassette under RecA/LexA SOS response regulation. This ImuA'-ImuB/DnaE2-encoded system constitutes a non-orthologous

replacement of the PolV mutasome, UmuD'2C•RecA•ATP, which has been extensively characterized in *E. coli.*[54,177,17] ImuB is one of the putative Y-family polymerase homologs in *Mtb* genome although it lacks the invariant site acidic residues needed for catalysis; it also contains a supplementary C-terminal domain predicted to be involved in protein-protein interaction which contrasts with the canonical Y-family polymerases, DinB1 and DinB2.[102] ImuB is thought to act as a linker protein between ImuA' and DnaE2 via C-terminal domain and a β-clamp binding motif. ImuA is structurally related to RecA minus the C-terminal and RecA-homologous region.[102,123] It is able to bind both DNA and ImuB but its function remains to be solved. The mutasome is thought to be the main source of induced mutagenesis in *Mtb*. Indeed, deletion of *dnaE2* renders *Mtb* hypersensitive to DNA damage and eliminates induced mutagenesis.[68] It also reduces the frequency of drug resistance *in vivo*. Consistent with ImuB's inferred inability to catalyse any polymerase reaction, DnaE2 was identified as the main TLS polymerase. DnaE2 is a C-family DNA polymerase thought to allow mycobacteria to survive and adapt to harsh DNA-damaging conditions within the host.[68] It does not have a β-clamp binding motif and thus needs ImuB in order to gain access to the DNA. The associated gene is part of *the imuA'/imuB* operon involved in the SOS response and act as a linker between the mutasome and the β-clamp. Interaction with *dnaN* gene product is essential to recruit the full mutasome complex to the site of the DNA damage, inferring the essentiality of ImuB for the bacterium survival to DNA damage.

In this chapter, the role of DnaQ-UvrC (*Mtb* Rv2191; *Msm* MSMEG_6275) in the DNA damage response will be investigated. Other NER components such as UvrB and UvrC will also be studied. Although biochemical *in vitro* experiments have already been persued,[41,157,159,179,180] actual *in vivo* work in *Msm* remains limited.

5-2- Results:

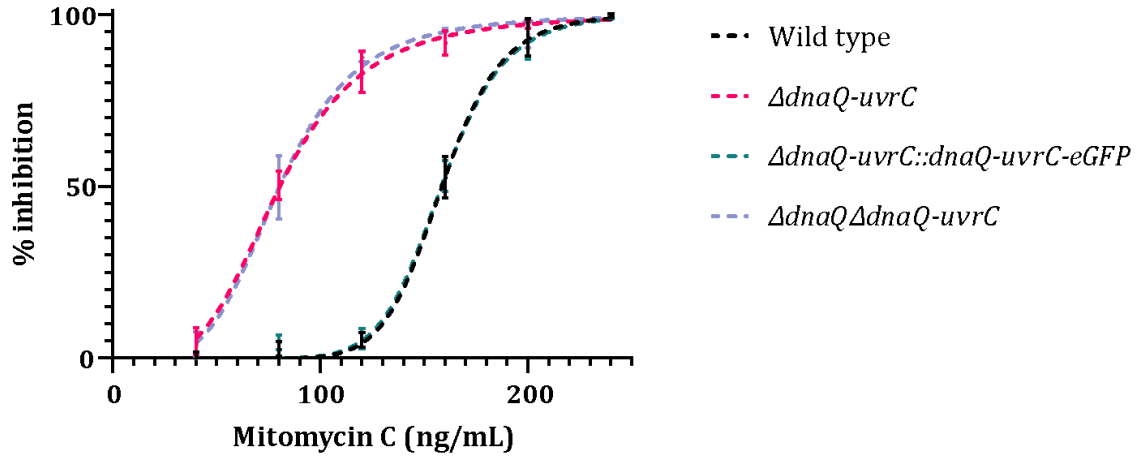
5-2-1- Deletion of MSMEG 4259 (*dnaQ-uvrC*) results in MMC hypersentivity:

Previous work in our laboratory[41] reported the hypersusceptibility of an *Msm* Δ*dnaQ-uvrC* (Δ*MSMEG_4259*) knockout mutant to MMC treatment. Surprisingly, this work also indicated that the damage-sensitivity phenotype was reversed by the deletion of *dnaQ* in the Δ*dnaQ-uvrC* background. To confirm those observations, the MMC sensitivity

assays were repeated at the start of this project. In addition, a complemented strain was included in which eGFP was fused at the C-terminal of the wild-type *dnaQ-uvrC* (Supplementary Figure III) and introduced into the Δ*dnaQ-uvrC* background. Survival was measured by spotting serial dilutions on solid medium containing 20 ng/mL MMC. In parallel, survival in liquid medium was determined in an MIC assay.

These experiments confirmed the hypersensitivity of the Δ*dnaQ-uvrC* mutant (Figure XXIX). Indeed, an IC_{50} of 158.3 ng/mL is observed for the wildtype whereas the *ΔdnaQ-uvrC* strain is at 80.13 ng/mL (Table XV). Notably, complementation with the *dnaQ-uvrC-eGFP* allele phenocopied wildtype. However, the reported reversal of the Δ*dnaQ-uvrC* damage sensitivity phenotype in a double Δ*dnaQ-uvrC* Δ*dnaQ* mutant was not observed. Instead, the double knockout exhibited similar MMC sensitivity to the Δ*dnaQ-uvrC* mutant. Given the discordance with prior observations, the genotypes of all strains were reconfirmed by PCR (Figure XXII, Section 4.2.1).

A



B

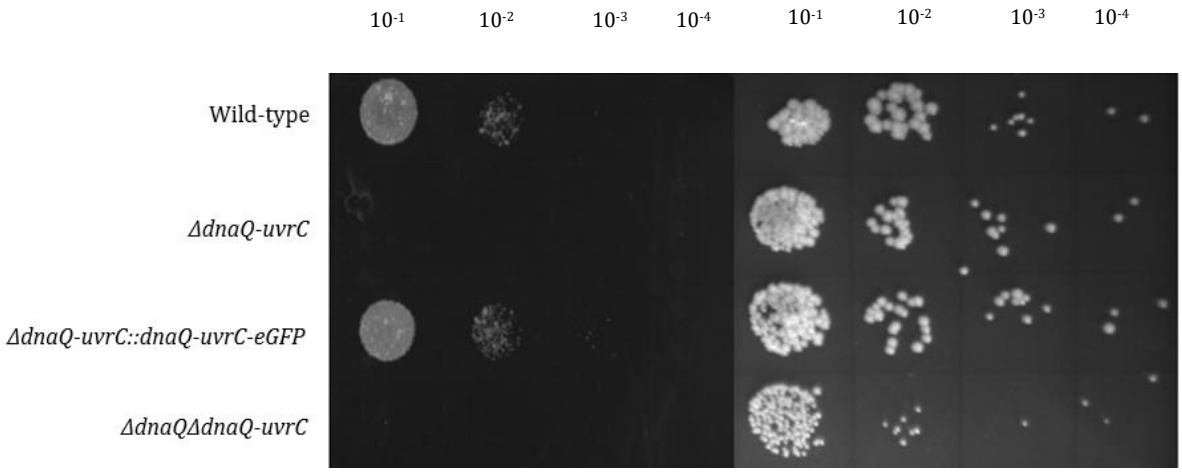


Figure XXIX- Inhibition of *Msm* strains in response to Mytomycin C (MMC). A- Liquid MIC assay. Percent of inhibition by MMC treatment. In liquid growth of strains was performed as described in Material and Methods. Results were obtained following resazurin coloration and optical density measurement. Values were then normalized to wild type values. Nonlinear fitted curves were drawn using the [inhibitor] vs normalized response—Variable slopes—tool of GraphPad Prism B- Solid medium spotting assays. Performed as described in Material and Methods. Right panel shows growth without MMC. Left panel shows growth on 20 ng/mL MMC medium. From left to right spots, successive 10-fold bacterial dilutions were performed and are indicated on top on the figure. The experimental designed of both experiments used three independent biological repeats for each assay of both experiments. For Figure B, a representative picture of three
independent experiments is shown.

		Strains			
	MMC (ng/mL)	Wild-type	Δ*dnaQ-uvrC*	Δ*dnaQ-uvrC::dnaQ-uvrC-eGFP*	Δ*dnaQ*Δ*dnaQ-uvrC*
Mean Difference with wild type strain (%)	40	Na	-4.333	0.333	-3.333
	80		-47.67*	-2	-47*
	120		-78**	-0.333	-82.33***
	160		-39**	-0.333	-39.333*
	200		-3.667	1	-3
	240		0	0	0
IC_{50}		158.3	80.13	158.1	80.04
95% Likelihood Interval		[155.8-159.3]	[77.07-83.13]	[155.4-160.5]	[76.29-83.63]

Table XV- Statistical analysis- Means and mean differences were determined by a two-way ANOVA test of three independent biological repeats. A Dunett's multiple comparisons test was used to compare means variation between strains at each concentration. Individual p-values are displayed as stars. Na = Not attributed. IC_{50} and 95% likelihood intervals were determined by non–linear regression.

5-2-2- Sequential targeted inactivation of DnaQ-UvrC interaction domains by site-directed mutagenesis:

As discussed in Chapter 3, DnaQ-UvrC (*Msm* MSMEG_4259; *Mtb* Rv2191) contains two predicted interaction sites for β-clamp and UVR binding. To determine the relative importance of these two binding sites to DnaQ-UvrC function, mutants were generated in which these sites were selectively disrupted (Table XVI), either by amino acid substitution (β-clamp binding site) or via domain deletion (UVR binding site).

As shown in Table VI, two pairs of primers (primers SA 101/102 And SA103/104) were designed to amplify the full length of the targeted gene along with its promotor located upstream of the gene. The upstream reverse and the downstream forward were designed to have their 5' sequence complementary to one an another. Because of the use of unusually long primers, unspecific primer annealing can be problematic. Thus, at each step, a gradient PCR was performed to determine the optimal temperature to obtain the desired PCR product. These overlapping primer sequences also contained the engineered mutation (Figure XXX).

The upstream forward and downstream reverse primers contained an engineered *Kpn*I cutting site which allowed the cloning of final PCR product into the integrative mycobacterial shuttle vector, *pMV306*.[58] Thus, both final PCR product and the vector were digested using *Kpn*I and ligated to generate *pMV306 attB::dnaQ-uvrC* $^{\Delta\beta\text{-}clamp}$. This construct was electroporated into the *dnaQ-uvrC* knockout to generate the *dnaQ-uvrC* $^{\Delta\beta\text{-}clamp}$ strain (Table XVI).

The UVR binding site mutant was generated using two pairs of primers: SA101/SA158 and SA159/SA104. The amino acids involved in the interaction are not clearly identified. So, the primers were designed to amplify all the proteins except the nucleotides coding for the UVR binding domain, generating *dnaQ-uvrC* $^{\Delta UVR}$. As the UVR binding site is the last identifiable structure of DnaQ-UvrC, its deletion should not affect the protein function other than the intended domain deletion. Finally, a third mutant combining the characteristics of both variants was generated. The process was the same as described above except that *dnaQ-uvrC* $^{\Delta\beta\text{-}clamp}$ strain was used as a template instead of the wildtype version of the gene. The final PCR product and the vector were digested with *Kpn*I and

ligated to generate *dnaQ-uvrC* $^{\Delta IS}$, which is a version of *dnaQ-uvrC* deprived of all its known interaction sites. Strains were sequenced to verify the deletion of the targeted sites (Supplementary Table I and Figure I).

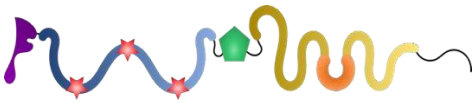
	DnaQ-UvrC		**Mutated DnaQ-UvrC**		
Interaction site	Nucleotide Sequence	Amino Acid Sequence	Nucleotide Sequence	Amino Acid Sequence	Variant
β-clamp binding site	CAG CTC TCG CTG GAC	QLSLD	GAG CAA GTG CAG AAC	EFVAY	*dnaQ-uvrC* $^{\Delta\beta\text{-clamp}}$
UVR binding domain	GCG GAC AGC … GTG GCG ATC	ADSA…VAID	Na	Na	*dnaQ-uvrC* $^{\Delta UVR}$
					*dnaQ-uvrC* $^{\Delta\beta\text{-clamp}}$
					*dnaQ-uvrC* $^{\Delta UVR}$
					*dnaQ-uvrC* $^{\Delta IS}$

Table XVI- Targeted inactivation of DnaQ-UvrC interaction sites.

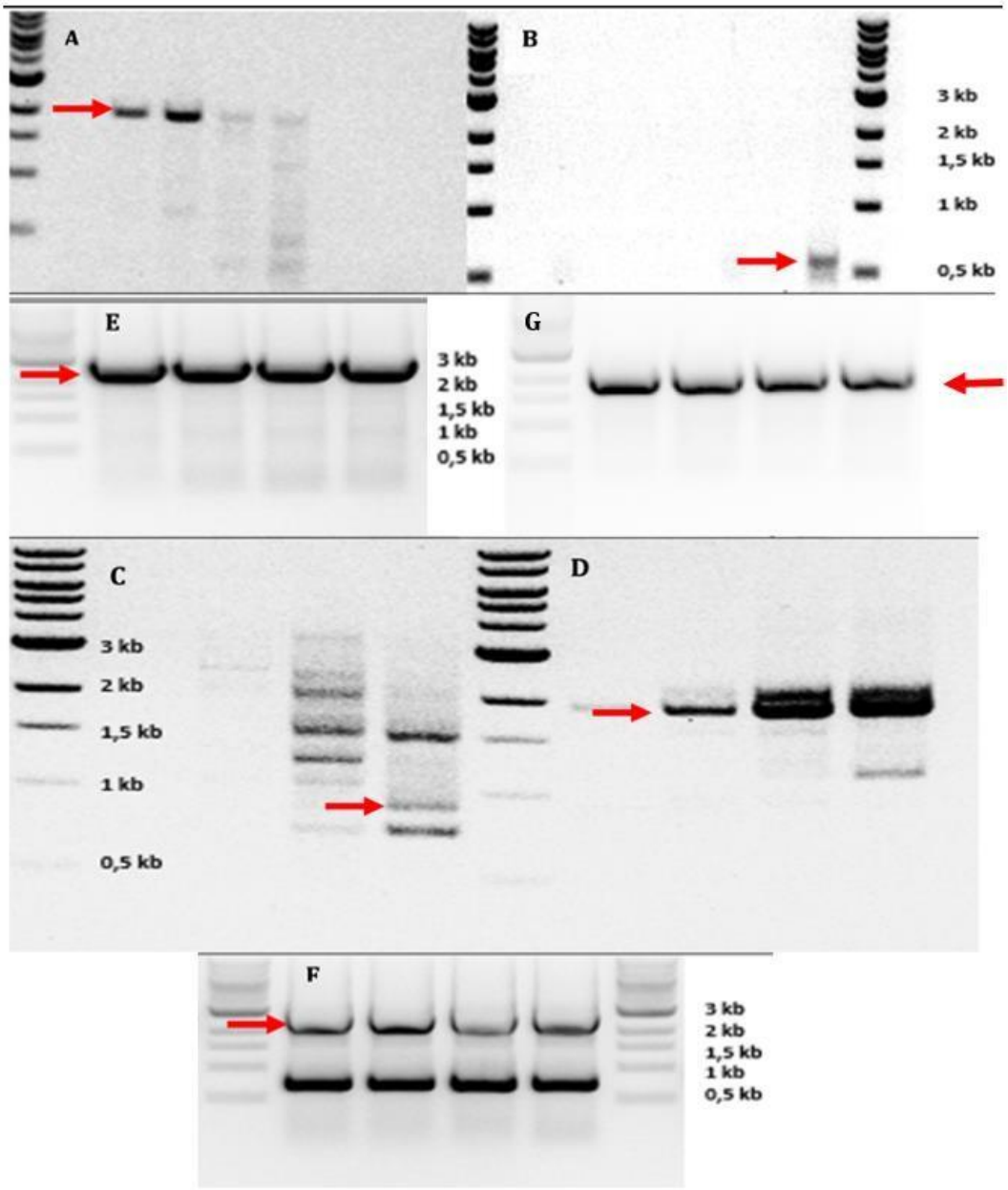


Figure XXX- d*naQ-uvrC* targeted mutagenesis by gradient overlap PCR. A- and B- UVR binding motif targeted mutagenesis. Primers were designed to have overlapping sequences holding the mutated UVR motif. A- amplification of the *dnaQ-uvrC* upstream part (1941 bp, red arrow). B- amplification of the *dnaQ-uvrC* downstream region (591 bp, red arrow). C- and D- β-clamp binding motif targeted mutagenesis. C- amplification of the *dnaQ-uvrC* upstream region (800 bp, red arrow). D- amplification of the *dnaQ-uvrC* downstream region (1800 bp, red arrow). E- UVR binding motif upstream and downstream regions amplified together (about 2.5kb). F- β-clamp binding motif upstream and downstream regions amplified together (about 2.5kb, red arrow). G- UVR binding motif upstream and downstream regions amplified together in *dnaQ-uvrC* $^{\Delta\beta\text{-}clamp}$ (about 1818bp, red arrow). The molecular weight marker is the NEB 1kb ladder.

5-2-3- β-clamp and UvrB binding domain are required for MMC-induced DNA damage tolerance by MSMEG 4259 (=*dnaQ-uvrC*):

The MMC damage sensitivity assay was used to determine the relative importance of the predicted DnaQ-UvrC interaction sites (Figure XXXI & Table XVII). The experimental design used three independent biological repeats for each assay. Survival was determined by the spotting serial dilutions on solid 7H10 medium containing 20 ng/mL MMC. In parallel, MMC sensitivity was also determined via MIC assay in liquid. As shown in Figure XXXI, inactivation of the β clamp-binding site resulted in an increased sensitivity to MMC when compared to the wildtype, yet the inhibition was not as strong as for the full knock-out (Table XVII). The inactivation of the UVR binding domain resulted in an increase sensitivity to MMC when compared to both wildtype and *dnaQ-uvrC* $^{\Delta\beta\text{-clamp}}$ strain but was less sensitive than the full knock-out. On the other hand, deletion of both interaction sites (*dnaQ-uvrC*$^{\Delta IS}$ strain) did mimic the *ΔdnaQ-uvrC* phenotype. Solid medium survival assay confirmed the results, although the strain seemed to be more sensitive to MMC while growing on solid media – perhaps because of the prolonged exposure to MMC on solid plates. As the inactivation of the UVR binding motif had the greatest impact on the phenotype in liquid MIC assays, it was decided to generate a number of NER gene deletion mutants: *uvrB* (MSMEG_3816) and *uvrC* (MSMEG_3078).

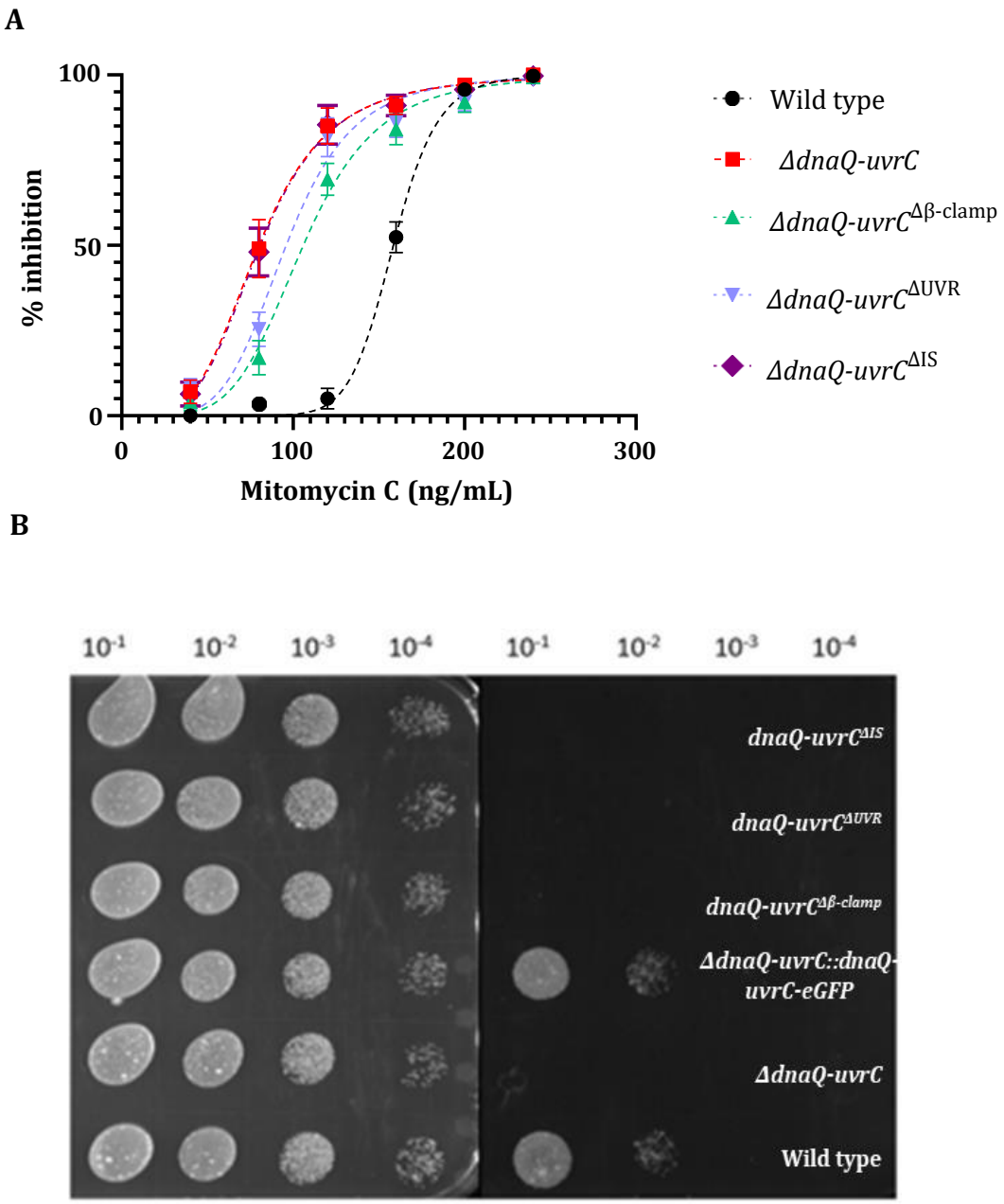


Figure XXXI- Inhibition of *Msm* strains in response to Mytomycin C (MMC). A- Liquid MIC assay. Percent of inhibition to MMC treatment. Fully complemented strain (Δ*dnaQ-uvrC*::*dnaQ-uvrC-eGFP*)data is not displayed for graph clarity as it overlapped wild type curve. In liquid growth of strains was performed as described in Material and Methods. Results were obtained following resazurin coloration and optical density measurement. Values were then normalized to wild type values. Nonlinea r fitted curves were drawn using the [inhibitor] vs normalized response — Variable slopes — tool of Graph Pad Prism B- Solid medium spotting assays. Performed as described in Material and Methods. Left Panel shows growth without MMC. Right panel shows growth on 20 ng/mL MMC medium. From left to right spots, successive 10-fold bacterial dilutions were performed.

	MMC conc. (ng/mL)	Strains				
		Wildtype	***ΔdnaQ-uvrC***	***dnaQ-uvr*CΔ$^{\beta\text{-clamp}}$**	***dnaQ-uvr*CΔ^{UVR}**	***dnaQ-uvr*CΔ^{IS}**
Mean Difference with Wildtype (%)	40	Na	-7	-4	-8.33	-6.33
	80		-45.67*	-13.67	-22*	-44.67
	120		-80***	-64.33**	-76.67**	-80.33**
	160		-38.67**	-31.67*	-33.67**	-38.67**
	200		-1.33	3.667	3.667	0
	240		0	0	1	0
IC_{50}	158.4	81.01	105.4	95.6	80.94	80.94
95% Likelihood Interval	[157.5-159.3]	[79.57-82.44]	[102.6-108.3]	[91.56-99.73]	[79.01-82.83]	[79.01-82.83]

Table XVII- Statistical analysis- Means and mean differences were determined by a two-way ANOVA test of three independent biological repeats. A Dunett's multiple comparisons test was used to compare means variation between strains at each concentration. Individual p-values are displayed as stars. Na = Not attributed. IC_{50} and 95% likelihood intervals were determined by non-linear regression.

5-2-4- Generation of unmarked *uvrC* KO mutant strain by gene replacement:

Generation of an unmarked KO strain is difficult because it requires two cross-over events after the cloning of the p2NIL suicide vector. This vector contained the upstream and downstream flanking sequences of *uvrC* without the gene sequence. To remove the gene, a first crossing-over event must happen to integrate the suicide vector into a bacterium genome. A second crossing-over is required that results in two different outcomes. Either the original genotype is restored, or the targeted gene is removed from the genome. The whole processed is described in Figure V & VI. These steps can be tricky as a crossover event happens in one out of 20 million cells. The genotype of the knock-out strain was checked by PCR analysis (Figure XXXII).

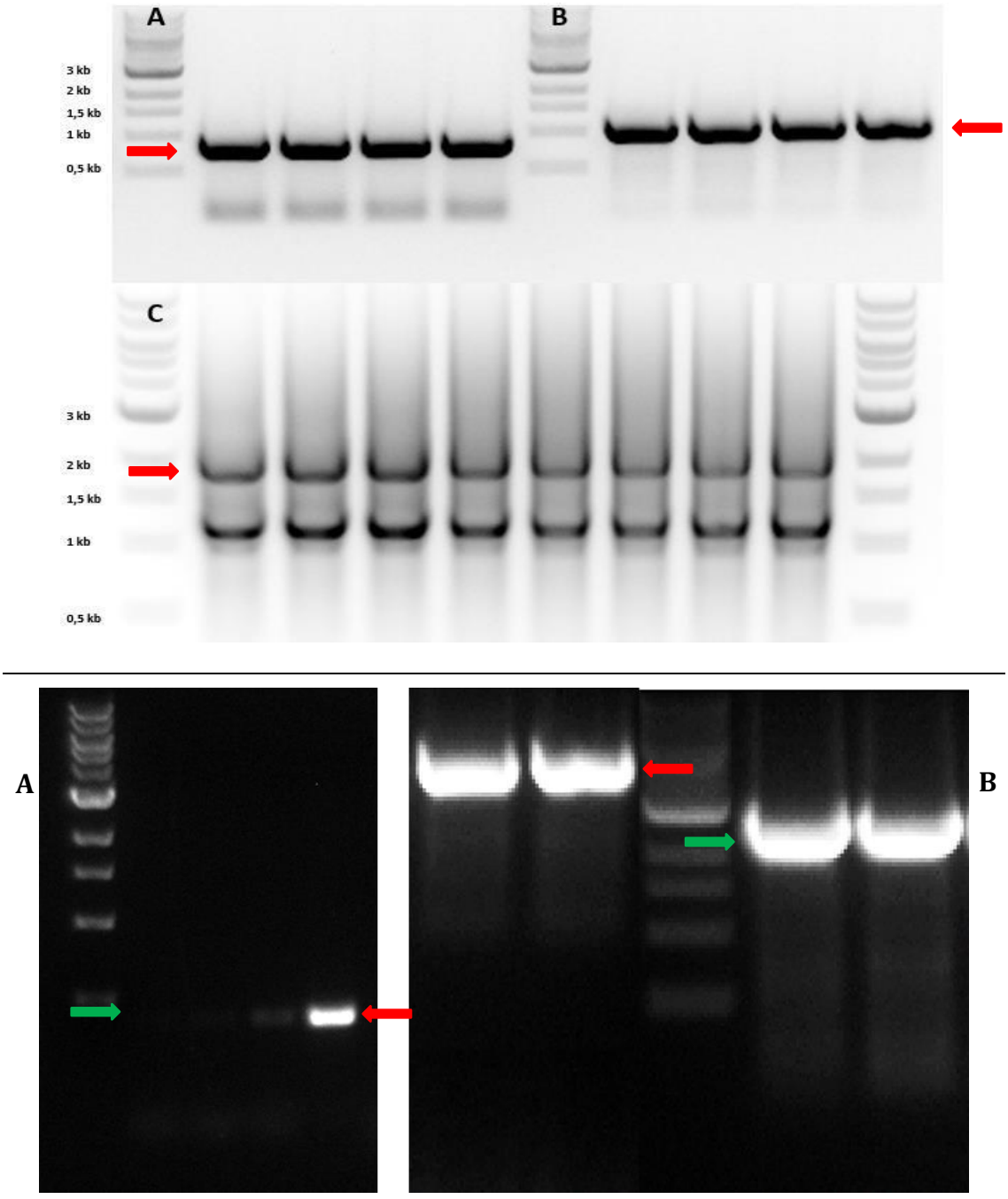


Figure XXXII- Construction of *ΔuvrC* strain- Top panel: *uvrC* flanking sequences amplification and ligation by gradient overlap PCR. A- Upstream flanking sequence amplification resulting in an 879 bp band (red arrow). B- Downstream flanking sequence amplification resulting in a 1044 bp band. For A- and B- there was no unspecific primer binding (red arrow). C - Upstream and Downstream sequence amplified together thanks to their overlapping sequences resulting in a 1923 bp band (red arrow). Non - specific primer binding amplified DNA sequences that are about 1 kb. The molecular weight marker is NEB 1kb ladder. Bottom panel: Genotypic confirmation of *uvrC* deletion mutant by PCR. For both figures, four clones are displayed (two with a valid *uvrC*, two with a deleted *uvrC*). A- Amplification of the *uvrC* inner part. Wild type (0.4 kb) (red arrow), *uvrC* deleted strain (no signal, green arrow). B- Amplification of *uvrC* flanking sequences. Wild type (4.5 kb) (red arrow), uvrC deleted (2.6 kb)(green arrow). Molecular weight marker is NEB 1kb ladde r.

5-2-5- No Δ*uvrC* phenotype under UV and MMC treatment:

Most of the research on NER uses ultraviolet (UV) light as the preferred DNA damaging agent.[161] Since NER has been shown to deal with several kinds of damage[184] and as MMC generates a broad range of DNA damage,[181] the deletion of genes involved in NER should result in an increased MMC sensitivity. Thus, both MMC and UV were used to phenotype the NER knock-out strains to compare them to Δ*dnaQ-uvrC*. UV exposure experiments used three independent biological repeats for each assay (Figure XXXIII). Survival on solid medium was measured by the size of colonies by plating on no drug plates. MMC survival on solid medium was determined by spotting on plates containing 20ng/mL MMC. Furthermore, MMC sensitivity in liquid medium was determined in an MIC assay. As shown in Figure XXXIII & Table XVIII, the deletion of *uvrB* almost phenocopied the Δ*dnaQ-uvrC* phenotype, consistent with the inferred importance of the DnaQ-UvrC UVR-binding domain. Surprisingly, deletion of *uvrC* had no effect on outgrowth after treatment and, if anything, appeared to confer a small but significant protective effect at high doses. When challenged with UV and then grown on solid media, deletion of *uvrC* still had no effect. Deletion of *dnaQ-uvrC* exhibited no sensitivity phenotype whereas *ΔuvrB* was very sensitive to both. The recurrent absence of phenotype for the *ΔuvrC* strain raised the question of the necessity of this gene in *Msm* to resolve thymine dimers induced by UV and other MMC-related damages.

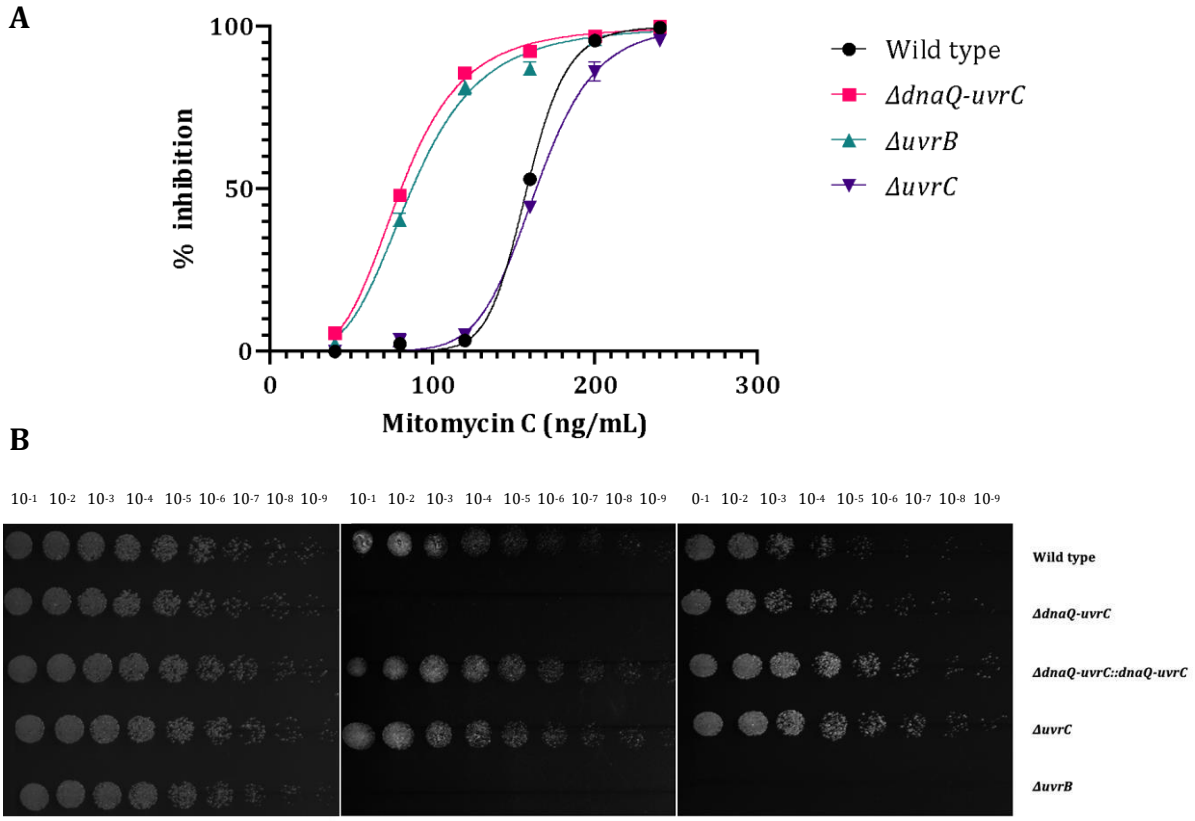


Figure XXXIII- Inhibition of *Msm* strains in response to Mytomycin C (MMC). A- Liquid MIC assay. Percent of inhibition to MMC treatment. In liquid growth of strains was performed as described in Material and Methods. Results were obtained following resazurin coloration and optical density measurement. Values were then normalized to wild type values. Nonlinear fitted curves were drawn using the [inhibitor] vs normalized response—Variable slopes—tool of GraphPad Prism B- Solid medium spotting assays. Performed as described in Material and Methods. Left panel shows growth on no drug medium. Middle panel shows growth on 20ng/mL MMC medium. Right panel shows growth on no drug medium after UV exposure as described in Material and Methods. From left to right spots, successive 10-fold bacterial dilutions were performed.

	MMC concentration (ng/mL)	Wildtype	Δ*dnaQ-uvrC*	Δ*uvrB*	Δ*uvrC*
Mean Difference with Wildtype (%)	*40*	Na	-5.66	-2.33	0
	80		-45.6****	-38***	-1
	120		-82.33****	-78****	-1.66
	160		-39.33***	-34****	8.66**
	200		-1.33	0.33	9.66*
	240		-0.33	-0.33	4
IC_{50}		158.4	81.01	87.70	164.1
95% Likelihood Interval		[157.5-159.3]	[79.57-82.44]	[84.93-90.4]	[162.4-165]

Table XVIII- Statistical analysis of MMC assay - Means and mean differences were determined by a two-way ANOVA test of three independent biological repeats. A Dunett's multiple comparisons test was used to compare means variation between strains at each concentration. Individual p-values are displayed as stars. Na = Not attributed. IC_{50} and 95% likelihood intervals were determined by non-linear regression.

5-2-6- Relative importance of *uvrC* and *dnaQ-uvrC* depends on the drug treatment:

To further investigate the role of *uvrC* and *dnaQ-uvrC*, the knock-out strains were challenged with other DNA damaging agents (Table IV) that targeted several components of the DNA metabolism either by blocking the interaction sites of the β-clamp (DnaN), preventing the DNA gyrase to unwind the DNA (GyrA/B), blocking the activity of the DNA polymerase (DnaE1), or inducing DNA damage in order to trigger DNA repair systems. Indeed, *dnaQ-uvrC* and *uvrB* are PafBC-regulated, whereas *uvrC* expression is constitutive.[32,33] As mentioned previously, mycobacteria exhibit two DNA damage responses. To control their importance in the inhibition assays, Δ*recA* and *lexA*ND (uncleavable LexA) strains were included (Table III). It was attempted to generate a *pafBC* knock-out strain, but it could not be achieved in time and will be further pursued. Since unrepaired DNA damage might be overcome by TLS polymerases,[158] a Δ*dnaE2* strain was also used in these assays in order to try and understand the balance between DNA repair and mutagenesis.

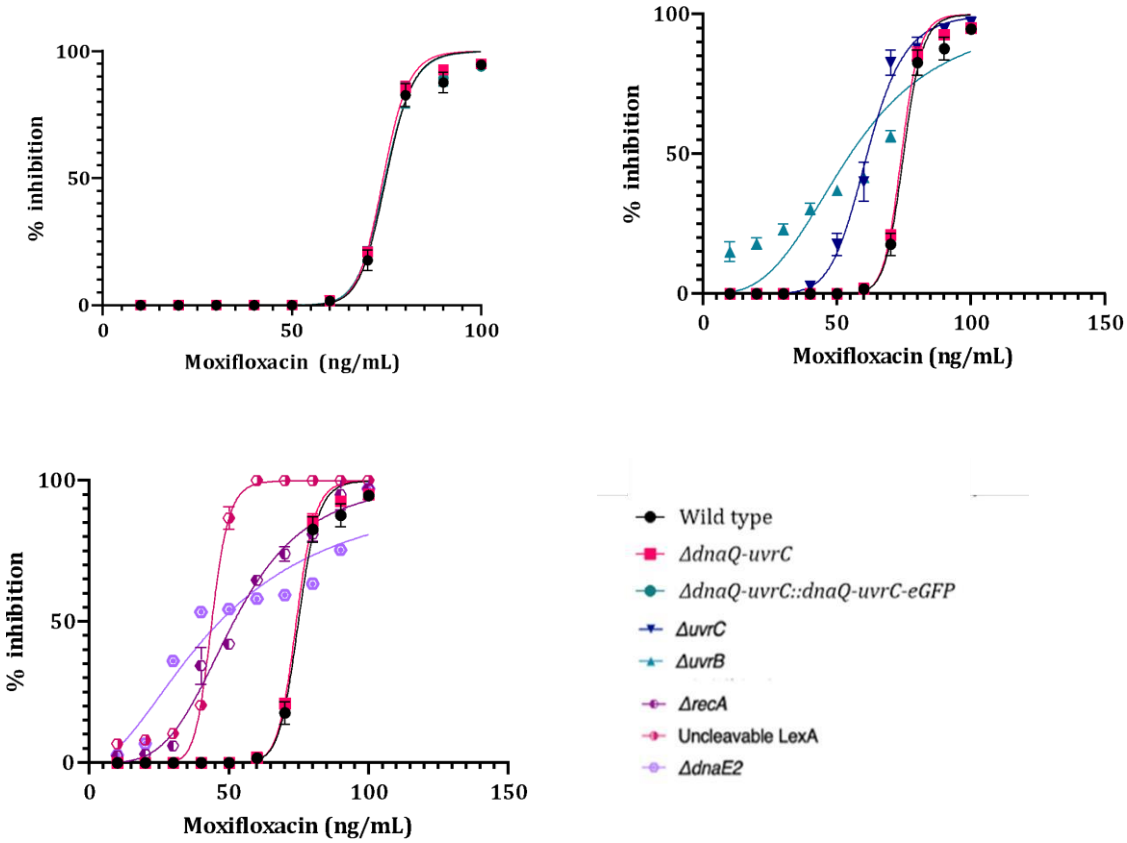


Figure XXXIV- Inhibition of *Msm* strains in response to Moxifloxacin. Top right panel highlights *ΔdnaQ-uvrC* strain and the complemented one. Top left panel highlights NER genes deletion strains. Bottom left panel highlights SOS response and mutagenesis genes deletion strains. The percent of inhibition to Moxifloxacin treatment was obtained after in liquid growth, performed as described in Material and Methods. Results were obtained following rezasurin coloration and optical density measureme nt. Values were then normalized to wild type values. Nonlinear fitted curves were drawn using the [inhibitor] vs normalized response —Variable slopes— tool of Graph Pad Prism.

The first drug tested was Moxifloxacin (Figure XXXIV), a fluoroquinolone antibiotic which inhibits GyrA activity which may induce transient dsDNA breaks.[144,218–220] Deletion of *dnaQ-uvrC* exhibited no sensitivity phenotype similar to the wild type (respectively IC_{50} = 74.18 ng/mL and IC_{50} = 74.94 ng/mL). This time the *ΔuvrC* strain exhibited a slightly more sensitive phenotype than the wildtype (IC_{50} = 61.35 ng/mL) along with the SOS-deficient strains, *ΔrecA* and LexAnd.

	Moxifloxacin concentration (ng/mL)	*ΔdnaQ-uvrC*	*ΔdnaQ-uvrC::dnaQ-uvrC-eGFP*	*ΔuvrB*	*ΔuvrC*	*ΔrecA*	*LexAND*	*ΔdnaE2*
Mean Difference with Wild type (%)	*10*	0	0	-15****	0	-2.33	-6.67**	-2.66
	20	0	0	-18****	0	-3	-8***	-6.67**
	30	0	0	-23****	0	-6*	-10****	-36****
	40	0	0	-30.33****	-2.67	-34.33****	-20.33****	-53.33****
	50	0	0	-37****	-17****	-42****	-86.67****	-54.33****
	60	0	-0.33	-40****	-38.33****	-63****	-98****	-56****
	70	-3.333	-2.33	-38.67****	-65****	-56.33****	-82.33****	-41.67****
	80	-3	0	-5	-5	1.667	-17.33****	19.33****
	90	-5	-1.667	-7	-7	-7.33***	-12.33****	12.33****
	100	-0.33	0.667	-2.33	-2.33	-2.33	-5.33*	-2.33
IC_{50}		74.18	74.67	55.15	61.35	51.82	43.92	46.77
95% Likelihood Interval		[73.67-74.68]	[73.95-75.40]	[49.55-60.41]	[60.29-62.40]	[50.09-53.56]	[43.05-44.81]	[42.55-51.14]

Table XIX- Statistical analysis of Moxifloxacin assay - Means and mean differences were determined by a two-way ANOVA test of three independent biological repeats. A Dunett's multiple comparisons test was used to compare means variation between strains at each concentration. Individual p-values are displayed as stars. Na = Not attributed. IC_{50} and 95% likelihood intervals were determined by non-linear regression.

The last group of strains (Δ*uvrB* and Δ*dnaE2*) were also more sensitive to the treatment than the wild ype (respectively IC_{50} of 55.15 ng/mL and 46.77ng/mL) (Table XIX). 4-Nitroquinolone is potent inducer of lethal DNA lesions and thus, is a known inducer of NER in other organisms (Figure XXXV).[221] Similar to the moxifloxacin results, three different sensitivity phenotypes were observed: deletion of *uvrC* did show a small yet reproducible protective effect; the SOS mutants (*recA*, *lexAND*) and Δ*dnaQ-uvrC* exhibited

a modest, yet significantly increased sensitivity to the treatment; while, deletion of *uvrB* was again the most sensitive (Table XX).

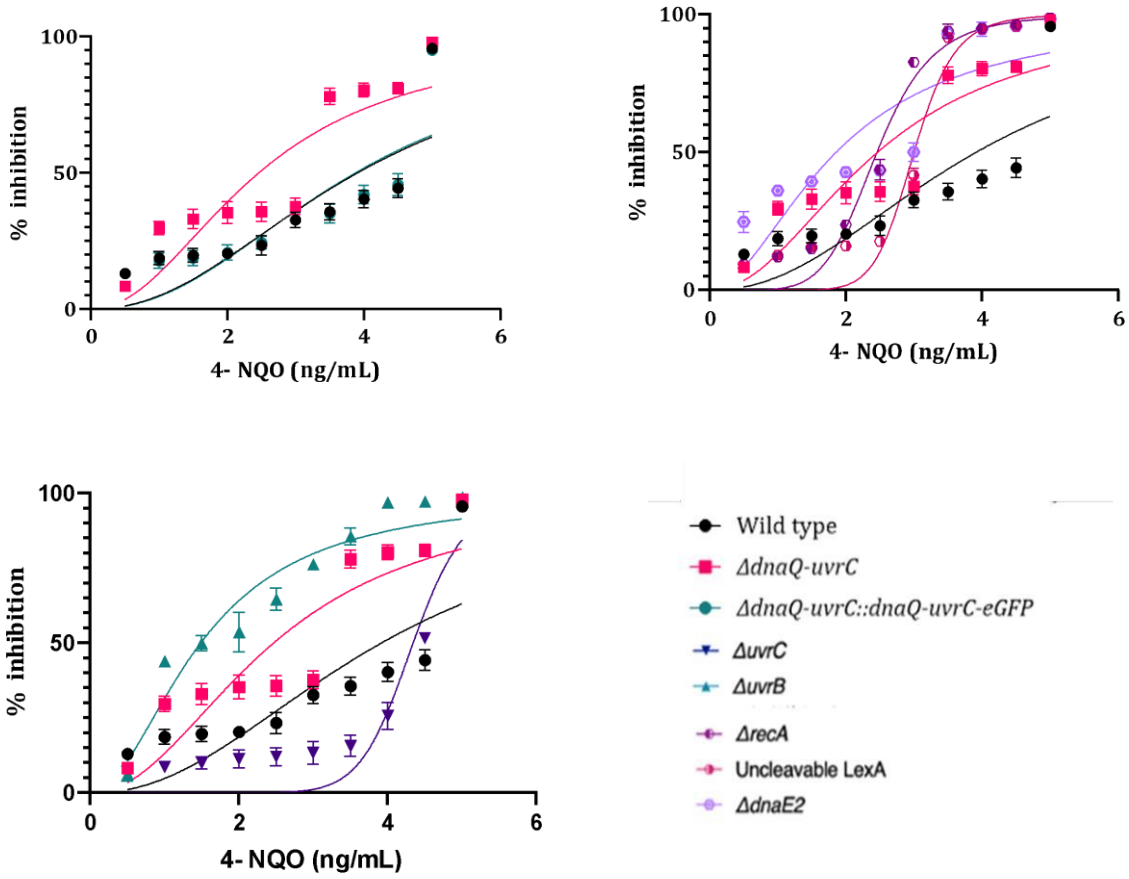


Figure XXXV- Inhibition of *Msm* strains in response to 4-Nitroquinoline. Top left panel highlights *ΔdnaQ-uvrC* strain and the complemented one. Top right panel highlights NER genes deletion strains. Bottom left panel highlights SOS response and mutagenesis genes deletion strains. The percent of inhibition to Moxifloxacin treatment was obtained after in liquid growth, performed as described in Material and Methods. Results were obtained following rezasurin coloration and optical density measurement. Values were then normalized to wild type values. Nonlinear fitted curves were drawn using the [inhibitor] vs normalized response —Variable slopes— tool of Graph Pad Prism.

	4-Nitroquinoline concentration (ng/mL)	*ΔdnaQ-uvrC*	*ΔdnaQ-uvrC::dnaQ-uvrC-eGFP*	*ΔuvrB*	*ΔuvrC*	*ΔrecA*	*LexAND*	*ΔdnaE2*
Mean Difference with Wildtype (%)	0.5	4.67	0	-6.67**	6.67**	3.67	5	-11.67****
	1	-11****	1	-25.33****	10****	6**	6.67*	-17.33****
	1.5	-13.33****	1	-30.33****	9.67****	4.67	4	-19.67****
	2	-15****	0.33	-33.33****	9****	-3.33	4.33	-22.33****
	2.5	-12.33****	-1	-41.33****	11.33****	-20.33***	5.67*	-20***
	3	-5	0	-43.67****	19.33****	-50****	-9****	-17.33****
	3.5	-42.33****	0.667	-50****	20****	-58.33****	-56****	-57.33****
	4	-40****	-1.667	-56.67****	14.67****	-54.67****	-54.33****	-54.33****
	4.5	-36.67****	-1.33	-53****	-7.33**	-52****	-51.33****	51.33****
	5	-2.33	0.667	-3	-1.33	-3	-3	-2.67
IC_{50}		2.467	3.885	1.487	4.344	2.458	3.002	1.860
95% Likelihood Interval		[2.119-2.8]	[3.448-4.568]	[1.328-1.65]	[4.199-4.47]	[2.353-2.558]	[2.877-3.114]	[1.456-2.256]

Table XX- Statistical analysis of 4-Nitroquinoline assay-Means and mean differences were determined by a two-way ANOVA test of three independent biological repeats. A Dunett's multiple comparisons test was used to compare means variation between strains at each concentration. Individual p-values are displayed as stars. Na = Not attributed. IC_{50} and 95% likelihood intervals were determined by non-linear regression

Griselimycin binds the *dnaN*-encoded encoded β-sliding clamp at the interaction site with the PolIIα subunit of the replisome or with other proteins[41]. When treated with that drug, the strains exhibited no significant phenotypes except the Δ*uvrB* strain that was more sensitive to the treatment than the wildtype (Figure XXXVI & Table XXI).

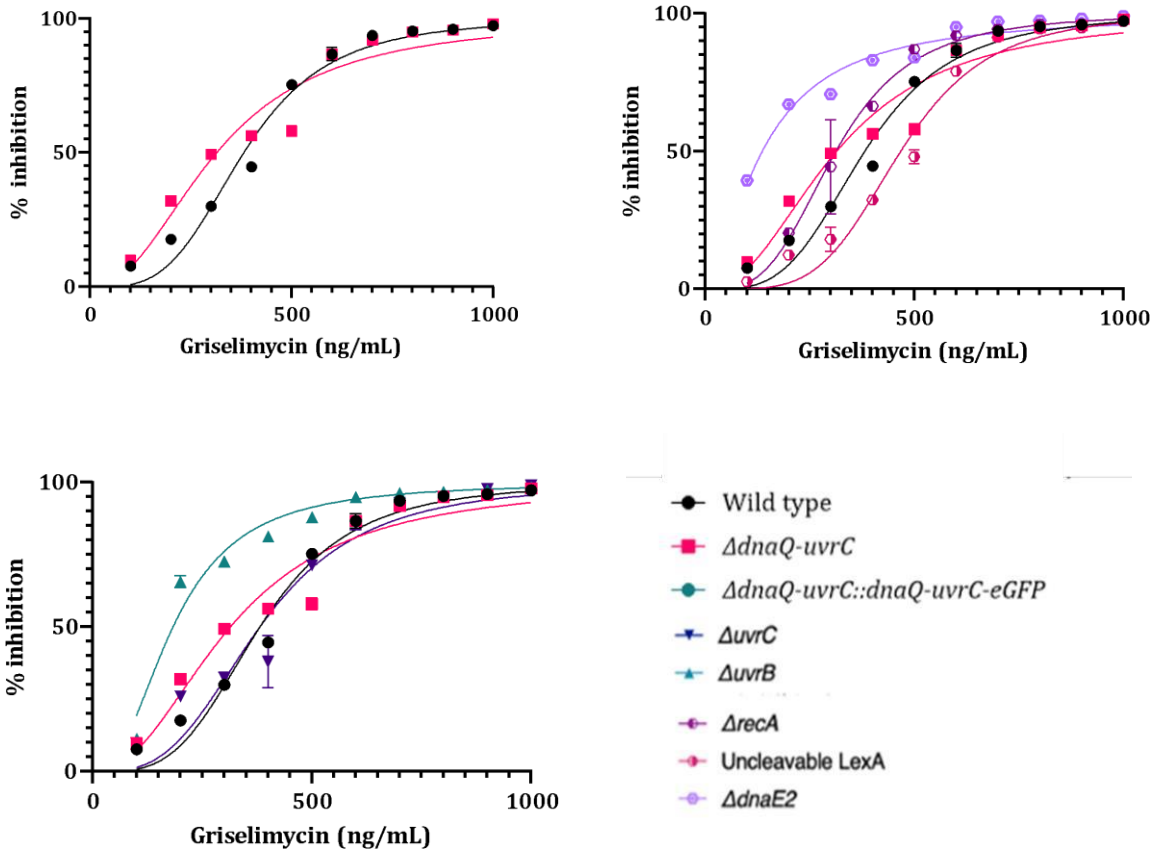


Figure XXXVI- Inhibition of *Msm* strains in response to Griselimycin. Top right panel highlights *ΔdnaQ-uvrC* strain and the complemented one. Top left panel highlights NER genes deletion strains. Bottom left panel highlights SOS response and mutagenesis genes deletion strains. The percent of inhibition to Griselimycin treatment was obtained after in liquid growth, performed as described in Material and Methods. Results were obtained following resazurin coloration and optical density measurement. Values were then normalized to wild type values. Nonlinear fitted curves were drawn using the [inhibitor] vs normalized response—Variable slopes— tool of Graph Pad Prism.

	Griselimycin concentration (ng/mL)	*ΔdnaQ-uvrC*	*ΔdnaQ-uvrC::dnaQ-uvrC-eGFP*	*ΔuvrB*	*ΔuvrC*	*ΔrecA*	*lexAND*	*ΔdnaE2*
Mean Difference with Wild type (%)	100	-2	0	-3.67	0.33	0	5	-31.67****
	200	-14.33****	0	-48****	-8.33**	-2.67	5.33	-49.33****
	300	-19.33****	0	-42.67****	-2.33	-14.33****	12****	-40.67****
	400	-11.67****	-1.667	-36.67****	6.67*	-21.67****	12.33****	-38.33****
	500	17.33****	0.33	-12.67****	4	-11.67***	27.33****	-8.67***
	600	0	0	-8.33**	1	-5.33	7.67**	-8.33**
	700	-1.67	0.667	-2.67	0.67	-0.67	2.33	-3.33
	800	0.33	-0.33	-1.33	0.33	-0.67	1	-2
	900	0.33	0	-1.33	-1.67	-0.33	1.33	-2
	1000	-0.67	0	-1	-1.67	0	0.33	-1.67
IC_{50}		314.6	382	1.487	183.8	382	312.6	467.8
95% Likelihood Interval		[286.8-342.5]	[364.5-399.2]	[1.328-1.65]	[171.4-196]	[350.3-412]	[296.4-328.7]	[449.1-486]

Table XXI- Statistical analysis of Griselimycin assay- Means and mean differences were determined by a two-way ANOVA test of three independent biological repeats. A Dunett's multiple comparisons test was used to compare means variation between strains at each concentration. Individual p-values are displayed as stars. Na = Not attributed. IC_{50} and 95% likelihood intervals were determined by non-linear regression.

5-2-7- UvrB is mandatory to deal with various DNA damages:

Figure XXXVII shows the IC_{50} of every strain against all the drugs tested; values are presented as a percentage of wildtype IC_{50}. Strains that had an unmarked knock-out of *uvrC* or *dnaQ-uvrC* never exhibited the same phenotype and even opposite phenotypes across treatments. Furthermore, deletion of *uvrC* appeared to result in a slight but

statistically significant protective effect under MMC and 4-NQO treatments. The complemented strain managed to restore the wildtype phenotype in most cases. However, some issues were noted with 4-NQO treatment. The gene deletions that had the greatest impact on phenotypes across treatments were *dnaE2* and *uvrB* deletions (Table XIX to XXI and Figure XXXVII).

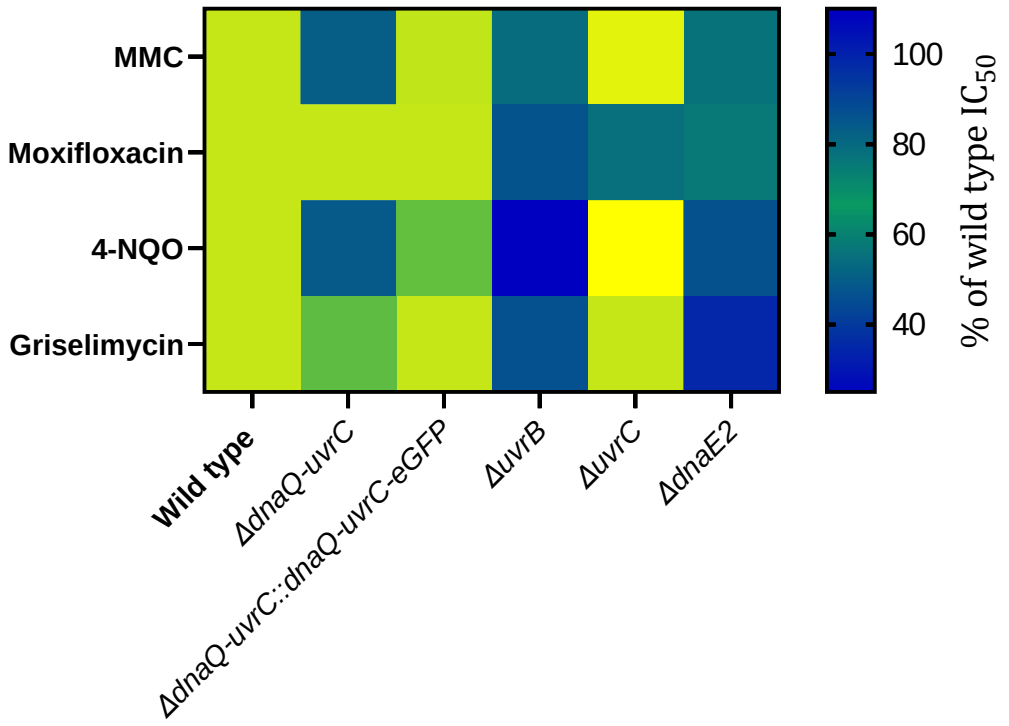


Figure XXXVII- IC_{50} of the *Msm* strains for the tested drugs relative to the Wild type strain. The scale gives the correspondence between color and percentage of wild type IC_{50} for *ΔdnaQ-uvrC,* complemented strain (*ΔdnaQ-uvrC::dnaQ-uvrC-eGFP)*, *ΔuvrC* and *ΔuvrB.* IC_{50} value were calculated using the Nonlinear fitted curves tool using the [inhibitor] vs normalized response —Variable slopes— on GraphPad Prism.

5-2-8- Use of CRSIPR interference to confirm results:

To confirm previous results for the NER genes, two CRISPRi constructs were generated to silence either *uvrC* or *uvrB*. As shown in Figures XXXVIII, XXXIX and Table XXII, similar results were observed for both strains when ATc was added to the medium at the concentration of 100 ng/mL to induce the CRISPRi system. Without proper

induction, CRISPRi strains exhibited similar behaviour to the wild type. To confirm the integrity of the results, the strains carrying CRSIPRi plasmids were sequenced (Supplementary Figure II).

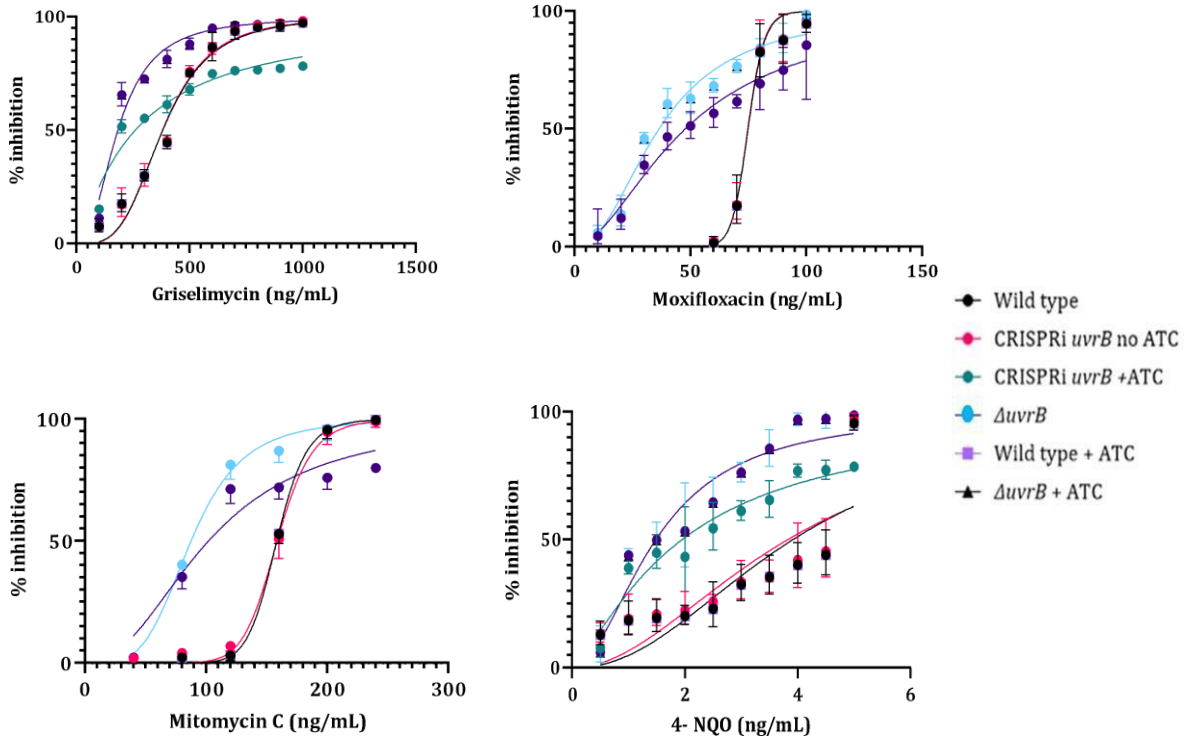


Figure XXXVIII- Inhibition of *uvrB*-CRISPRi strain. Top left panel displays treatment by griselimycin. Top right displays treatment by moxifloxacin. Bottom left displays MMC treatment. Bottom left displays 4-NQO. Δ*uvrB*. strain is added as a comparison. ATC was at 100 ng/mL. The percents of inhibition to were obtained after in liquid growth, performed as described in Material and Methods. Results were obtained following resazurin coloration and optical density measurement. Values were then normalized to wild type values.

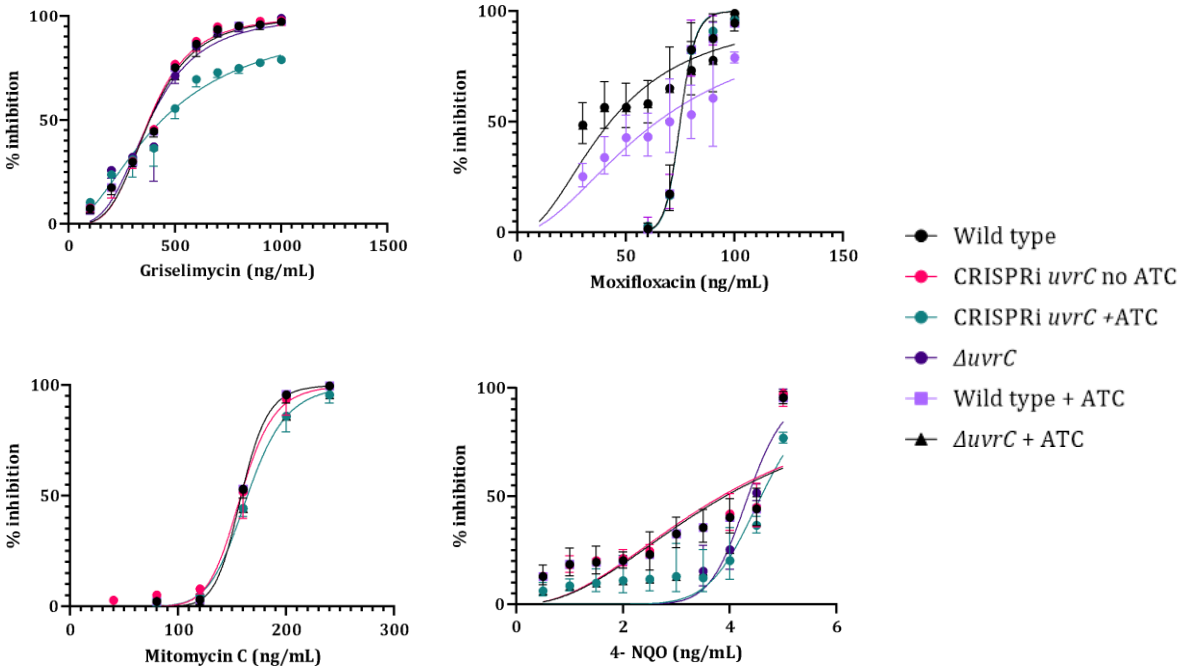


Figure XXXIX- Inhibition of *uvrC* - CRISPRi strain. Top left panel displays treatment by griselimycin . Top right displays treatment by moxifloxacin. Bottom left displays MMC treatment. Bottom left displays 4 - NQO. Δ *uvrB* . strain is added as a comparison. ATC was at 100 ng/ m L. The percents of inhibition to were obtained after in liquid growth, performed as described in Material and Methods. Results were obtained following resazurin coloration and optical density measurement. Values were then normalized to wild type values.

Drugs	Strains	Wild type	CRISPRi *uvrB*		*ΔuvrB*	CRISPRi *uvrC*		*ΔuvrC*
			Atc +	Atc -		Atc +	Atc -	
Mitomycin C	**IC_{50}**	158.4	95.89	159.2	87.7	165.7	157.2	164.1
	95% Likelihood Interval	[157.5-159.3]	[156.5-161]	[156.5-161]	[84.93-90.46]	[162.4-165.8]	[155-160.8]	[162.4-165.8]
Moxifloxacin	**IC_{50}**	74.94	47.93	74.83	36.55	64.52	74.98	42.79
	95% Likelihood Interval	[74.13-75.75]	[45.1-50.8]	[74.03-75.6]	[34.37-38.75]	[59.9-69.63]	[74.37-75.6]	[37.99-47.8]
4-NQO	**IC_{50}**	3.916	1.950	3.824	1.487	4.59	3.845	4.344
	95% Likelihood Interval	[3.451-4.7]	[1.790-1.498]	[3.33-4,461]	[1.328-1.65]	[4.446-4.754]	[3.377-4.588]	[4.199-4.476]
Griselimycin	**IC_{50}**	382	261.1	381.1	183.8	439.9	377.9	382
	95% Likelihood Interval	[364.5-399.2]	[237.3-284.9]	[364.1-397.8]	[171.4-196.4]	[416-463.8]	[360.1-395.3]	[350.3-412.6]

Table XXII- IC_{50} for MMC treatment and 95% likelihood intervals (L.I) of *Msm* Strains. Values shown represent IC_{50} were generated with a non-linear regression of [inhibitor]VS normalized response and 95% likelihood intervals of experiments.

5-2-9- Deletion of *uvrB* inhibits the evolution of antibiotic resistance:

Mutagenesis has been implicated in the development of drug resistance *in vivo,*[68,102] as indicated in Section 5-1-2. 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-clamp via a β-clamp binding motif.[102] To investigate the potential link between *uvrB* and *dnaE2* in the mutagenesis pathway, an antibiotic evolution assay was designed based on

the previous work of Ragheb et al.[69] This assay was performed as described in the Materials and Methods section. Cells were cultivated in the presence of drugs in a 96-well plate, similarly to what was done in Section 5-2-1 to 5-2-8. Although, for this assay, mutant strains that grew to ~50% relative to wildtype in the no drug control were passaged to a repeat of the assay. A total of six passages were performed and each assay had six independent biological repeats.

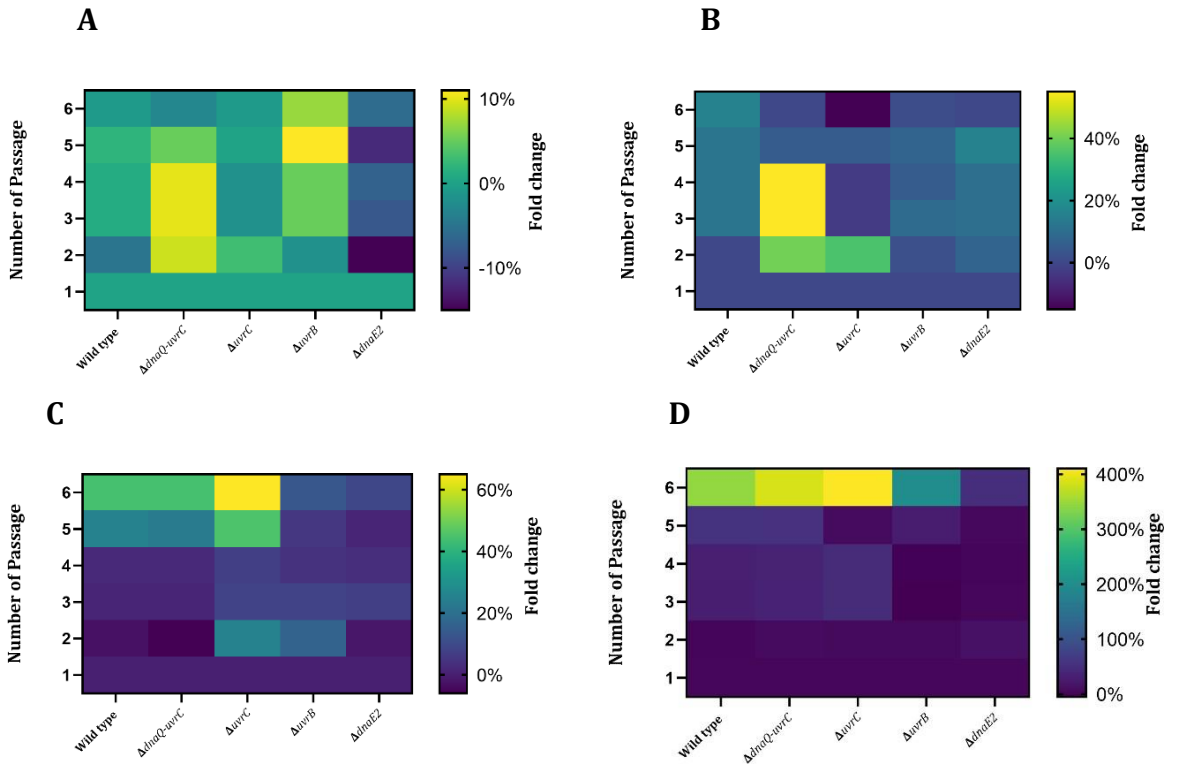


Figure XL- Antibiotic evolution assay-A-4-Nitroquinoline assay. B-Mitomycin C assay. C-Moxifloxacin assay. D-Rifampicin assay. Plot shows IC_{50} for strains at each passage. Statistical significance was determined using a two-tailed Mann-Whitney U test (∗p value < 0.10). Number of replicates for each strain of *Msm* is 6.

Strains were grown in the presence of MMC, moxifloxacin or rifampicin. Rifampicin is the frontline anti-TB drug; it inhibits RNA synthesis by the DNA-dependent RNA polymerase and it the main antibiotic used to study mutagenesis and antibiotic adaptation. It is the only antibiotic used in this study that does not directly target the DNA or the replisome. However, the acquisition of resistance to rifampicin is commonly used to study mutagenesis.[34] The IC_{50} was determined using the percent inhibition that was calculated based on normalised data relative to wild type. Data were generated by measuring the

$OD_{600?}$ with FLUOstar OPTIMA plate reader (BMG Labtech). Treatment with MMC or 4-NQO generated similar results. Regardless of individual strain sensitivities to the drug tested, the IC_{50} remained the same across passages. When treated with rifampicin, it was noticed that, around the fifth passage, the IC_{50} was increased for all strains except *dnaE2* KO strain. Interestingly, though, the strain with a deletion of *uvrB* reduced the increase of the IC_{50} while not preventing it. In the case of moxifloxacin, the adaptation phenotype was present but not as strong as for the rifampicin treatment. As previously observed, Δ*uvrB* is mandatory to face the damage caused by moxifloxacin. Moreover, its deletion completely inhibits the adaptation, similarly to *dnaE2* KO strain (Figure XL & Table XXIII).

Drugs		Mitomycin C				4-Nitroquinoline			
Strains		Δ*dnaQ-uvrC*	Δ*uvrC*	Δ*uvrB*	Δ*dnaE2*	Δ*dnaQ-uvrC*	Δ*uvrC*	Δ*uvrB*	Δ*dnaE2*
PASSAGE	**1**	-74.11****	9.65****	-66.7****	-66****	-2.023	0.155	-2.5	4
	2	-77.9****	6.01****	-70.5****	-70.2****	-1.9	0.32	-2.6	3.99
	3	-78.2****	6.01****	70.5****	-70.3****	-1.82	0	-2.4	4.027
	4	-78.2****	5.2****	-70.7****	-71.1****	-.1967	-0.4	-2.786	3.227
	5	-77.8****	5.4****	-70.7****	-71.2****	-2.117	0	-2.53	3.186
	6	-78.2****	5.23****	-70.9****	-7.9****	-2,16	0.2	-2.7	3.42

Drugs		Rifampicin				Moxifloxacin			
Strains		Δ*dnaQ-uvrC*	Δ*uvrC*	Δ*uvrB*	Δ*dnaE2*	Δ*dnaQ-uvrC*	Δ*uvrC*	Δ*uvrB*	Δ*dnaE2*
PASSAGE	**1**	-0.117	-0.08	-0.1	-0.08	-0.793	-32****	-38.9****	-28.9****
	2	0.17	-0.01	-0.07	-0.08	-0.8	-32****	-39****	-29****
	3	-0.5	-0.03	-4.1****	-4.567****	-0.8	-32****	-39****	-29****
	4	0.03	0.03	-3.67****	-4.55****	-0.8****	-32****	-39.3****	-29****
	5	-0.16	0.16	-4.2****	-7.23****	-0.8	-32****	-44.5****	-34.5****
	6	-0.31	-0.9	-6.83****	-39.63****	-6.8****	-44****	-62.9****	-52.8****

Table XXIII- Statistical analysis of antibiotic evolution assay-Means (ng/mL) and mean differences were determined by a two-way ANOVA test of three independent biological repeats. A Dunett's multiple comparisons test was used to compare means variation between strains at each concentration. Individual p-values of 95% confidence intervals are displayed as stars (<0.05= *; <0.005=**; <0.0005=***; <0.0001=****. Na = Not attributed.

5-2-10- Deletion of *uvrB* decreased mutation frequency following UV exposure:

The standard mutagen to study mutagenesis is UV radiation because of its convenience of handling and ubiquity of distribution in samples, unlike chemicals that can be non-uniformly distributed in liquids. UV induces thymine dimers that prevent the replication from being performed.[151] Indeed, classical polymerases cannot work past obstacles such as thymine dimers. If these obstacles cannot be solved, TLS polymerases are summoned as they can replicate DNA past thymine dimers.[182] However due to their nature, they can generate errors -mutations- that can lead to the uprising of drug resistant bacteria.[124] In this work, UV was used in two distinct assays. The data presented in Figure XLI is an adaptation of the Luria-Delbrück fluctuation assay that measures the probability of a bacteria sustaining a mutation (mutation/base pair/generation) in the *rpoB* locus. As mentioned in section 4-3-1, the fluctuation assay set-up strictly requires that the bacteria are exposed to the mutagenic agent (UV or antibiotic) constantly during the experiment; therefore, the use of a once-off/transient perturbation such as UV treatment contravenes a core principle of the fluctuation assay. Since the experiments carried out in this work were exposed to UV at a single time point (as described in section 2.8.4), they will be referred to as "mutation frequency" assays. Bacteria following the UV exposure are allowed to rest in 7H9 medium with no drugs and) they are parallel plated on Rif and no drug plate after 4 hours of recovery. For this experiment the survival is measured after counting colonies on no drug plate (Supplementary Table V). The rates of spontaneous acquisition of RIF resistance was then used as a mutation frequency approximation.

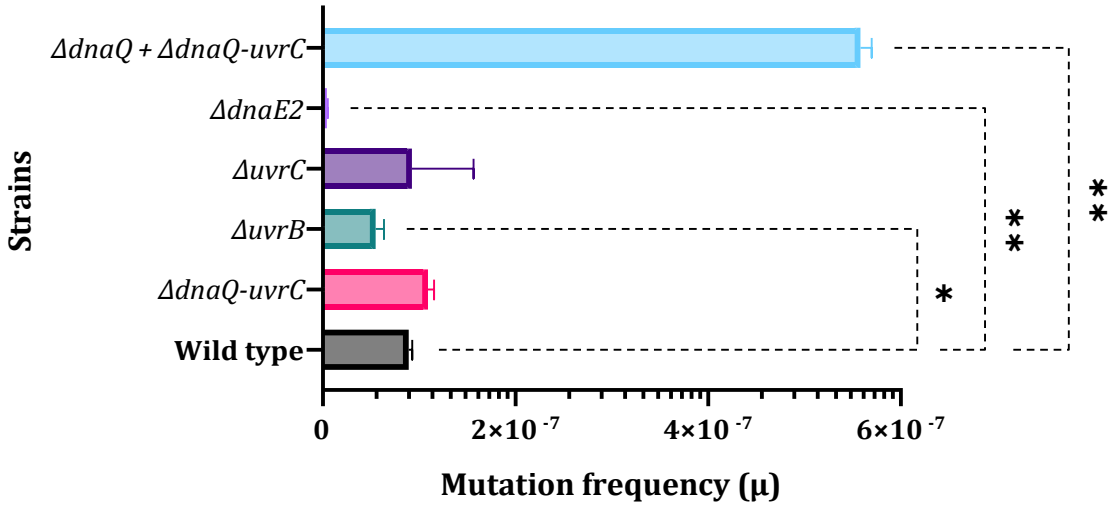


Figure XLI- UV-induced mutation frequency. Mutation frequency was calculated with the software developed in the MMRU by Z. Martin and J. Kent (unpublished) based on the results of three independent experiments. Experiment was carried out as an adaptation of the Luria Delbruck fluctuation assay described in the materials and methods.

The experimental design utilized a total of 3 parallel cultures per strain in each of the repeats of the adaptation of the mutation frequency assay, and the number of RIF resistant mutants was enumerated by plating on rifampicin solid medium. For this experiment, the survival was measured after counting colonies on no drug plate (Supplementary Table V). Deletion of *uvrC* did not result in a significant fold change in the mutation frequency (Figure XLI & Table XXIV). *ΔdnaE2* strain was used as control; Boshoff et al[58] showed that *dnaE2* is critical to UV induced-mutagenesis thus, its deletion was expected to significantly decrease the mutation rate. Indeed, *ΔdnaE2* exhibited only 4.5% of the wild type mutation rate. Deletion of *uvrB* resulted in $6.12E^{-1}$ decrease in the fold change compared to the wild type. On the other hand, the double *dnaQ dnaQ-uvrC* knock-out resulted into 6 times increase in the mutation rate fold change.

Strains	*ΔdnaQ-uvrC*	*ΔuvrB*	*ΔuvrC*	Double KO	*ΔdnaE2*
Mean of Fold change relative to wild type	1.21e0	6.12e-1*	1.06	6.27***	4.59e-2**

Table XXIV- Mean of fold change in the mutation rate for the adaptation of the Luria Delbruck fluctuation Assay. The fold change in the mutation rate was calculated by dividing the mutation rate in the presence of mutant strains by the mutation rate of the wild type. Significance was calculated by a Brown-Forsythe and Welsh Anova test.

5-2-11- Temporal activity of NER proteins correlated to the growth phase:

The adaptation of the Luria-Delbrück fluctuation assay as performed previously has inherent limitations. Indeed, it only gives an overview of the phenotypic state of the organism at a single time-point. However, it was shown by Gill *et al*[97] that the replication dynamic is not in a static equilibrium but rather can vary in a temporal manner, depending on the bacterial state or resources availability. Thus, another assay was conducted. This assay displayed in Figure XLII is a similar mutation frequency assay but this time; a temporal factor was added to the experiment. In other words, bacteria following the UV exposure are allowed to rest in 7H9 medium with no drugs and every 4 hours (up to 24 hours) they are parallel plated on Rif and no drug plate. For this experiment the survival is measured after counting colonies on no drug plate (Supplementary Table VI). The rates of spontaneous acquisition of RIF resistance were estimated at several time points. The experimental design utilized a total of 6 parallel cultures per strain in each of the repeat of the assay, and the number of RIF resistant mutants was enumerated by plating on rifampicin solid medium. The spontaneous mutation rates were determined for each strain relative to wild type. At four hours, similar results to first assay were observed (Table XXV). However, over time, variations occurred. Indeed, the mutation of *ΔdnaQ-uvrC* relative to the wild type went from being superior at eight hours to being inferior at 16 hours. The double knock-out strain had a far superior mutation rate at the beginning of the experiment that became inferior to the wild type around 16 hours. Deletion of *uvrC*

resulted in a more important mutation rate than the wild type but only after 24 hours. The negative control Δ*dnaE2* exhibited a very low mutation rate during the full 24 hours, confirming its role as the main driver of mutagenesis. The contribution of *uvrB* to the mutation rate remained very low too.

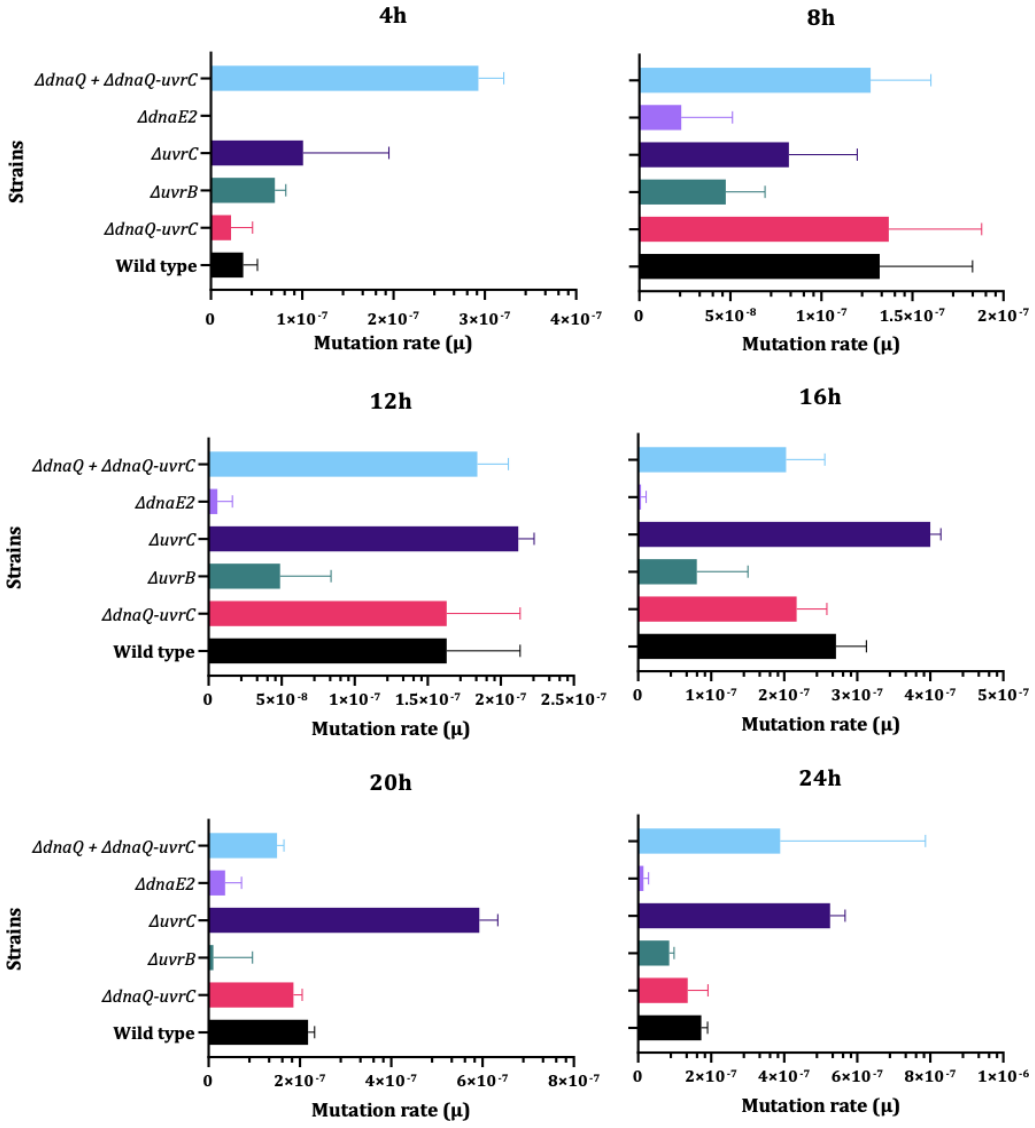


Figure XLII- Temporal UV- induced mutation frequency. Mutation frequency were calculated as indicated in Materials and Methods, based on the results of three independent experiments. Experiment was carried out as a UV-induced mutation frequency assay described in the materials and methods over a period of 24 hour

Strains	Time Point	*ΔdnaQ-uvrC*	*ΔuvrB*	*ΔuvrC*	*Double KO*	*ΔdnaE2*
Mean of Fold change	4h	5.05E-01 *	2.16	2,33	9.85 ***	0
	8h	1,28	4.52E-1	7.82E-1	8.49E-1	2.78E-1
	12h	1,14	3.70E-1	1.41	1.24	3.91E-2
	16h	7.98E-1	2,74E-1 *	1.49	7.39E-1	1.91E-2 ***
	20h	8.52E-1	1.73E-1 **	2.72 ****	6.93E-1	2.34E-1*
	24h	7.70E-1	4,92E-1	3.08 ****	2.13 **	4.08E-2

Table XXV- Mean of fold change in the mutation rate for the adaptation of the Luria Delbruck fluctuation Assay.. The fold change in the mutation rate was calculated by dividing the mutation rate in the presence of mutant strains by the mutation rate of the Wild type. Significance was calculated by a Brown-Forsythe and Welsh Anova test.

5-3- Discussion:

5-3-1- Nucleotide Excision repair in mycobacteria:

Nucleotide Excision Repair (NER) is well conserved mechanism of paramount importance to maintain genomic integrity at all times. Its importance is even greater in pathogenic organisms that are constantly harassed by the host immune system. Its greatest strength is its ability to deal with a broad range of DNA damage in and out of the replication. The UvrA, UvrB and UvrC were considered so far, as the main proteins involved. Regarding many aspects of its metabolism, *Mtb* and mycobacteria in general, are fundamentally different from other organisms used as prokaryotic model. Previous work performed at the MMRU identified a natural hybrid protein, DnaQ-UvrC that had its C-terminal sharing homology with the canonical UvrC. Its deletion resulted in a hyper-sensitivity to MMC. Consequently, this chapter focused on the possibility that an external factor might contribute to mycobacterial NER.

5-3-2- Confirmation of DnaQ-UvrC requirement for MMC-induced DNA damage repair in *in vitro* experiments:

Before a role for DnaQ-UvrC in NER could be investigated, previous work performed within the MMRU by Zanele Ditse had to be repeated. Indeed, a hypersensitivity phenotype for *dnaQ-uvrC* knock-out strain was identified, and it was reversed by deletion of *dnaQ* in the same background. Even though, the hypersensitivity phenotype was confirmed, the work performed here is not withstanding by the phenotype reversion by *dnaQ* deletion. On the contrary, the MMC hypersensitivity remained even with the deletion of both DnaQ homologs genes. However, PCR screening of the strains was performed and ascertained their genotype.

5-3-3- MMC-induced damage repair is dependent on interaction with the replisome and UvrB:

The data generated in Chapter 3, suggested a possibility for DnaQ-UvrC to interact with either the β-clamp or UvrB. The MMC hypersensitivity was used to explore both options. Indeed, DnaQ-UvrC variants were generated to investigate the contribution of

both interacting domains to the phenotype observed. Inactivation of each domain resulted in an intermediate sensitivity phenotype. Disruption of both sites did mimic the KO phenotype. If both sites seem important to DnaQ-UvrC activity, this sensitivity seemed stronger when the UVR domain was impaired. Because of replication and NER are not triggered the same way and because of the temporal limitation of the experiments, these findings, however, unprecedented, do not allow to draw more subtle conclusions. Nevertheless, the conclusions drawn were in accordance with the *in-silico* study performed in Chapter 3. Even though the exact role of DnaQ-UvrC remain elusive, the finding of this study links it to DNA damage repair which is in accordance with its homology to UvrC and to the bioinformatic data displayed in Chapter 3. Consequently, its role in UvrABC-related pathway was studied.

5-3-4- No apparent role for UvrC for MMC and UV-induced damage repair:

At end of the Chapter 3, it was wondered if UvrC was necessary in mycobacteria. However central, NER has not been thoroughly studied outside of *E. coli* and most of the work focus on UvrB and its ability to scan and detect damage in dsDNA.[188–191,224] Nevertheless, UvrA, UvrB and UvrC are supposed to be able to deal with any kind of damage. Since UV and MMC targets DNA, NER genes deletion should result in a MMC sensitivity. Surprisingly, the knocked-out *uvrC* strain generated in this study did not differ from the wild type phenotype. It even had a small but significant protective effect at high doses. (two-way Anova on means, p-value <0,0001). Lack of sensitivity was also observed after UV-exposure for both *ΔdnaQ-uvrC* and Δ*uvrC* suggesting a potential overlapping activity in the specific DNA-damaging condition that is UV. PCR screening of the *Msm uvrC* KO strains confirmed the phenotype of this strain. This apparent inconstancy regarding *uvrC* expected role in other organism requires that this result is handled with caution. To ascertain these results, a Tet-inducible CRISPRi targeting *uvrC* was generated. It confirmed the unmarked knocked-out strain results. The protective effect at high doses for MMC could also be detected.

5-3-5- Putative damage specificity of mycobacterial NER incision nucleases:

Expression regulation of DnaQ-UvrC and UvrC advocates for a damage specific NER in mycobacteria. Indeed, the expression of the natural hybrid protein is driven by the PafBC regulated DNA damage response while UvrC expression is constitutive in mycobacteria. This could explain the protective effect that resulted from *uvrC* deletion during MMC and 4-NQO treatments. Indeed, the deletion may allow more UvrB proteins to be available and bind DnaQ-UvrC. This might suggest a competition between DnaQ-UvrC and UvrC Tto bind UvrB. In the case of UV-induced DNA damage, a combined absence of phenotype for both *ΔdnaQ-uvrC* and *ΔuvrC* strains was observed. This questions a potential counterbalancing between each other activity for UV-induced DNA damaging conditions. For most of the drug tested, DnaQ-uvrC seems to be the main protein involved for the damaged nucleotide removal. Except for the Moxifloxacin treatment, in which deletion of *uvrC* exhibited a stronger sensitivity *than ΔdnaQ-uvrC*.

5-3-6- UvrB is the central actor of NER:

As mentioned earlier, NER is able to repair DNA damages in an unspecific way. In some cases, it is even able to act as a BER backup.[156] UvrB is the link that allows to switch from damage detection to its resolution. Without it, no damaged DNA incision can be performed. In order to avoid unneeded DNA cleavage, its expression is dependent on the PafBC regulated DNA damage response. Data generated in this work shows that UvrB is mandatory to deal with any kind of damage sustained by DNA. Indeed, its deletion systematically resulted in a sensitivity phenotype. The fact that UvrB is critical in NER is not new, therefore, the data generated in this work are in accordance with previous research and is reproducible for a broad range of DNA-damaging conditions.

5-3-7- Deletion of *uvrB* inhibits the evolution of antibiotic resistance:

Mutagenesis was proven to be implicated in the development of drug resistance *in vivo*. 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-clamp via a β-clamp binding motif.[102] To

investigate the potential link between *uvrB* and *dnaE2* in the mutagenesis pathway, an antibiotic evolution was designed based on the previous work of Ragheb et al. the assay was performed as described in the material and methods section. By selecting cultures that survived the highest drug concentration and serially passaging them, it was hoped to interrogate their role in the adaptation to antibiotic treatments. No adaptation was observed when bacteria were treated with MMC or 4-NQO, even for the wild type. It was likely these drugs didn't allow any adaptation from the strains because of the various effects on DNA they have. For the other drugs (namely, moxifloxacin and rifampicin) tested, deletion of *uvrB*, similarly to the deletion of *dnaE2* completely prevents bacteria to adapt to the drugs across passages, highlighting a putative role in adaptative mutagenesis from UvrB.

5-3-8- Deletion of *uvrB* reduces the mutation frequency following UV exposure:

The standard mutagen to study organism mutagenesis is UV radiation because of its convenience of handling and ubiquity of distribution in samples, unlike chemicals that can be non-uniformly distributed in liquids. UV induces thymine dimers that prevent the replication from being performed. Indeed, classical polymerase cannot work past obstacles such as thymine dimers. If these obstacles cannot be solved, translesion polymerases are summoned as they can replicate DNA past thymine dimers. However due to their nature, they will generate a lot of errors -mutations- that can lead to the uprising of drug resistant bacteria.[124] The mutation rate is usually estimated by the Luria-Delbrück fluctuation assay. The spontaneous mutation rates were determined for each strain relative to wild type. Deletion of *uvrB* resulted in a small but significant decreasing in the mutation rate. On the other hand, *ΔdnaE2* exhibited only 4.5% of the wild type mutation rate. Boshoff et [68] showed that *dnaE2* is critical to UV induced-mutagenesis thus, its deletion was expected to significantly decrease the mutation rate. These two assays confirmed UvrB potential role in mutagenesis, however, the adapted Luria-Delbrück fluctuation assay has a number of limitations. It gives an overview of the phenotypic state of the organism at a single time-point. However, it was shown by Gill et al, that the replication dynamic is not in a static equilibrium but rather can vary in a temporal

manner, depending on the bacterial state or resources availability. Thus, another kind of assay was conducted that include a temporal factor.

5-3-9- Temporal activity of NER correlated to the growth phase:

Adding a temporal factor to experiments is a way to consider the bacterial growth as a parameter of a gene's activity. The classic Luria-Delbrück fluctuation assay does not consider the time parameter which is why the UV-induced mutation frequency was designed. During the first part of the experiment, genes involved in the replication seem to have a greater impact on the mutation rate, relative to the wild type. However, starting at 12 hours until the end of the experiments *uvrC* knock-out is taking an increasing importance until a maximum is reached at 24 hours. As the main driver of mutagenesis, DnaE2 has the greater impact of the mutation rate. Indeed, its deletion almost inhibit it. Similarly, *uvrB* knock-out had its role in mutagenesis confirmed in this assay.

5-3-10- Synergy between DnaQ homologs is dependent on the replication state:

Adding a temporal factor to experiments is a way to consider the bacterial growth as a parameter of genes activity, which is not considered by the classic Luria-Delbrück fluctuation assay. *M. smegmatis* is a slow growing bacterium with about a 4-hour doubling time. The replication is made slower in DNA damaging conditions or during antibiotic treatment.[97] During the first 4 hours of the experiment, the replication is fully operational which results in an increase of the mutation rate for the strains exhibiting the double *dnaQ dnaQ-uvrC* knock-out. At 8 hours, the replication seemed to be reduced. Indeed, the mutation rate of DnaQ homologs did not differ from wild type. Similarly, for the rest of the time points, the replication appeared to be on hold as there was no further significant variation in the mutation. However, at 24 hours, the replication is starting to resume as the double *dnaQ dnaQ-uvrC* knock-out background exhibit about a 40-fold increase in the mutation rate. These results however, statically significant and consistent across repeats must be considered carefully because of the inherent limitation of the assay. Moreover, the temporal factor is not the most precise proxy to measure the replicative state of bacteria. However, apparent correlation between replication state and *dnaQ dnaQ-uvrC* knock-out strain mutation frequency gives credit to the potential synergy of DnaQ homologs to maintain the replication fidelity in the condition of the assay.

5-3-11- Revision of mycobacterial NER model:

Phylogenetic analysis revealed highly conserved NER proteins among bacteria[155] and thus the model organism *E. coli* could largely be used to generate new knowledge about this mechanism. Even though most bacteria exhibit only the highly conserved UvrC, some species, like Actinobacteria were shown to possess both a canonical UvrC along with a DnaQ-UvrC protein. It is not clear what its role is yet, but it seems to be required to deal with some damages generated by DNA-damaging drugs. Moreover, its activity was shown to be dependent on both its binding to the β-clamp and UvrABC interactions. Two hypotheses about DnaQ-UvrC can be considered: either it's a part of DNA-damage specific pathway of NER which is not consistent with the current literature, or it plays a role in TCR which is a NER pathway occurring during the replication. Given the critical importance of maintaining genome for pathogenic organism and given that DnaQ-UvrC is able to bind the replisome, it might be the most plausible possibility. However, as such a phenomenon was never reported in other organism, a revision of the existing model is likely necessary when studying Actinobacteria organism. The findings highlighted in this work emphasize the need for further work regarding DNA repair mechanism in mycobacteria, including both TCR and GGR and considering the existence of a new possible player, DnaQ-UvrC.

5-3-12- UvrB might be involved in mutagenesis:

Mutagenesis has been thoroughly studied in our research unit. The mycobacterial mutasome was shown to be composed of accessory protein ImuA' and ImuB along with the TLS polymerase DnaE2, which is the main player in mutagenesis. Indeed, in this work, deletion of *dnaE2* rendered *Msm* highly sensitive to DNA-damage and it also reduced mutagenesis which is in accordance with the literature. Surprisingly, deletion of *uvrB* generated similar results. The knocked-out strain was sensitive to the DNA-damaging drugs used, it reduced the mutation rate and prevented the adaptation to Rifampicin and Moxifloxacin. Recently, a work from Deng et al.[192] highlighted a collaboration between NER and DnaE2. Indeed, they showed that NER is responsible for the generation of single-stranded DNA gaps, which are imperative for the functionality of DnaE2 in the context of

DNA repair. As UvrB is the critical factor for DNA incision in NER, it might also explain why its deletion severely impairs mutation frequency in the UV-induced mutation frequency assays performed in this thesis. These findings should help direct further studies on NER/mutagenesis interactions

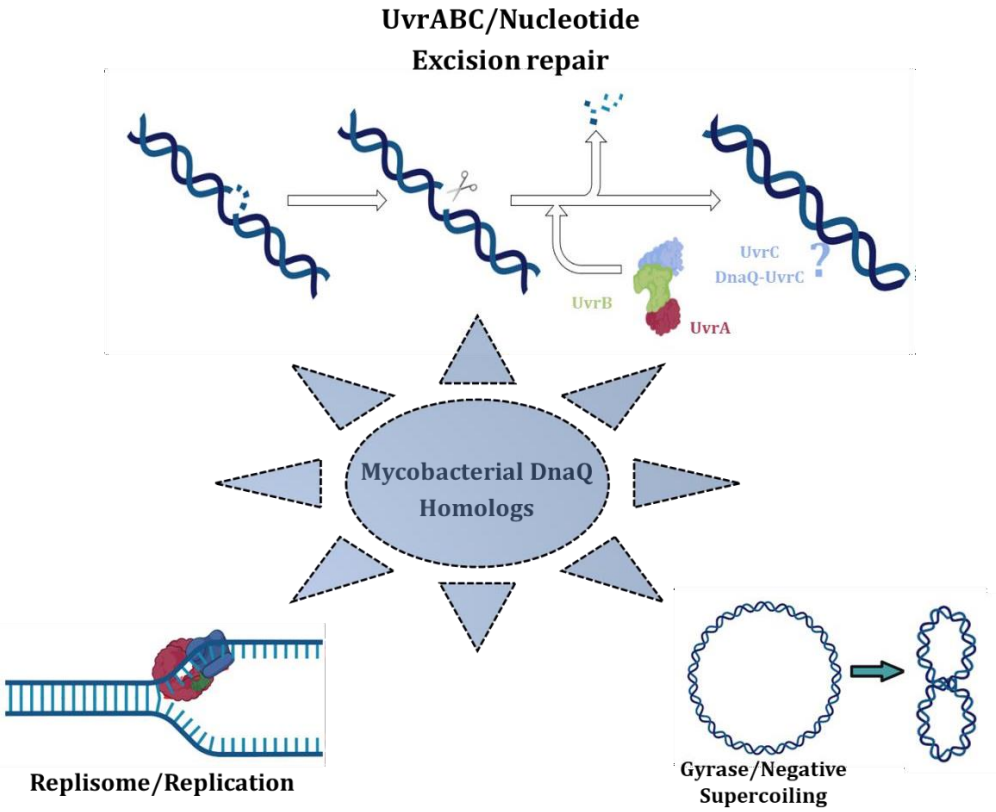
5-3-13- Recommendations for future research:

This study provides evidence that UvrC deletion do not jeopardize the survival chance of *Msm* strains in most of the conditions tested tested. However, such an absence of phenotype was highly unexpected and is not consistent with the current literature, even though almost all the work performed on UvrC was done in an organism in which no protein similar to DnaQ-UvrC exists. To ascertain the relative importance of UvrC vs DnaQ-UvrC to deal with various DNA damages further research must be carried out. This current work highlighted that different dynamics can be observed that depends on the growth phase. Therefore, further research should explore the process of DNA repair in the light of exponential growth phase vs stationary phase. Moreover, it is possible that more factors play roles in DNA repair that were not studied in this work.

This study reports an exclusively genetic approach to study NER. However, a biochemistry approach was initiated during this project by using eGFP-fused DnaQ-UvrC in a *dnaQ-uvrC* knock-out background. The eGFP is located at the C-terminal of the protein (Supplementary Figure III). Protein structure study performed in Chapter III showed that DnaQ-UvrC protein had about 200 unstructured amino acids after the last identified domain, the UVR binding domain. This, coupled to the flag-C tag should give the protein enough flexibility so the eGFP does not interfere with the protein normal activity. The work performed in this Chapter showed that this fluorescent tagged protein was able to restore the wild type phenotype.

CHAPTER VI:

Concluding remarks:



DNA metabolism is a broad subject that is incompletely understood in the human pathogen, *Mycobacterium tuberculosis*, despite its obvious importance in bacterial viability and adaptative evolution. DnaQ homologs are poorly studied and the knowledge about them relies solely on homology with the *E. coli* model organism. However, by using a dual, bio-informatic and molecular biology approach, this thesis highlights the need for a specific mycobacterial model, *Mycobacterium smegmatis* while bringing to light new information about the potential role(s) of the mycobacterial DnaQ homologs in DNA metabolism.

Through the use of bio-informatic tools, the work carried out in this thesis reports for the first time that mycobacterial DnaQ homologs possess protein domain organizations that differ from their *E. coli* counterparts. Even though they exhibit functional nucleases, similar to either DnaQ-like 3' – 5' proofreading exonuclease or UvrC GIYYIG endonuclease, their functional differences might be explained by the existence of unexpected interaction sites or conserved interaction sites with an unusual localization on the protein that are first reported in this work. Interestingly, similar protein structures were only found in mycobacteria which advocates for the specificity of DNA maintenance in mycobacteria mechanisms. Furthermore, both DnaQ and DnaQ-UvrC were identical in *Msm* and *Mtb* meaning that the data and knowledge generated in this project can be transposed to *Mtb*.

Mycobacterial differences with *E. coli* model organism were first investigated in the replication fidelity mechanism. As it was previously observed [31,66], mycobacterial DnaQ (MSMEG 4259, Rv_2191) do not contribute to maintain the replication fidelity and rather depends on the DnaE1 PHP domain. This assessment, however, should be considered carefully because of the inherent limitation of the mutation frequency assays that only look at mutations on the single *rpoB* locus and thereby neglect neutral or lethal mutations and mutations in noncoding areas of the genome, which likely occur more frequently, are ignored.[107]Interestingly, after UV irradiation, mycobacterial DnaQ homologs tend to have a synergistic activity that helped maintaining the genome integrity. However, the context in which this synergy occurs remain to be elucidated, even though it seems to be dependent on the replicative state of the bacteria.

On its own, mycobacterial DnaQ (*Msm* MSMEG 4259, *Mtb* Rv2191) seems to be involved in the gyrase A pathway. Even though the observed phenotype was modest, it was

statistically significant, reproducible and consistent with the previous identification of BRCT domain in the protein structure. Furthermore, it is the first phenotype ever reported for a Δ*dnaQ* strain. Shortly before this thesis publication a pre-print from Deng et al [187] showed that DnaQ in *Msm* has a role in sustaining of DNA replication fidelity upon chromosomal topological stress and highlighted a ofloxacin sensitization for a *Msm* Δ*dnaQ* strain. Interestingly ofloxacin and moxifloxacin used in both projects are both gyrase inhibitors which seems to confirm DnaQ involvement with gyrase pathways in *Msm*.

Previous work[41] showed that DnaQ-UvrC (MSMEG_6275, Rv_3711c) deletion resulted in a MMC hypersensitivity. Results obtained in this project confirmed these results while determining that this phenotype was dependent on interaction with both the replication machinery (through the β clamp-binding motif) and with the NER (through UVR binding motif) which confirmed previous bioinformatic analysis. Interestingly, deletion of *uvrC* had no effect on phenotype following MMC exposure. It even had a small, yet robust protective effect against MMC which was further confirmed by CRISPRi silencing of the gene.

Following UV exposition, neither deletion of *dnaQ-uvrC* nor *uvrC* had any apparent effect on survival. This result must be considered carefully and compared to existing literature. However, UvrC's ability to incise damaged DNA *in vitro* has been shown numerous times, [88,158,180] there was no proper molecular biology experiments performed *in vivo*. Although a previous work has be considered[179] in which *uvrC* was silenced using antisense primers. This work showed that silencing of *uvrC* resulted in an increased sensitivity to UV after 10 minutes of exposure. However, given the close sequence similarity between UvrC and DnaQ-UvrC, the anti-primers designed for this experiment do not allow discrimination between DnaQ-UvrC from UvrC and thus, it is difficult to determine which one of them - or both- is responsible for this UV sensitivity. Another possibility is that DnaQ-UvrC and UvrC have some functional redundancy and thus, it could be interesting to generate a double *dnaQ-uvrC uvrC* knockout to explore this idea.

At the exception of moxifloxacin, all the DNA-damaging drugs used in this project showed that DnaQ-UvrC was the main protein involved to damaged DNA incisions. Unlike, in *E. coli*, *uvrC* expression is constitutive in mycobacteria whereas *dnaQ-uvrC* is SOS response

dependent, like the other NER components. To determine the expression following DNA damage, a RNA-seq pipeline was set-up (Supplementary Figure IV). However, the data collected highlighted an issue with drug concentration used and no conclusions could be drawn from this experiment regarding gene expression. However, the pipeline was proven functional[50] and this experiment will be repeated. Yet, the 24 hours long UV-induced mutation frequency assay provided the beginning of an answer, namely that in low-replicative state, *uvrC* seem to have an increasing importance for DNA repair.

The investigation of antibiotic adaptation was one of the main topics investigated in this thesis. Previous work identified that DnaQ is linked to the emergence of drug resistance in *Mtb.*[34] The antibiotic adaptation performed in this study provide new evidence that DnaQ is indeed involved in the emergence of drug resistance in mycobacteria. Yet, no clear effect could be observed in UV-induced mutagenesis nor in spontaneous mutation frequency assays. Thus, a clear molecular role remains to be identified for mycobacterial DnaQ. Similarly, to the deletion of *dnaE2*-the TLS polymerase in mycobacteria which is mainly responsible for mutagenesis- deletion of *uvrB* resulted in an impaired antibiotic adaptation. Its deletion also resulted in a reduced UV-induced mutation rate which seem to link UvrB to the mutagenesis pathways.

Despite its crucial importance mycobacterial NER is a rather understudied subject. Since the discovery of the *dnaQ-uvrC* gene[176] it has never been carefully studied and the scientific community assumed that mycobacterial NER was identical to *E. coli*'s based on protein conservation. This reinforced by the difficulty inherent to mycobacterial organism studies, especially *Mtb*. However, the idea that mycobacteria have a unique metabolism that is linked to their human pathogens' nature is gaining acceptance and will surely better direct new research that will help defeat this foe that plagued humanity since its beginning.

There are already approaches to carry on with this project. For both DnaQ (MSMEG 4259) and DnaQ-UvrC (MSMEG_6275), eGFP fused proteins were able to supplement the native protein function. This is thought to be due to the presence of a FLAG-C tag linker between the eGFP and the rest of the protein. This means that they can be used in fluorescent microscopic experiments to track the live position of the protein in *Msm* bacilli. Preliminary experiments showed that these EGFP-fused proteins were well expressed

(Supplementary Figure V). Both eGFP and FLAG-C tag can be used for future protein pull down experiments that will allow the identification of any direct proteins directly interacting with mycobacterial DnaQ homologs.

One limitation of the work in Chapter III is that mutation frequency assays pick up mutations at only the *rpoB* locus. This limitation can be overcome by utilising whole genome sequencing (WGS) to study mutagenesis in more detail; *e.g*., the mutation frequency across the whole genome WGS in the *dnaQ* mutant strains generated in this project. An adaptation of WGS[192] could also be used, namely RADAR-seq (RAre DAmage and Repair sequencing).[191] This technique allows for the identification and localisation of a wide variety of DNA lesions throughout the genome. This would allow for activity comparisons to be made between DnaQ-UvrC and UvrC to study their potential repair function specificity at the level of the genome, something that can't not be detected using growth and resazurin-based assays.

References

1. WHO. *Global Tuberculosis Report, 2020. Geneva, Switzerland: WHO, 2020*. *ISBN 9789240037021.[Google Scholar]* (2021).
2. *Who Global Tuberculosis Report 2019*. (2020). doi:10.1787/f494a701-en.
3. WHO. *Global Tuberculosis Report 2022*.
4. Chiok, K. R., Dhar, N. & Banerjee, A. Mycobacterium tuberculosis and SARS-CoV-2 co-infections: The knowns and unknowns. *iScience* **26**, 106629 (2023).
5. Żukowska, L. *et al.* An overview of tuberculosis outbreaks reported in the years 2011–2020. *BMC Infect Dis* **23**, 1–11 (2023).
6. M Cristina, G. *et al.* Ancient Origin and Gene Mosaicism of the Progenitor of Mycobacterium tuberculosis. *PLoS Pathog* **1**, e5 (2005).
7. Hershberg, R. *et al.* High Functional Diversity in Mycobacterium tuberculosis Driven by Genetic Drift and Human Demography. *PLoS Biol* **6**, e311 (2008).
8. Supply, P. *et al.* Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosis. *Nature Genetics 2013 45:2* **45**, 172–179 (2013).
9. Gagneux, S. Ecology and evolution of Mycobacterium tuberculosis. *Nat Rev Microbiol* **16**, 202–213 (2018).
10. Orgeur, M. & Brosch, R. Evolution of virulence in the Mycobacterium tuberculosis complex. *Curr Opin Microbiol* **41**, 68–75 (2018).
11. Stritt, C. & Gagneux, S. How do monomorphic bacteria evolve? The Mycobacterium tuberculosis complex and the awkward population genetics of extreme clonality. *Peer Community Journal* Volume 3 (2023), article no. e92.
12. Boritsch, E. C. & Brosch, R. Evolution of Mycobacterium tuberculosis: New Insights into Pathogenicity and Drug Resistance. *Microbiol Spectr* (2016).
13. Delogu, G., Sali, M. & Fadda, G. The Biology of Mycobacterium Tuberculosis Infection. *Citation: Mediterr J Hematol Infect Dis* **2013**, 2013070 (2013).

14. Dinkele, R. *et al.* Aerosolization of Mycobacterium tuberculosis by Tidal Breathing. *Am J Respir Crit Care Med* **206**, 206–216 (2022).

15. Bussi, C. & Gutierrez, M. G. Mycobacterium tuberculosis infection of host cells in space and time. *FEMS Microbiol Rev* **43**, 341–361 (2019).

16. Sia, J. K. & Rengarajan, J. Immunology of Mycobacterium tuberculosis Infections . *Microbiol Spectr* **7**, 3–22 (2019).

17. Shaler, C. R., Horvath, C. N., Jeyanathan, M. & Xing, Z. Within the enemy's camp: Contribution of the granuloma to the dissemination, persistence and transmission of Mycobacterium tuberculosis. *Front Immunol* **4**, 1–8 (2013).

18. Warner, D. F. & Mizrahi, V. *Tuberculosis Chemotherapy: The Influence of Bacillary Stress and Damage Response Pathways on Drug Efficacy*. *Clinical Microbiology Reviews* vol. 19 558–570 (American Society for Microbiology (ASM), 2006).

19. Ndlovu, H. & Marakalala, M. J. Granulomas and inflammation: Host-directed therapies for tuberculosis. *Frontiers in Immunology* vol. 7 Preprint at https://doi.org/10.3389/fimmu.2016.00434 (2016).

20. Marakalala, M. J. *et al.* Inflammatory signaling in human tuberculosis granulomas is spatially organized. *Nat Med* **22**, 531–538 (2016).

21. Bishai, W. R. Rekindling old controversy on elusive lair of latent tuberculosis. *Lancet* **356**, 2113–2114 (2000).

22. Feldman, W. H. & Baggenstoss, A. H. The residual infectivity of the primary complex of tuberculosis. *Am J Pathol* **14**, 473 (1938).

23. Kasik, J. E. & Peacham, L. Properties of beta-lactamases produced by three species of mycobacteria. *Biochem J* **107**, 675–682 (1968).

24. Chambers, H. F. *et al.* Can penicillins and other beta-lactam antibiotics be used to treat tuberculosis? *Antimicrob Agents Chemother* **39**, 2620–2624 (1995).

25. Brennan, P. J. Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis. *Tuberculosis* **83**, 91–97 (2003).

26. Hett, E. C. & Rubin, E. J. Bacterial Growth and Cell Division: a Mycobacterial Perspective. *Microbiol Mol Biol Rev* **72**, 126 (2008).

27. Smith, T., Wolff, K. A. & Nguyen, L. Molecular biology of drug resistance in Mycobacterium tuberculosis. *Curr Top Microbiol Immunol* **374**, 53–80 (2013).

28. Kohanski, M. A., DePristo, M. A. & Collins, J. J. Sublethal antibiotic treatment leads to multidrug resistance via radical-induced mutagenesis. *Mol Cell* **37**, 311–320 (2010).

29. Warner, D. F., Tønjum, T. & Mizrahi, V. DNA Metabolism in Mycobacterial Pathogenesis. in *Current topics in microbiology and immunology* vol. 374 27–51 (2013).

30. Ditse, Z., Lamers, M. H. & Warner, D. F. DNA Replication in Mycobacterium tuberculosis. *Microbiol Spectr* **5**, (2017).

31. Rock, J. M. *et al.* DNA replication fidelity in Mycobacterium tuberculosis is mediated by an ancestral prokaryotic proofreader. *Nat Genet* **47**, 677–681 (2015).

32. Warner, D. F., Rock, J. M., Fortune, S. M. & Mizrahi, V. DNA replication fidelity in the mycobacterium tuberculosis complex. in *Advances in Experimental Medicine and Biology* vol. 1019 247–262 (2017).

33. Gu, S. *et al.* The β2 clamp in the Mycobacterium tuberculosis DNA polymerase III αβ2ε replicase promotes polymerization and reduces exonuclease activity. *Sci Rep* **6**, 18418 (2016).

34. Farhat, M. R. *et al.* Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. *Nat Genet* **45**, 1183–1189 (2013).

35. Green, A. G. *et al.* A convolutional neural network highlights mutations relevant to antimicrobial resistance in Mycobacterium tuberculosis. *Nat Commun* **13**, (2022).

36. Kesavan, A. K., Brooks, M., Tufariello, J. A., Chan, J. & Manabe, Y. C. Tuberculosis genes expressed during persistence and reactivation in the resistant rabbit model. *Tuberculosis (Edinb)* **89**, 17 (2009).

37. Olivencia, B. F. *et al.* Mycobacterium smegmatis PafBC is involved in regulation of DNA damage response. *Sci Rep* **7**, (2017).

38. Müller, A. U., Imkamp, F. & Weber-Ban, E. The Mycobacterial LexA/RecA-Independent DNA Damage Response Is Controlled by PafBC and the Pup-Proteasome System. *Cell Rep* (2018).

39. Rand, L. *et al.* The majority of inducible DNA repair genes in Mycobacterium tuberculosis are induced independently of RecA. *Mol Microbiol* **50**, 1031–1042 (2003).

40. Ford, C. B. *et al.* Use of whole genome sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection. *Nat Genet* **43**, 482–488 (2011).

41. Ditse, Z. Replication fidelity in the microevolution of mycobacteria, *PhD Thesis*, University of Cape Town(2015).

42. J M Reyrat, D. K. Mycobacterium smegmatis: an absurd model for tuberculosis? *Trends Microbiol* (2001).

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

44. Nagaraja, V., Godbole, A. A., Henderson, S. R. & Maxwell, A. DNA topoisomerase I and DNA gyrase as targets for TB therapy. *Drug Discov Today* **22**, 510–518 (2017).

45. Dawson, R. *et al.* Efficiency and safety of the combination of moxifloxacin, pretomanid (PA-824), and pyrazinamide during the first 8 weeks of antituberculosis treatment: a phase 2b, open-label, partly randomised trial in patients with drug-susceptible or drug-resistant pulmonary tuberculosis. *The Lancet* **385**, 1738–1747 (2015).

46. Kling, A. *et al.* Targeting DnaN for tuberculosis therapy using novel griselimycins. *Science (1979)* **348**, 1106–1112 (2015).

47. Reiche, M. A., Warner, D. F. & Mizrahi, V. Targeting DNA Replication and Repair for the Development of Novel Therapeutics against Tuberculosis. *Front Mol Biosci* **4**, 75 (2017).

48. Pidot, S. J. & Rizzacasa, M. A. The Nargenicin Family of Oxa-Bridged Macrolide Antibiotics. *Chemistry* **26**, 2780–2792 (2020).

49. Dhakal, D. *et al.* Characterization of Tailoring Steps of Nargenicin A1 Biosynthesis Reveals a Novel Analogue with Anticancer Activities. *ACS Chem Biol* **15**, 1370–1380 (2020).

50. Chengalroyen, M. D. *et al.* DNA-Dependent Binding of Nargenicin to DnaE1 Inhibits Replication in Mycobacterium tuberculosis. *ACS Infect Dis* **8**, 612–625 (2022).

51. Brooks, P. C., Movahedzadeh, F. & Davis, E. O. Identification of some DNA damage-inducible genes of mycobacterium tuberculosis: Apparent lack of correlation with lexA binding. *J Bacteriol* (2001).

52. Khisimuzi Mdluli *et al.* Mycobacterium tuberculosis DNA Gyrase as a Target for Drug Discovery. *Infect Disord Drug Targets* **7**, 159–168 (2008).

53. Painter, R. E. *et al.* Elucidation of DnaE as the antibacterial target of the natural product, nargenicin. *Chem Biol* **22**, 1362–1373 (2015).

54. Jarosz, D. F., Godoy, V. G., Delaney, J. C., Essigmann, J. M. & Walker, G. C. A single amino acid governs enhanced activity of DinB DNA polymerases on damaged templates. *Nature* **439**, 225–228 (2006).

55. Mrugała, B. *et al.* A study on the structure, mechanism, and biochemistry of kanamycin B dioxygenase (KanJ)—an enzyme with a broad range of substrates. *FEBS J* **288**, 1366–1386 (2021).

56. J. M. Pardo, F. Malpartida, M. Rico, A. Jimènes. Biochemical Basis of Resistance to Hygromycin B in Streptomyces hygroscopicus-the Producing Organism. *J Gen Microbiol* **131**, 1289–1290 (1985).

57. Castle, S.S. (2007) 'Ampicillin,' in Elsevier eBooks, pp. 1–6.

58. Parish, T. & Stoker, N. G. Use of flexible cassette method to generate a double unmarked Mycobacterium tuberculosis tlyA plcABC mutant by gene replacement. *Microbiology (N Y)* **146**, 1969–1975 (2000).

59. Wong, A. I. & Rock, J. M. CRISPR Interference (CRISPRi) for Targeted Gene Silencing in Mycobacteria. *Methods in Molecular Biology* **2314**, 343–364 (2021).

60. Kapopoulou, A., Lew, J. M. & Cole, S. T. The MycoBrowser portal: A comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis* **91**, 8–13 (2011).

61. Frickey, T. & Lupas, A. CLANS: A Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* **20**, 3702–3704 (2004).

62. Henikoff, S. & Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A* **89**, 10915–10919 (1992).

63. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **50**, D439–D444 (2022).

64. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature 2021 596:7873* **596**, 583–589 (2021).

65. Schrödinger, L. & DeLano, W., 2020. PyMOL, Available at: http://www.pymol.org/pymol.

66. Martin, Z. The roles of DNA damage responses and replication fidelity in mycobacterial evolution. *PhD Thesis*, University of Cape Town, (2022).

67. Rosche, W. A. & Foster, P. L. Determining mutation rates in bacterial populations. *Methods* **20**, 4–17 (2000).

68. Boshoff, H. I. M., Reed, M. B., Barry, C. E. & Mizrahi, V. DnaE2 polymerase contributes to in vivo survival and the emergence of drug resistance in Mycobacterium tuberculosis. *Cell* **113**, 183–193 (2003).

69. Ragheb, M. N. *et al.* Inhibiting the Evolution of Antibiotic Resistance. *Mol Cell* **73**, 157-165.e5 (2019).

70. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173 (2013).

71. de Wet, T. J., Winkler, K. R., Mhlanga, M., Mizrahi, V. & Warner, D. F. Arrayed crispri and quantitative imaging describe the morphotypic landscape of essential mycobacterial genes. *Elife* **9**, 1–36 (2020).

72. Sigrist, C. J. A. *et al.* ProRule: a new database containing functional and structural information on PROSITE profiles. *Bioinformatics* **21**, 4060–4066 (2005).

73. Sigrist, C. J. A. *et al.* PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief Bioinform* **3**, 265–274 (2002).

74. Sigrist, C. J. A. *et al.* New and continuing developments at PROSITE. *Nucleic Acids Res* **41**, (2013).

75. de Castro, E. *et al.* ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res* **34**, W362–W365 (2006).

76. Hulo, N. *et al.* The 20 years of PROSITE. *Nucleic Acids Res* **36**, 245–249 (2008).

77. Repšys, V., Margelevičius, M. & Venclovas, K. Re-searcher: A system for recurrent detection of homologous protein sequences. *BMC Bioinformatics* **9**, 1–6 (2008).

78. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389 (1997).

79. Margelevičius, M. & Venclovas, Č. PSI-BLAST-ISS: An intermediate sequence search tool for estimation of the position-specific alignment reliability. *BMC Bioinformatics* **6**, 1–10 (2005).

80. Koski, L. B. & Golding, G. B. The closest BLAST hit is often not the nearest neighbor. *J Mol Evol* **52**, 540–542 (2001).

81. Enright, A. J. & Ouzounis, C. A. BioLayout - An automatic graph layout algorithm for similarity visualization. *Bioinformatics* **17**, 853–854 (2001).

82. Gabler, F. *et al.* Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr Protoc Bioinformatics* **72**, e108 (2020).

83. Zimmermann, L. *et al.* A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol* **430**, 2237–2243 (2018).

84. Leroux, M., Soubry, N. & Reyes-Lamothe, R. Dynamics of Proteins and Macromolecular Machines in Escherichia coli. *EcoSal Plus* **9**, 6200 (2021).

85. Beattie, T. R. & Reyes-Lamothe, R. A replisome's journey through the bacterial chromosome. *Front Microbiol* **6**, 1–12 (2015).

86. Reyes-Lamothe, N. K. R. A quest for coordination among activities at the replisome. *Biochem Soc Trans* **47**, (2019).

87. Baños-Mateos, S. *et al.* High-fidelity DNA replication in Mycobacterium tuberculosis relies on a trinuclear zinc center. *Nat Commun* **8**, (2017).

88. Karakas, E. *et al.* Structure of the C-terminal half of UvrC reveals an RNase H endonuclease domain with an Argonaute-like catalytic triad. *EMBO J* **26**, 613 (2007).

89. Yang, W. Nucleases: Diversity of Structure, Function and Mechanism. *Q Rev Biophys* **44**, 1 (2011).

90. Nishino, T. & Morikawa, K. Structure and function of nucleases in DNA repair: shape, grip and blade of the DNA scissors. *Oncogene 2002 21:58* **21**, 9022–9032 (2002).

91. Gui, W. J. *et al.* Crystal structure of DNA polymerase III β sliding-clamp from Mycobacterium tuberculosis. *Biochem Biophys Res Commun* **405**, 272–277 (2011).

92. Toste Rêgo, A., Holding, A. N., Kent, H. & Lamers, M. H. Architecture of the Pol III-clamp-exonuclease complex reveals key roles of the exonuclease subunit in processive DNA synthesis and repair. *EMBO Journal* **32**, 1334–1343 (2013).

93. Wijffels G, Dalrymple BP, Prosselkov P, Kongsuwan K, Epa VC, Lilley PE, Jergic S, Buchardt J, Brown SE, Alewood PF, Jennings PA, Dixon NE. Inhibition of protein interactions with the beta 2 sliding clamp of Escherichia coli DNA polymerase III by peptides from beta 2-binding proteins. Biochemistry. 2004 May 18;43(19):5661-71.

94. Yu, X., Chini, C. C. S., He, M., Mer, G. & Chen, J. The BRCT domain is a phospho-protein binding domain. *Science* **302**, 639–642 (2003).

95. Timinskas, K., Balvočiute, M., Timinskas, A. & Venclovas, Č. Comprehensive analysis of DNA polymerase III α subunits and their homologs in bacterial genomes. *Nucleic Acids Res* **42**, 1393–1413 (2014).

96. Ekundayo, B. & Bleichert, F. Origins of DNA replication. *PLoS Genet* **15**, 1–21 (2019).

97. Gill, W. P. *et al.* A replication clock for Mycobacterium tuberculosis. *Nat Med* **15**, 211–214 (2009).

98. Warner, D. F. The role of DNA repair in M. tuberculosis pathogenesis. *Drug Discov Today* **7**, (2010).

99. Oakley, A. J. A structural view of bacterial DNA replication. *Protein Science* **28**, 990–1004 (2019).

100. Li, Y., Schroeder, J. W., Simmons, L. A. & Biteen, J. S. Visualizing bacterial DNA replication and repair with molecular resolution. *Curr Opin Microbiol* 38–45 (2019)

101. Masai, H. & Arai, K. DnaA- and PriA-dependent primosomes: two distinct replication complexes for replication of Escherichia coli chromosome. *Front Biosci* **1**, (1996).

102. Warner, D. F. *et al.* Essential roles for imuA′- and imuB-encoded accessory factors in DnaE2-dependent mutagenesis in Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **107**, 13093–13098 (2010).

103. S. Uphoff, Real-time dynamics of mutagenesis reveal the chronology of DNA repair and damage tolerance responses in single cells, Proc. Natl. Acad. Sci. U.S.A.

115 (28) E6516-E6525,

104. Ford, C. B. *et al.* Mycobacterium tuberculosis mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat Genet* **45**, 784–790 (2013).

105. Wanner, R. M., Güthlein, C., Springer, B., Böttger, E. C. & Ackermann, M. Stabilization of the genome of the mismatch repair deficient Mycobacterium tuberculosis by context-dependent codon choice. *BMC Genomics* **9**, 1–11 (2008).

106. Castañeda-Garciá, A. *et al.* A non-canonical mismatch repair pathway in prokaryotes. *Nat Commun* **8**, 1–10 (2017).

107. Nishant, K. T., Singh, N. D. & Alani, E. Genomic mutation rates: What high-throughput methods can tell us. *BioEssays* **31**, 912–920 (2009).

108. Machowski, E. E., Barichievy, S., Springer, B., Durbach, S. I. & Mizrahi, V. In vitro analysis of rates and spectra of mutations in a polymorphic region of the Rv0746 PE_PGRS gene of Mycobacterium tuberculosis. *J Bacteriol* **189**, 2190–2195 (2007).

109. Güthlein, C. *et al.* Characterization of the mycobacterial NER system reveals novel functions of the uvrDl helicase. *J Bacteriol* **191**, 555–562 (2009).

110. Warner DF, Evans JC, Mizrahi V. Nucleotide Metabolism and DNA Replication. Microbiol Spectr. 2014 Oct;2(5).

111. Fijalkowska, I. J., Schaaper, R. M. & Jonczyk, P. DNA replication fidelity in Escherichia coli: A multi-DNA polymerase affair. *FEMS Microbiol Rev* **36**, 1105–1121 (2012).

112. Maki, H. & Kornberg, A. Proofreading by DNA polymerase III of Escherichia coli depends on cooperative interaction of the polymerase and exonuclease subunits. *Proceedings of the National Academy of Sciences* **84**, 4389–4392 (1987).

113. McHenry, C. S. Breaking the rules: Bacteria that use several DNA polymerase IIIs. *EMBO Reports* vol. 12 408–414.

114. Ozawa, K. *et al.* Proofreading exonuclease on a tether: The complex between the E. coli DNA polymerase III subunits α, ε, θ and β reveals a highly flexible arrangement of the proofreading domain. *Nucleic Acids Res* **41**, 5354–5367 (2013).

115. Jergic, S. *et al.* A direct proofreader-clamp interaction stabilizes the Pol III replicase in the polymerization mode. *EMBO Journal* **32**, 1322–1333 (2013).

116. Barros T, Guenther J, Kelch B, Anaya J, Prabhakar A, O'Donnell M, Kuriyan J, Lamers MH. A structural role for the PHP domain in E. coli DNA polymerase III. BMC Struct Biol. 2013 May 14; 13:8.

117. Scheuermann, R., Tam, S., Burgers, P. M., Lu, C. & Echols, H. Identification of the epsilon-subunit of Escherichia coli DNA polymerase III holoenzyme as the dnaQ gene product: a fidelity subunit for DNA replication. *Proc Natl Acad Sci U S A* **80**, 7085–9 (1983).

118. Whatley, Z. & Kreuzer, K. N. Mutations that Separate the Functions of the Proofreading Subunit of the Escherichia coli Replicase. *G3: Genes|Genomes|Genetics* **5**, 1301–1311 (2015).

119. Woodgate, R., Bridges, B. A., Herrera, G. & Blanco, M. Mutagenic DNA repair in Escherichia coli XIII Proofreading exonuclease of DNA polymerase III holoenzyme is not operational during UV mutagenesis. *Mutation Research DNA Repair Reports* **183**, 31–37 (1987).

120. S.S. Dawes & V. Mizrahi. DNA metabolism in Mycobacterium leprae . *Lepr Rev* **72**, (2001).

121. Foster, P. L. & Sullivan, A. D. Interactions between epsilon, the proofreading subunit of DNA polymerase III, and proteins involved in the SOS response of Escherichia coli. *MGG Molecular & General Genetics* **214**, 467–473 (1988).

122. Dos Vultos, T., Mestre, O., Tonjum, T. & Gicquel, B. DNA repair in Mycobacterium tuberculosis revisited. in *FEMS Microbiology Reviews* vol. 33 471–487 (2009).

123. Sophia Gessner, Zela Alexandria-Mae Martin, Michael A Reiche, Joana A Santos, Ryan Dinkele, Atondaho Ramudzuli, Neeraj Dhar, Timothy J de Wet, Saber Anoosheh, Dirk M Lang, Jesse Aaron, Teng-Leong Chew, Jennifer Herrmann, Rolf Müller, John D McKinney; Roger Woodgate, Valerie Mizrahi, Česlovas Venclovas, Meindert H Lamers, Digby F

Warner (2023) Investigating the composition and recruitment of the mycobacterial ImuA′–ImuB–DnaE2 mutasome eLife 12:e75628.

124. Shibai, A. *et al.* Mutation accumulation under UV radiation in Escherichia coli. *Scientific Reports 2017 7:1* **7**, 1–12 (2017).

125. Frenoy, A. & Bonhoeffer, S. Death and population dynamics affect mutation rate estimates and evolvability under stress in bacteria. *PLoS Biol* **16**, 1–23 (2018).

126. Geisinger, E. *et al.* Antibiotic susceptibility signatures identify potential antimicrobial targets in the Acinetobacter baumannii cell envelope. Nat Commun. 2020 Sep 9;11(1):4522. Erratum in: Nat Commun. 2020 Nov 24;11(1):6107.

127. Faheem I, Gupta R, Nagaraja V. Distinct subunit architecture and assembly pattern of DNA gyrase from mycobacteria. Mol Microbiol. 2023 Jun;119(6):728-738

128. Lu, T. Effect of chloramphenicol, erythromycin, moxifloxacin, penicillin and tetracycline concentration on the recovery of resistant mutants of Mycobacterium smegmatis and Staphylococcus aureus. *Journal of Antimicrobial Chemotherapy* **52**, 61–64 (2003).

129. Ishino, S. *et al.* Identification of a mismatch-specific endonuclease in hyperthermophilic Archaea. *Nucleic Acids Res* **44**, 2977–2986 (2016).

130. Lang, G. I. Measuring mutation rates using the luria-delbrück fluctuation assay. in *Methods in Molecular Biology* vol. 1672 21–31 (Humana Press Inc., 2018).

131. Unniraman, S., Chatterji, M. & Nagaraja, V. DNA gyrase genes in Mycobacterium tuberculosis: A single operon driven by multiple promoters. *J Bacteriol* **184**, 5449–5456 (2002).

132. Pandey, B. *et al.* Double Mutants in DNA Gyrase Lead to Ofloxacin Resistance in Mycobacterium tuberculosis. *J Cell Biochem* **118**, 2950–2957 (2017).

133. Manjunatha, U. H., Somesh, B. P., Nagaraja, V. & Visweswariah, S. S. A Mycobacterium smegmatis gyrase B specific monoclonal antibody reveals association of gyrase A and B subunits in the cell. *FEMS Microbiol Lett* **194**, 87–92 (2001).

134. Smollett, K. L. *et al.* Global analysis of the regulon of the transcriptional repressor LexA, a key component of SOS response in Mycobacterium tuberculosis. *Journal of Biological Chemistry* **287**, 22004–22014 (2012).

135. Davis, E. O., Dullaghan, E. M. & Rand, L. Definition of the mycobacterial SOS box and use to identify LexA-regulated genes in Mycobacterium tuberculosis. *J Bacteriol* (2002).

136. Shinagawa, H. SOS response as an adaptive response to DNA damage in prokaryotes. *EXS* **77**, 221–235 (1996).

137. Papavinasasundaram, K. G. *et al.* Slow induction of RecA by DNA damage in Mycobacterium tuberculosis. *Microbiology (N Y)* **147**, 3271–3279 (2001).

138. Joo, C. *et al.* Real-Time Observation of RecA Filament Dynamics with Single Monomer Resolution. *Cell* **126**, 515–527 (2006).

139. Little, J. W., Edminston, S. H., Pacelli, L. Z. & Mount, D. Cleavage of the Escherichia coli lexA protein by the recA protease. *Proceedings of the National Academy of Sciences* **77**, 3225–3229 (1980).

140. Singh, A. Guardian of the nycobacterial genome: A review on DNA repair systems in Mycobacterium tuberculosis. *Microbiology (Reading)* **163**, 1740–1758 (2017).

141. Gamulin, V., Cetkovic, H. & Ahel, I. Identification of a promoter motif regulating the major DNA damage response mechanism of Mycobacterium tuberculosis. *FEMS Microbiol Lett* **238**, 57–63 (2004).

142. Ahel I, Vujaklija D, Mikoc A, Gamulin V. Transcriptional analysis of the recA gene in Streptomyces rimosus: identification of the new type of promoter. FEMS Microbiol Lett. 2002 Mar 19;209(1):133-7

143. Erill, I., Campoy, S., Mazon, G. & Barbé, J. Dispersal and regulation of an adaptive mutagenesis cassette in the bacteria domain. *Nucleic Acids Res* **34**, 66–77 (2006).

144. Campoy, S. *et al.* Role of the High-Affinity Zinc Uptake znuABC System in Salmonella enterica Serovar Typhimurium Virulence. *Infect Immun* **70**, 4721 (2002).

145. Kenyon, C. J. & Walker, G. C. Expression of the E. coli uvrA gene is inducible. *Nature 1981 289:5800* **289**, 808–810 (1981).

146. Schendel, P. F., Fogliano, M. & Strausbaugh, L. D. Regulation of the Escherichia coli K-12 uvrB operon. *J Bacteriol* **150**, 676–685 (1982).

147. Gifford, C. M., Blaisdell, J. O. & Wallace, S. S. Multiprobe RNase protection assay analysis of mRNA levels for the Escherichia coli oxidative DNA glycosylase genes under conditions of oxidative stress. *J Bacteriol* **182**, 5416–5424 (2000).

148. Courcelle, J., Khodursky, A., Peter, B., Brown, P. O. & Hanawalt, P. C. Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient Escherichia coli. *Genetics* **158**, 41–64 (2001).

149. Wang Y, Huang Y, Xue C, He Y, He ZG. ClpR protein-like regulator specifically recognizes RecA protein-independent promoter motif and broadly regulates expression of DNA damage-inducible genes in mycobacteria. J Biol Chem. 2011 Sep 9;286(36):31159-67.

150. Setlow R. B. and Carrier W. L. the Disappearance of Thymine Dimers From DNA: an Error-Correcting. *Proceedings of the National Academy of Sciences of the United States of* **51**, 226–231 (1964).

151. Boyce, R. P. & Howard-Flanders, P. Release of ultraviolet light-induced thymine dimers from DNA in E. coli K-12. 1964. *DNA Repair (Amst)* **2**, 1280–1287 (2003).

152. Wirth, N. *et al.* Conservation and divergence in nucleotide excision repair lesion recognition. *Journal of Biological Chemistry* **291**, 18932–18946 (2016).

153. Van Houten, B. & Kad, N. Investigation of bacterial nucleotide excision repair using single-molecule techniques. *DNA Repair (Amst)* **20**, 41–48 (2014).

154. Stracy, M. *et al.* Single-molecule imaging of UvrA and UvrB recruitment to DNA lesions in living Escherichia coli. *Nat Commun* **7**, (2016).

155. Truglio, J. J., Croteau, D. L., van Houten, B. & Kisker, C. Prokaryotic nucleotide excision repair: The UvrABC system. *Chemical Reviews* vol. 106 233–252.

156. Kisker, C., Kuper, J. & Van Houten, B. Prokaryotic nucleotide excision repair. *Cold Spring Harb Perspect Biol* **5**, 1–18 (2013).

157. Pakotiprapha, D., Samuels, M., Shen, K., Hu, J. H. & Jeruzalmi, D. Structure and mechanism of the UvrA-UvrB DNA damage sensor. *Nat Struct Mol Biol* **19**, 291–298 (2012).

158. Sohi, M. *et al.* Crystal structure of Escherichia coli UvrB C-terminal domain, and a model for UvrB-UvrC interaction. *FEBS Lett* **465**, 161–164 (2000).

159. Theis, K. *et al.* The nucleotide excision repair protein UvrB, a helicase-like enzyme with a catch. *Mutation Research - DNA Repair* **460**, 277–300 (2000).

160. Grossman, L. & Thiagalingam, S. Nucleotide excision repair, a tracking mechanism in search of damage. *Journal of Biological Chemistry* **268**, 16871–16874 (1993).

161. B, V. H. Nucleotide excision repair in Escherichia coli. *Microbiol Rev* **54**, 18–51 (1990).

162. Thakur, M., Kumar, M. B. J. & Muniyappa, K. Mycobacterium tuberculosis UvrB Is a Robust DNA-Stimulated ATPase That Also Possesses Structure-Specific ATP-Dependent DNA Helicase Activity. *Biochemistry* **55**, 5865–5883 (2016).

163. Thilly, W. G. Have environmental mutagens caused oncomutations in people? *Nat Genet* **34**, 255–259 (2003).

164. Spivak, G. & Ganesan, A. K. The complex choreography of transcription-coupled repair. *DNA Repair (Amst)* **19**, 64–70 (2014).

165. Hanawalt, P. C. & Spivak, G. Transcription-coupled DNA repair: Two decades of progress and surprises. *Nat Rev Mol Cell Biol* **9**, 958–970 (2008).

166. Cole, S. T. *et al.* Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393**, 537–544 (1998).

167. Dos Vultos, T. *et al.* Evolution and diversity of clonal bacteria: The paradigm of Mycobacterium tuberculosis. PLOS ONE 3(2): e1538 (2008).

168. Selby, C. P., Witkin, E. M. & Sancar, A. Escherichia coli mfd mutant deficient in 'mutation frequency decline' lacks strand-specific repair: In vitro complementation with purified coupling factor. *Proc Natl Acad Sci U S A* **88**, 11574–11578 (1991).

169. Kamarthapu, V. & Nudler, E. Rethinking transcription coupled DNA repair. *Curr Opin Microbiol* **24**, 15–20 (2015).

170. Cohen, S. E. & Walker, G. C. New discoveries linking transcription to DNA repair and damage tolerance pathways. *Transcription* **2**, 37–40 (2011).

171. Cohen, S. E. *et al.* Roles for the transcription elongation factor NusA in both DNA repair and damage tolerance pathways in Escherichia coli. *Proc Natl Acad Sci U S A* **107**, 15517–15522 (2010).

172. Putta, S. *et al.* Structural insights into the molecular mechanisms of the Mycobacterium evolvability factor Mfd. BioRxiv preprint, August 7, 2019.

173. Epshtein, V. *et al.* UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature* **505**, 372 (2014).

174. Goodman, M. F. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu Rev Biochem* **71**, 17–50 (2002).

175. Ohmori, H. *et al.* The Y-family of DNA polymerases. *Mol Cell* **8**, 7–8 (2001).

176. Mizrahi, V. & Andersen, S. J. DNA repair in Mycobacterium tuberculosis. What have we learnt from the genome sequence? *Mol Microbiol* **29**, 1331–1339 (1998).

177. Jiang, Q., Karata, K., Woodgate, R., Cox, M. M. & Goodman, M. F. The active form of DNA polymerase V is UmuD′2C–RecA–ATP. *Nature 2009 460:7253* **460**, 359–363 (2009).

178. Kana, B. D. *et al.* Role of the DinB homologs Rv1537 and Rv3056 in Mycobacterium tuberculosis. *J Bacteriol* **192**, 2220–2227 (2010).

179. Prammananan, T., Phunpruch, S., Jaitrong, S. & Palittapongarnpim, P. Mycobacterium tuberculosis uvrC essentiality in response to UV-induced cell damage. *Southeast Asian J Trop Med Public Health* **43**, 370–375 (2012).

180. Verhoeven, E. E. A., Van Kesteren, M., Moolenaar, G. F., Visse, R. & Goosen, N. Catalytic sites for 3' and 5' incision of Escherichia coli nucleotide excision repair are both located in UvrC. *Journal of Biological Chemistry* **275**, 5120–5123 (2000).

181. Lee, Y. J., Park, S. J., Ciccone, S. L. M., Kim, C. R. & Lee, S. H. An in vivo analysis of MMC-induced DNA damage and its repair. *Carcinogenesis* **27**, 446–453 (2006).

182. Sinha, R. P. & Häder, D. P. UV-induced DNA damage and repair: A review. *Photochemical and Photobiological Sciences* vol. 1 225–236, (2002).

183. Zatopek KM, Potapov V, Maduzia LL, Alpaslan E, Chen L, Evans TC Jr, Ong JL, Ettwiller LM, Gardner AF. RADAR-seq: A RAre DAmage and Repair sequencing method for detecting DNA damage on a genome-wide scale. DNA Repair (Amst). 2019 Aug;80:36-44.

184. Sabin S, Herbig A, Vågene ÅJ, Ahlström T, Bozovic G, Arcini C, Kühnert D, Bos KI. A seventeenth-century Mycobacterium tuberculosis genome supports a Neolithic

emergence of the Mycobacterium tuberculosis complex. Genome Biol. 2020 Aug 10;21(1):201.

185. Bellerose MM, Proulx MK, Smith CM; Baker RE, Ioerger TR, Sassetti CM. 2020. Distinct bacterial pathways influence the efficacy of antibiotics against Mycobacterium tuberculosis. mSystems 5:e00396-20.

186. Xu W, DeJesus MA, Rücker N, Engelhart CA, Wright MG, Healy C, Lin K, Wang R, Park SW, Ioerger TR, Schnappinger D, Ehrt S. 2017. Chemical genetic interaction profiling reveals determinants of intrinsic antibiotic resistance in Mycobacterium tuberculosis. Antimicrob Agents Chemother 61:e01334-17.

187. Ming-Zhi Deng, Qingyun Liu, Shu-Jun Cui, Han Fu, Mingyu Gan, Yuan-Yuan Xu, Xia Cai, Wei Sha, Guo-Ping Zhao, Sarah M. Fortune, Liang-Dong Lyu. Mycobacterial DnaQ is an Alternative Proofreader Ensuring DNA Replication Fidelity. bioRxiv 2023.10.24.563508.

188. Dániel MolnárÉva Viola SurányiTamás TrombitásDóra FüzesiRita HirmondóJudit Toth (2024) Genetic stability of Mycobacterium smegmatis under the stress of first-line antitubercular agents eLife 13:RP96695.

189. Poulton NC, Rock JM. Unraveling the mechanisms of intrinsic drug resistance in *Mycobacterium tuberculosis*. Front Cell Infect Microbiol. 2022 Oct 17;12:997283.

190. Sparks IL, Derbyshire KM,Jacobs WR, Morita YS, 2023. Mycobacterium smegmatis: The Vanguard of Mycobacterial Research. J Bacteriol 205:e00337-22.

191. Zatopek KM, Potapov V, Maduzia LL, Alpaslan E, Chen L, Evans TC Jr, Ong JL, Ettwiller LM, Gardner AF. RADAR-seq: A RAre DAmage and Repair sequencing method for detecting DNA damage on a genome-wide scale. DNA Repair (Amst). 2019 Aug;80:36-44.

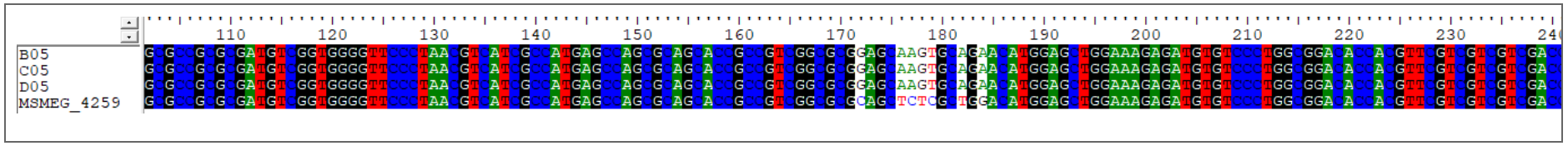
192. Deng MZ, Liu Q, Cui SJ, Wang YX, Zhu G, Fu H, Gan M, Xu YY, Cai X, Wang S, Sha W, Zhao GP, Fortune SM, Lyu LD. An additional proofreader contributes to DNA replication fidelity in mycobacteria. Proc Natl Acad Sci U S A. 2024 Aug 20;121(34):e2322938121.

193. Marina Della et al. Mycobacterial Ku and Ligase Proteins Constitute a Two-Component NHEJ Repair Machine.Science306,683-685(2004).

Supplementary Information:

List of Supplementary Figures:

List of Supplementary Tables:



Supplementary Figure I- - Screenshot from MSMEG_4259 and 3 dnaQ-uvrC Δβ-clamp clones sequence alignment. Oligo SA101 was used as a sequencing primer. Targeted mutagenesis was used twice on the D05 clone. First to remove the uvrC-mid domain (Figure XXIII-G) and secondly to remove the β clamp-binding site. The Full sequencing data can be found at: https://drive.google.com/drive/folders/1i3N3dVYLUoTm7lxyn6_yJ2lYqGOY51x7?usp=drive_link .

Strains		Δβ-clamp sequence				
MSMEG_4259		CAG	CTC	TCG	CTG	GAC
dnaQ-uvrC $^{\Delta\beta\text{-clamp}}$	B05	GAG	CAA	GTG	CAG	AAC
	C05	GAG	CAA	GTG	CAG	AAC
	D05	GAG	CAA	GTG	CAG	AAC

Supplementary Table I- Sequencing of three *dnaQ-uvrC* $^{\Delta\beta\text{-clamp}}$ clones. Oligo SA101 was used as a sequencing primer. Targeted mutagenesis was used twice on the D05 clone. First to remove the uvrC-mid domain (Figure XXIII-G) and secondly to remove the β clamp-binding site. The Full sequencing data can be found at: https://drive.google.com/drive/folders/1i3N3dVYLUoTm7lxyn6_yJ2lYqGOY51x7?usp=drive_link

Oligo ID	Sequence (5'-3')	Details	Restriction site
SA101	ATGGTACCGCCTCAAGTTTTCGTCCTC	Forward 1st segment MSMEG_4259	MluI
1834	TTCCTGTGAAGAGCCATTGATAATG	From Rock et al[62]	Na

Supplementary Table II- Primers used for sequencing

PAM sequence	PAM score	sgRNA_ sequence	sgRNA length	sgRN start coordinate	sgRNA_end_coordinate	Gene	Top_oligo	Bottom_oligo
CGGGCAC	24	GGGCCGGTACGTCGCTGGAT	20	3150235	3150254	MSMEG_3078	GGGAGGGCCGGTACGTCGCTGGAT	AAACATCCAGCGACGTACCGGCCC
ACGGAAT	12	ATGACCCGGCCGTGGGGGTC	20	3150300	3150319	MSMEG_3078	GGGAATGACCCGGCCGTGGGGGTC	AAACGACCCCCACGGCCGGGTCAT
ATAGGAC	15	AGCGTGACGGCCAGCACCGG	20	3150537	3150556	MSMEG_3078	GGGAAGCGTGACGGCCAGCACCGG	AAACCCGGTGCTGGCCGTCACGCT
CCAGCAC	19	ACTCCTCGTTGAGCGTGACGG	21	3150547	3150567	MSMEG_3078	GGGAACTCCTCGTTGAGCGTGACGG	AAACCCGTCACGCTCAACGAGGAGT
TCAGCAG	13	GCGCGCAGGGAAGACGCGGG	20	3150670	3150689	MSMEG_3078	GGGAGCGCGCAGGGAAGACGCGGG	AAACCCCGCGTCTTCCCTGCGCGC
AGGGAAG	4	ACTCCCGCCGAGCACGTGCGCGC	23	3150684	3150706	MSMEG_3078	GGGAACTCCCGCCGAGCACGTGCGCGC	AAACGCGCGCACGTGCTCGGCGGGAGT
CAGGGAA	18	ACTCCCGCCGAGCACGTGCGCG	22	3150685	3150706	MSMEG_3078	GGGAACTCCCGCCGAGCACGTGCGCG	AAACCGCGCACGTGCTCGGCGGGAGT
CGAGCAC	19	GTGCCGCTTGAACACTCCCGC	21	3150699	3150719	MSMEG_3078	GGGAGTGCCGCTTGAACACTCCCGC	AAACGCGGGAGTGTTCAAGCGGCAC

CCAGCAG	13	GCCGAGCACTTGTCGATGTAGC	22	3150748	3150769	MSMEG_3078	GGGAGCCGAGCACTTGTCGATGTAGC	AAACGCTACATCGACAAGTGCTCGGC
CGAGCAC	19	ACCCGCCCCACACACGGCGC	20	3150768	3150787	MSMEG_3078	GGGAACCCGCCCCACACACGGCGC	AAACGCGCCGTGTGTGGGGCGGGT
CGAGCAC	19	GCCCAGGAAGTCGCAGAAAT	20	3150820	3150839	MSMEG_3078	GGGAGCCCAGGAAGTCGCAGAAAT	AAACATTTCTGCGACTTCCTGGGC
GCAGAAA	**13**	**GTCTTGCCGCCCAGGAAGTC**	**20**	**3150828**	**3150847**	**MSMEG_3078**	**GGGAGTCTTGCCGCCCAGGAAGTC**	**AAACGACTTCCTGGGCGGCAAGAC**
CAGGAAG	4	AGCCGGTCGGTCTTGCCGCC	20	3150837	3150856	MSMEG_3078	GGGAAGCCGGTCGGTCTTGCCGCC	AAACGGCGGCAAGACCGACCGGCT
CCAGGAA	14	GCCAGCCGGTCGGTCTTGCCGC	22	3150838	3150859	MSMEG_3078	GGGAGCCAGCCGGTCGGTCTTGCCGC	AAACGCGGCAAGACCGACCGGCTGGC
GTGGAAG	4	GCGCACACGGCCGCCACGCAC	21	3151053	3151073	MSMEG_3078	GGGAGCGCACACGGCCGCCACGCAC	AAACGTGCGTGGCGGCCGTGTGCGC
CGAGGAC	15	GTCACCCGTCGCCGCAGGCTC	21	3151107	3151127	MSMEG_3078	GGGAGTCACCCGTCGCCGCAGGCTC	AAACGAGCCTGCGGCGACGGGTGAC
GAGGAAC	11	GCCTGCTCGCCGTAGAACTGGGT	23	3151176	3151198	MSMEG_3078	GGGAGCCTGCTCGCCGTAGAACTGGGT	AAACACCCAGTTCTACGGCGAGCAGGC

TGAGGAA	14	GCCTGCTCGCCGTAGAACTGGG	22	3151177	3151198	MSMEG_3078	GGGAGCCTGCTCGCCGTAGAACTGGG	AAACCCCAGTTCTACGGCGAGCAGGC
GTAGAAC	5	GCCGAGTTCGGCCTGCTCGCC	21	3151188	3151208	MSMEG_3078	GGGAGCCGAGTTCGGCCTGCTCGCC	AAACGGCGAGCAGGCCGAACTCGGC
CTGGCAC	24	GCCTCGTCGCCGCCGGTGTCG	21	3151217	3151237	MSMEG_3078	GGGAGCCTCGTCGCCGCCGGTGTCG	AAACCGACACCGGCGGCGACGAGGC
ACGGGAT	17	GCACCAGCACCTGCCGCGGT	20	3151250	3151269	MSMEG_3078	GGGAGCACCAGCACCTGCCGCGGT	AAACACCGCGGCAGGTGCTGGTGC
CCAGCAC	19	GTTGGGCGGCAGCACCGGCA	20	3151267	3151286	MSMEG_3078	GGGAGTTGGGCGGCAGCACCGGCA	AAACTGCCGGTGCTGCCGCCCAAC
CCGGCAC	24	GTCGGCGTTGGGCGGCAGCA	20	3151273	3151292	MSMEG_3078	GGGAGTCGGCGTTGGGCGGCAGCA	AAACTGCTGCCGCCCAACGCCGAC
GCAGCAC	19	GCCAGCTCGTCGGCGTTGGGCG	22	3151279	3151300	MSMEG_3078	GGGAGCCAGCTCGTCGGCGTTGGGCG	AAACCGCCCAACGCCGACGAGCTGGC
GCGGCAG	22	GGCCAGCTCGTCGGCGTTGG	20	3151282	3151301	MSMEG_3078	GGGAGGCCAGCTCGTCGGCGTTGG	AAACCCAACGCCGACGAGCTGGCC
CAGGCAC	24	GCGCGCGTTTGTCACCGCGCA	21	3151351	3151371	MSMEG_3078	GGGAGCGCGCGTTTGTCACCGCGCA	AAACTGCGCGGTGACAAACGCGCGC

GGGGCAG	22	GTGCCGGTAGTC GGATTTGC	20	3151588	3151607	MSME G_3078	GGGAGTGCCGGTAGT CGGATTTGC	AAACGCAAATCCGAC TACCGGCAC
CAGGAAC	11	GCTGCGAGTCGG ACGTGTGCCG	22	3151689	3151710	MSME G_3078	GGGAGCTGCGAGTCG GACGTGTGCCG	AAACCGGCACACGTC CGACTCGCAGC
GCAGGAA	14	GCTGCGAGTCGG ACGTGTGCC	21	3151690	3151710	MSME G_3078	GGGAGCTGCGAGTCG GACGTGTGCC	AAACGGCACACGTCC GACTCGCAGC
CTGGGAT	17	GGCGCGGACGCT GCTCGGCG	20	3151721	3151740	MSME G_3078	GGGAGGCGCGGACGC TGCTCGGCG	AAACCGCCGAGCAGC GTCCGCGCC
CGAGCAC	19	GTCGTCGACACCG AGATCCT	20	3151816	3151835	MSME G_3078	GGGAGTCGTCGACAC CGAGATCCT	AAACAGGATCTCGGT GTCGACGAC
ACGGCAC	24	GATCACCGGGTCC GGCTCCG	20	3151885	3151904	MSME G_3078	GGGAGATCACCGGGT CCGGCTCCG	AAACCGGAGCCGGAC CCGGTGATC
GCAGGAT	9	ACAGACCCTCACT GTTGCGCG	21	3151909	3151929	MSME G_3078	GGGAACAGACCCTCA CTGTTGCGCG	AAACCGCGCAACAGT GAGGGTCTGT
GCGGCAG	22	GTACAGACCCTCA CTGTTGC	20	3151912	3151931	MSME G_3078	GGGAGTACAGACCCT CACTGTTGC	AAACGCAACAGTGAG GGTCTGTAC
GCAGCAG	13	GCGCCTCGTCACG CACGCGCT	21	3151939	3151959	MSME G_3078	GGGAGCGCCTCGTCA CGCACGCGCT	AAACAGCGCGTGCGT GACGAGGCGC

CGGGCAC	24	GCGCGGTGGCGACGCCGATGC	21	3152131	3152151	MSMEG_3078	GGGAGCGCGGTGGCGACGCCGATGC	AAACGCATCGGCGTCGCCACCGCGC
CCGGGAG	16	GTGTTTGCGGGTCGCCGGTC	20	3152216	3152235	MSMEG_3078	GGGAGTGTTTGCGGGTCGCCGGTC	AAACGACCGGCGACCCGCAAACAC
GCGGCAG	22	GCTCGAACTGCAGGTCCCGC	20	3885437	3885456	MSMEG_3816	GGGAGCTCGAACTGCAGGTCCCGC	AAACGCGGGACCTGCAGTTCGAGC
GCGGCAT	21	GAGGTCGGCCAACTCGGCGC	20	3885492	3885511	MSMEG_3816	GGGAGAGGTCGGCCAACTCGGCGC	AAACGCGCCGAGTTGGCCGACCTC
CGAGGAT	9	GGCCTCGCGGTAGACCTGGT	20	3885642	3885661	MSMEG_3816	GGGAGGCCTCGCGGTAGACCTGGT	AAACACCAGGTCTACCGCGAGGCC
GCAGGAA	14	GATCAGGCTGCGCGTGGACC	20	3885846	3885865	MSMEG_3816	GGGAGATCAGGCTGCGCGTGGACC	AAACGGTCCACGCGCAGCCTGATC
CAGGAAG	4	ATCAGGCTGCGCGTGGACCG	20	3885847	3885866	MSMEG_3816	GGGAATCAGGCTGCGCGTGGACCG	AAACCGGTCCACGCGCAGCCTGAT
CGAGGAT	9	GAAGCCCTCCTTGTCGGCGT	20	3885870	3885889	MSMEG_3816	GGGAGAAGCCCTCCTTGTCGGCGT	AAACACGCCGACAAGGAGGGCTTC
CGGGCAG	22	GATCGACACCAGCGACACCT	20	3885894	3885913	MSMEG_3816	GGGAGATCGACACCAGCGACACCT	AAACAGGTGTCGCTGGTGTCGATC

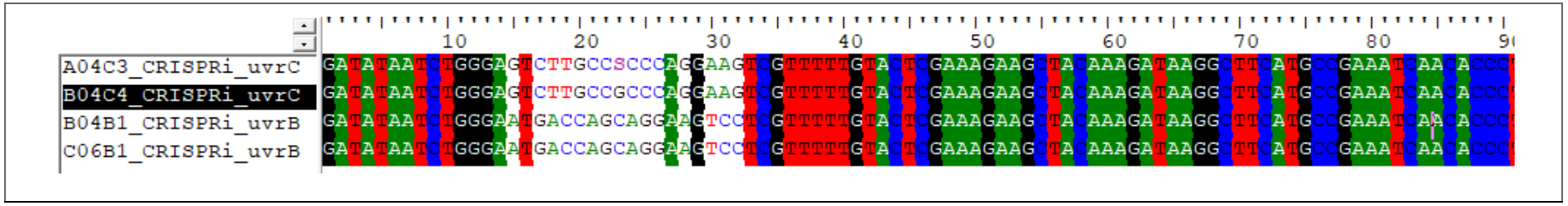
GCAGCAG	13	GGGCAGGTCGAG GCCCTCGC	20	3885915	3885934	MSME G_3816	GGGAGGGCAGGTCGA GGCCCTCGC	AAACGCGAGGGCCTC GACCTGCCC
CCAGCAC	19	GCGCAGCAGGTT GATGCCCA	20	3885933	3885952	MSME G_3816	GGGAGCGCAGCAGGT TGATGCCCA	AAACTGGGCATCAAC CTGCTGCGC
GCAGCAG	13	ACTCACCGAGCCG CAGCTGCC	21	3885965	3885985	MSME G_3816	GGGAACTCACCGAGC CGCAGCTGCC	AAACGGCAGCTGCGG CTCGGTGAGT
CGAGCAG	13	ACCGGACCCGGAT GCCCATCT	21	3886031	3886051	MSME G_3816	GGGAACCGGACCCGG ATGCCCATCT	AAACAGATGGGCATC CGGGTCCGGT
CCAGCAC	19	ATCTTCTTGGTCA GCGTCGTGA	22	3886081	3886102	MSME G_3816	GGGAATCTTCTTGGT CAGCGTCGTGA	AAACTCACGACGCTG ACCAAGAAGAT
ACGGCAG	22	GCGGACGGTTGT CGACCGCGG	21	3886325	3886345	MSME G_3816	GGGAGCGGACGGTTG TCGACCGCGG	AAACCCGCGGTCGAC AACCGTCCGC
CCGGAAG	4	GTTGTCGACCGCG GACGGCAG	21	3886332	3886352	MSME G_3816	GGGAGTTGTCGACCG CGGACGGCAG	AAACCTGCCGTCCGC GGTCGACAAC
GCGGCAC	24	GCCCTCGTACATA CCGCCGATCT	23	3886395	3886417	MSME G_3816	GGGAGCCCTCGTACA TACCGCCGATCT	AAACAGATCGGCGGT ATGTACGAGGGC
CCAGCAG	13	GTCACGTGCGACT CGTCGATGA	22	3886426	3886447	MSME G_3816	GGGAGTCACGTGCGA CTCGTCGATGA	AAACTCATCGACGAG TCGCACGTGAC

GCAGGAA	14	GTGCGACTCGTCG ATGACCA	20	3886431	3886450	MSME G_3816	GGGAGTGCGACTCGT CGATGACCA	AAACTGGTCATCGAC GAGTCGCAC
CAGGAAG	4	GTGCGACTCGTCG ATGACCAG	21	3886431	3886451	MSME G_3816	GGGAGTGCGACTCGT CGATGACCAG	AAACCTGGTCATCGA CGAGTCGCAC
CGGGGAA	18	GATGACCAGCAG GAAGTCCT	20	3886443	3886462	MSME G_3816	GGGAGATGACCAGCA GGAAGTCCT	AAACAGGACTTCCTG CTGGTCATC
GGGGAA G	**14**	**ATGACCAGCAGG AAGTCCTC**	**20**	**3886444**	**3886463**	**MSME G_381 6**	**GGGAATGACCAGCA GGAAGTCCTC**	**AAACGAGGACTTCC TGCTGGTCAT**
CCAGCAG	13	GAAGTCCTCGGG GAAGTAGT	20	3886455	3886474	MSME G_3816	GGGAGAAGTCCTCGG GGAAGTAGT	AAACACTACTTCCCC GAGGACTTC
CGAGCAG	13	GCGCGAGTAGTT CTCGATGCC	21	3886521	3886541	MSME G_3816	GGGAGCGCGAGTAGT TCTCGATGCC	AAACGGCATCGAGAA CTACTCGCGC
GCAGAAC	5	GAGTAGTTCTCG ATGCCCGA	20	3886525	3886544	MSME G_3816	GGGAGAGTAGTTCTC GATGCCCGA	AAACTCGGGCATCGA GAACTACTC
CCAGCAG	13	GCATGCGCAACCG CTGCGCCT	21	3886589	3886609	MSME G_3816	GGGAGCATGCGCAAC CGCTGCGCCT	AAACAGGCGCAGCGG TTGCGCATGC
CCGGCAA	23	GCCGCCGCCATCC GCTCCGGA	21	3886678	3886698	MSME G_3816	GGGAGCCGCCGCCAT CCGCTCCGGA	AAACTCCGGAGCGGA TGGCGGCGGC

CCGGGAA	18	ACCGGCAACGTA GTGCGTGG	20	3886698	3886717	MSME G_3816	GGGAACCGGCAACGT AGTGCGTGG	AAACCCACGCACTAC GTTGCCGGT
CGGGAAG	4	ACCGGCAACGTA GTGCGTGGC	21	3886698	3886718	MSME G_3816	GGGAACCGGCAACGT AGTGCGTGGC	AAACGCCACGCACTA CGTTGCCGGT
GAAGAAC	5	AGCGCCTCGATCT CGTCGCC	20	3886783	3886802	MSME G_3816	GGGAAGCGCCTCGAT CTCGTCGCC	AAACGGCGACGAGAT CGAGGCGCT
ACGGGAT	17	GCACCGCGAGTTC CTCGTACG	21	3886817	3886837	MSME G_3816	GGGAGCACCGCGAGT TCCTCGTACG	AAACCGTACGAGGAA CTCGCGGTGC
ACGGAAC	11	ATCTCCACGGTGT CGCCGCGCAC	23	3886846	3886868	MSME G_3816	GGGAATCTCCACGGT GTCGCCGCGCAC	AAACGTGCGCGGCGA CACCGTGGAGAT
CCAGCAG	13	GCGGTTGTACTGC ACGTCCA	20	3886902	3886921	MSME G_3816	GGGAGCGGTTGTACT GCACGTCCA	AAACTGGACGTGCAG TACAACCGC
TGAGCAG	13	ACTGCACGTCCAC CAGCAGTC	21	3886910	3886930	MSME G_3816	GGGAACTGCACGTCC ACCAGCAGTC	AAACGACTGCTGGTG GACGTGCAGT
GCGGCAC	24	GCAGTCTGAGCA GCCCGTCGC	21	3886925	3886945	MSME G_3816	GGGAGCAGTCTGAGC AGCCCGTCGC	AAACGCGACGGGCTG CTCAGACTGC
ACAGCAG	13	ACCACCACGACAT CGCGGCGCG	22	3887035	3887056	MSME G_3816	GGGAACCACCACGAC ATCGCGGCGCG	AAACCGCGCCGCGAT GTCGTGGTGGT

GCGGAAG	4	GAAGTACTCGAC GGCGTTGT	20	3887184	3887203	MSME G_3816	GGGAGAAGTACTCGA CGGCGTTGT	AAACACAACGCCGTC GAGTACTTC
GAAGCAT	7	GTACTCGACGGCG TTGTGCG	20	3887187	3887206	MSME G_3816	GGGAGTACTCGACGG CGTTGTGCG	AAACCGCACAACGCC GTCGAGTAC
CGAGCAA	10	ATTTACCCGTACC GGTGGCGC	21	3887315	3887335	MSME G_3816	GGGAATTTACCCGTA CCGGTGGCGC	AAACGCGCCACCGGT ACGGGTAAAT
CGAGCAC	19	GGCCCGGTACTCC GAATGTG	20	3887469	3887488	MSME G_3816	GGGAGGCCCGGTACT CCGAATGTG	AAACCACATTCGGAG TACCGGGCC
ACAGGAT	9	GGTACTCCGAAT GTGCGAGC	20	3887474	3887493	MSME G_3816	GGGAGGTACTCCGAA TGTGCGAGC	AAACGCTCGCACATT CGGAGTACC
GTAGAAC	5	GCGAAGGCCATG TTCACCAGG	21	3887509	3887529	MSME G_3816	GGGAGCGAAGGCCAT GTTCACCAGG	AAACCCTGGTGAACA TGGCCTTCGC

Supplementary Table III- List of sgRNA sequence with PAM scores. sgRNA used in this study are highlighted in bold.



Supplementary Figure II- Screenshot clones carring PJR962 plasmid with either *uvrB* or *uvrC* sgRNA sequence alignment. Oligo 1834 was used as a sequencing primer. The full sequencing data can be found at: https://drive.google.com/drive/folders/1FrLs6KxICFqK6fQbznrPdEQhN-J_vf8l?usp=drive_link

Supplementary Figure III- Map of plasmids used in this study- pJR962 was used for CRISPRi silencing experiments, p2NIL was used for unmarked knock-out gene deletion and pMCpAINT was used to express eGFP-fused MSMEG_4259 and MSMEG_6275.

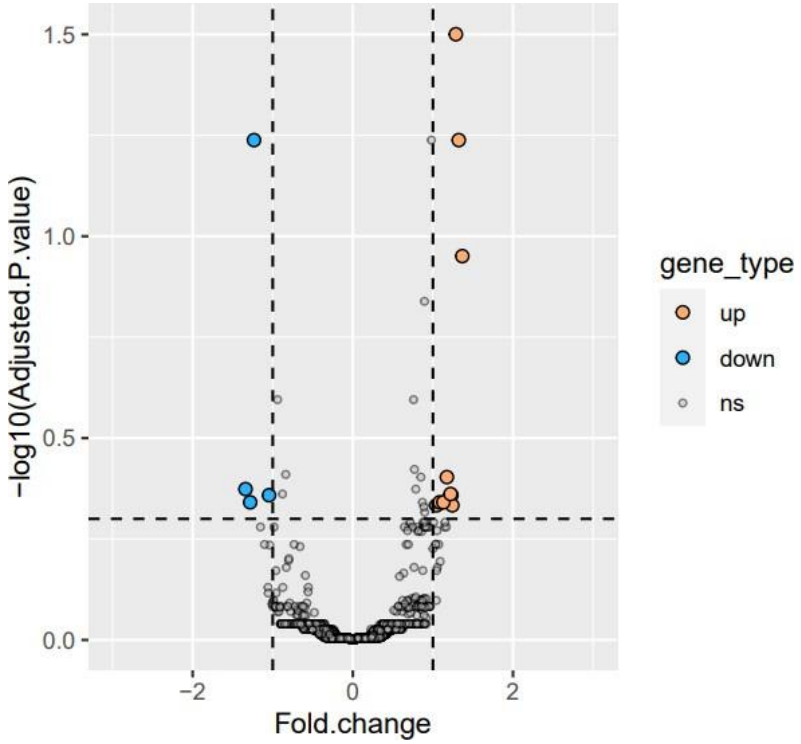
Supplementary Table IV- Geninfo identifier and distribution of the genomes used for the clustering. Column name is explained below and can be visualised on Figure XIX in section 3-2-7: Subgroup MDQ-A corresponds to mycobacterial DnaQ (MSMEG_6275), Exonuclease domain + BRCT, Exonuclease domain + BRCT + BRCT + TerD, Exonuclease domain only. MDQ-B corresponds to Exonuclease domain + TerB + BRCT. MDQ-C corresponds to shorter exonuclease domain + BRCT. DQUC-A corresponds to mycobacterial DnaQ-UvrC (MSMEG_4259). DQUC-B corresponds to exonuclease domain, GIYYIG nuclease, shortened UvrC-mid domain. DQUC-C corresponds to Fanconi associated nuclease + Exonuclease domain. The full table can be found at: https://drive.google.com/drive/folders/1BRgELZtXp3DLODgH58_C7yPfKn53kN3L?usp=drive_link

Strain colony counts	wild type			*ΔdnaQ-uvrC*			*ΔdnaQ +ΔdnaQ-uvrC*			*ΔuvrB*			*ΔuvrC*			*ΔdnaE2*		
Dilution	10^{-5}	10^{-6}	10^{-7}	10^{-5}	10^{-6}	10^{-7}	10^{-5}	10^{-6}	10^{-7}	10^{-5}	10^{-6}	10^{-7}	10^{-5}	10^{-6}	10^{-7}	10^{-5}	10^{-6}	10^{-7}
Pre UV-exposure	159	16	2	132	18	4	132	14	1	137	12	2	142	6	1	202	24	2

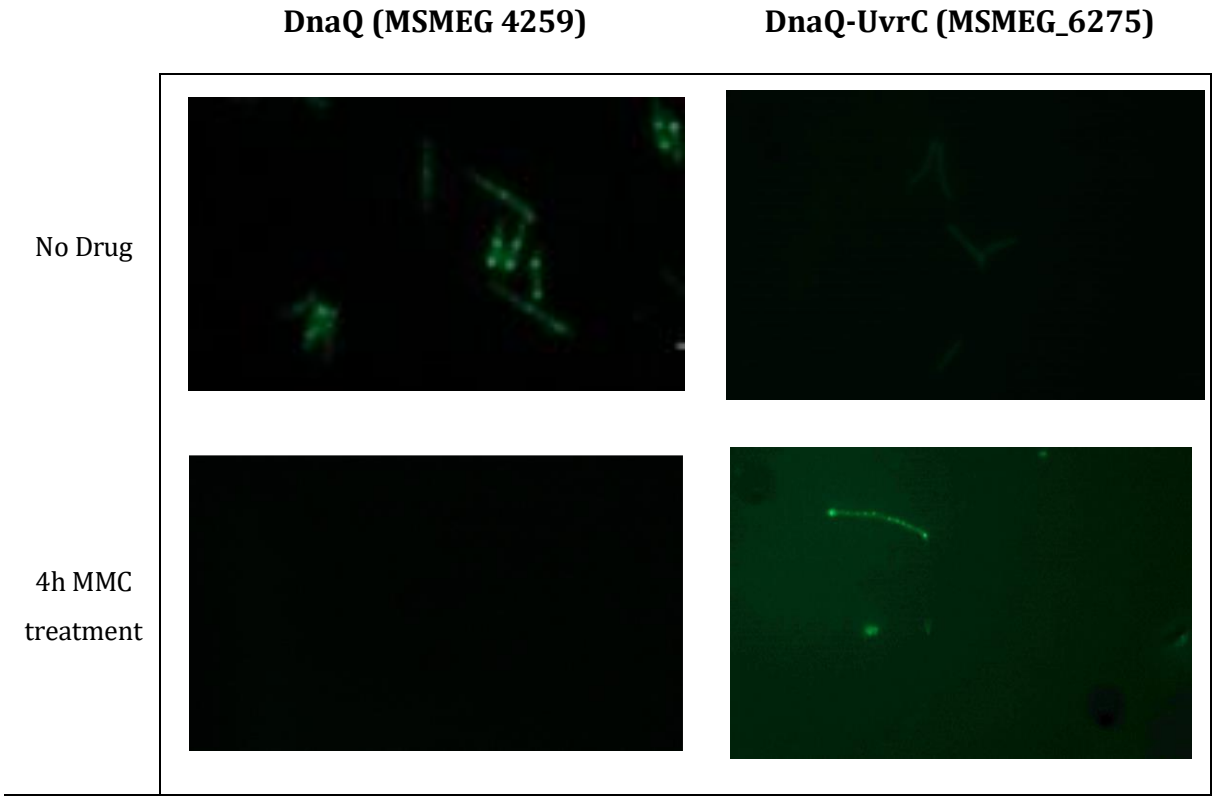
Supplementary Table V- Bacteria survival in for the adaptation of the Luria Delbruck fluctuation Assay - Bacteria were diluted as indicated and plated on 7H10 plates without any drugs.

trains	Time Point	wild type			*ΔdnaQ-uvrC*			*ΔuvrB*			*ΔuvrC*			*Double KO*			*ΔdnaE2*		
Dilution		Neat	10^{-1}	10-2	Neat	10^{-1}	10-2	Neat	10^{-1}	10-2	Neat	10^{-1}	10-2	Neat	10^{-1}	10-2	Neat	10^{-1}	10-2
Colony count	4h	56	4	0	58	6	0	34	3	0	56	8	0	44	5	0	27	2	0
	8h	75	8	0	88	8	1	58	5	0	91	8	0	67	8	0	39	3	0
	12 h	124	13	2	152	17	1	101	11	1	171	16	2	133	14	1	69	8	0
	16 h	TMTC	33	2	TMTC	37	4	TMTC	19	1	TMTC	38	4	TMTC	33	4	127	11	2
	20 h	TMTC	77	6	TMTC	81	7	TMTC	34	4	TMTC	86	10	TMTC	71	7	TMTC	25	2
	24 h	TMTC	TMTC	17	TMTC	TMTC	19	TMTC	TMTC	9	TMTC	TMTC	21	TMTC	TMTC	17	TMTC	61	7

Supplementary Table VI- Bacteria survival in for the UV-induced mutation frequency. Bacteria were diluted as indicated and plated on 7H10 plates without any drugs.



Supplementary Figure IV- Volcano plot RNA sequencing data -As mentioned previously, the drug concentration was not adapted and the transcriptional response was too low for any conclusions to be drawn.



Supplementary Figure V- Fluorescent microscopy proof of concept. As DnaQ-UvrC (MSMEG_6275) is SOS response inducible, a MMC treatment was performed.