UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL

Unveiling the physiological and molecular basis of *Mycobacterium tuberculosis* complex biofilms: the potential role of *Rv2488c*

Mestrado em Biologia Molecular e Genética

Tiago Filipe Serras Baeta

Dissertação orientada por:
Doutora Mónica Vieira Cunha

2015

"*Nothing in Biology makes sense except in the light of Evolution.*"

Theodosius Dobzhansky

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Nunca imaginei chegar a este momento, em que estou prestes a dar um passo final neste percurso que começou há anos, mesmo antes de entrar na Universidade. Peço desde já desculpa se omitir alguém, mas muitos se cruzaram comigo e não cabem todos aqui.

Antes de mais quero deixar um gigantesco agradecimento à minha orientadora, Doutora Mónica Vieira Cunha, pelo voto de confiança dado há mais de um ano, pelas palavras encorajadoras, pelos puxões de orelhas quando eram necessários, pelas conversas informais e por todo o apoio (material, financeiro, humano) e pelos conselhos quando o trabalho parecia ter problemas constantes.

Agradeço ao INIAV IP, particularmente ao Laboratório de Bacteriologia e Micologia da UEISPSA, como instituição de acolhimento do trabalho experimental que esteve na base da escrita da presente dissertação. Agradeço o apoio financeiro prestado pela Fundação para a Ciência e Tecnologia (FCT, IP), no âmbito do projeto com a referência PTDC/CVT/117794/2010 e com enquadramento no Projecto 3599 – *Promover a Produção Científica e Desenvolvimento Tecnológico e a Constituição de Redes Temáticas*.

Ao meu grupo de amigos mais próximo: Tânia, Gil, Inês, Feijoca, Hannah, Saphira, Luzinha, Lila, Borbinha, Teresinha, Joaquim, Catarina Dias, Jaime, aos meus amigos da "roupa de pinguim", Diogo Ribeiro, Lourenço (Slowpoke), Sofia(s), Rochinha, Nuno, Lemos, Viegas, Castelo, Fábio, Mameri, Duarte, Gui, Cátia, CR, Stitch, Maria, Gus & Companhia, à "Comissão de Velhos do Restelo", Macro, Carvalho, Barateiro e Crazy (e novos membros, 30 e Sílvia), à minha madrinha Dora Dias pelos conselhos, risos e conversas, aos meus afilhados Narciso, Serafim, Gisela e Soromenho ... e a tanta gente mais que não consigo colocar aqui os nomes todos. Ao pessoal do Secundário, e principalmente à Pires, Filipa & Diogo e Marco, obrigado pelos cafés para desanuviar, os jantares, as saídas e as conversas. Ao meu grupo de trabalho, Marta Vaz, Ana Reis e Ana Prata, por todos os lanches da manhã, da tarde, pelos jantares no INIAV à pressa e com alguma dificuldade porque o senhor da pizza hut não sabia onde era o INIAV, pelos almoços, pelos risos, pelas conversas, pelas idas à padaria portuguesa, quando tudo estava a correr mal mas não deixávamos que ninguém fosse ao chão. Melhores colegas não podia ter pedido. À D. Conceição e à D. Ana dos meios de cultura do INIAV de Benfica, por todo o apoio, material, ajudas e conselhos que me deram. Este trabalho tinha sido completamente diferente sem a vossa ajuda, e por isso estou muitíssimo grato! Aos meus amigos de mestrado, Catarina Brás e Tatiana Morais principalmente, mas a tantos outros, Bruno, Fábio, Catarina R., Teresa, Inês Direito, Cristiana, Vanessa, foi apenas um ano mas foi um ano bem passado. À minha família, aos meus pais, aos meus primos e primas, tios, padrinhos, que sempre estiveram lá e me apoiaram no caminho todo, e a ti Joel por tudo.

E por onde esta estrada que percorro me leve, carrego isto tudo e muito mais às costas.

"Façam o favor de ser felizes"

ABSTRACT

The *Mycobacterium tuberculosis* complex (MTC) includes several closely-related pathogenic species, namely *M. tuberculosis*, the main etiological agent of human tuberculosis (TB), and *M. bovis*, the causative agent of bovine tuberculosis (bTB), one of the most relevant zoonosis in the world. The complex pathogenesis of MTC bacteria is influenced by several factors, including host and bacterial genetic backgrounds and environmental factors. On infection, a proportion of infecting bacilli is able to survive and strive among the challenging conditions present in the host's macrophage. Current models of mycobacterial persistence invoke the presence of dormant populations of non-growing cells that reactivate during host immunosuppression. Bacterial biofilms, described as single or multispecies communities triggered by cell-density dependent Quorum-Sensing (QS) events, exhibit heterogeneous and stress-tolerant populations that thrive under inhospitable niches and display recalcitrant characters, such as increased xenobiotic tolerance and resistance to host's immune response. *M. tuberculosis* infections in humans have similarities to microbial biofilms, displaying antibiotic resistance, asymptomatic latency and a persistence phenotype associated to a dormant and non-replicating state of infecting microorganisms that enables evasion from the host's immune system. The clinical phenotype of TB thus raises the question of whether or not *M. tuberculosis* latent infections are associated to the biofilm phenotype *in vivo* and if these multicellular structures may allow persistence of infection and recalcitrance to xenobiotic through time.

Biofilm formation and maturation is modulated by cell density and the accumulation of signaling molecules (autoinducers) that bind a cognate response regulator and initiate a transduction cascade, triggering QS-dependent gene expression. Contrasting with the well-studied QS circuits of many pathogenic bacterial species, mycobacterial QS mechanisms remain poorly explored. The recent bioinformatic analysis of 53 actinobacterial genomes highlighted the presence of several putative LuxR response regulator determinants in mycobacteria. Interestingly, pathogenic members of MTC exhibit three coding DNA sequences (CDS), one of which is *Rv2488c*, presenting an unprecedented architectural organization that simultaneously associate adenylate cyclase and LuxR domain sequences. Being coupled with adenylate cyclase activity, we hypothesized that LuxR regulators may potentially respond to a wider range of environmental signals, enabling the cell to more efficiently cope with deleterious stress.

This study aimed to identify which environmental conditions favor the establishment and maturation of biofilms by MTC bacteria *in vitro* and to explore the microdiversity of LuxR-like *Rv2488c* within circulating *M. tuberculosis* and *M. bovis* strains, also evaluating *Rv2488c* recruitment by the cell through the assessment of its transcriptional profile under the conditions that mimic host microenvironment.

The bioinformatic analyses of *Rv2488c* sequences from 81 MTC publicly available genomes denoted a highly heterogeneous pattern among *M. tuberculosis* strains, specifically through the accumulation of single nucleotide polymorphisms (SNPs) and/or INDELs that were more prevalent at the DNA-binding and cyclase homology domains, while the 25 *M. bovis* isolates under scrutiny depicted high degree of conservation. The polymorphisms found were used to simulate three-dimensional protein models, showing that the correspondent structural architecture is altered when compared to reference *M. tuberculosis* H37Rv, also suggesting functional impairment of the encoded protein in several clinical *M. tuberculosis* strains due to frameshift-causing INDELs. The neighbor joining cladograms inferred from the analyzed *M. bovis* and *M. tuberculosis* sequences thus suggest different evolutionary pressure and microevolutionary history of *Rv2488c* among these two MTC ecotypes, possibly reflecting host niche adaptation.

Batch culture growth of *M. bovis* BCG Tokyo, *M. smegmatis* mc^2 155 and two field *M. bovis* strains (I and II) under *in vitro* environmental conditions mimicking host's microenvironment and phagosomal maturation, including low pH, magnesium or iron starvation, reactive oxygen intermediates, acetate as alternative carbon source and toxic rifampicin challenge, was tested in modified Middlebrook 7H9 broth to disclose the associated physiological and transcriptional responses. Viable cell count, survival percentage, specific growth rate and OD_{590} of crystal violet-stained attached cells, were determined to compare mycobacterial physiology upon stress. Exposure to acid (pH 5.5) using sodium pyruvate as carbon source impaired *M. bovis* growth, leading to slower metabolism (two-fold reduction in specific growth rate) and severely affecting the final amount of biomass in suspension (reduction of viable cell count in five orders of magnitude and survival percentage to 0.00127%, as compared to control). Growth of *M. bovis* under the presence of rifampicin 25 μg/ml decreased survival percentage to 50%, although the observed effect in growth rate was less pronounced than the one induced by low pH.

Mycobacterial cells formed reticulated pellicles at the air-medium interface and attached to the periphery of 12 well-microtitre plates under specific stresses. Considering the effect of rifampicin challenge on attached growth, although survival percentage significantly decreased for three Mb strains upon antibiotic exposure, cell attachment was similar to control, which, considering the biofilm/survival ratio, suggests that rifampicin may enhance biofilm formation by MTC bacteria as a mean to cope with stress. When cells were grown under magnesium or iron starvation, slower growth patterns and final biomass yields were registered, while mycobacterial suspensions depicted macroscopically visible cord-like structures, suggesting that competition for oligoelements that are crucial for fundamental cellular activities stimulate cell aggregation and thus may stimulate biofilm coordinated responses.

Transcriptional profiling of *Rv2488c* in the established stress conditions revealed differential transcriptional responses prone to variation among two *M. bovis* field strains and dependent on environmental cues, with pH 5.5 exerting 131-fold induction of *Rv2488c*. Growth under iron or magnesium starvation or upon rifampicin challenge did not affect *Rv2488c* expression.

The construction by specialized transduction of a knock-out isogenic mutant to study *Rv2488c* role in mycobacteria physiology was also envisaged and advanced, but not fully completed in the scope of this work. Comparison of *in vitro* growth of wild-type and isogenic knock-out strains in batch assays and macrophage cell lines and, also *in vivo* using an animal model of infection, would enable unveiling the possible role of *Rv2488c* to bacilli survival inside the host, considering the mycobacteria-host axis.

Altogether, findings from this work suggest that, in contrast with *M. bovis*, *Rv2488c* has accumulated several mutations in *Mycobacterium tuberculosis* contemporary strains, some of them damaging its regulatory domains and destroying its protein-coding potential, which suggest pseudogenization of this genomic region. However, in *M. bovis*, *Rv2488c* is transcribed and preliminary findings from our group suggest that its transcription may be articulated with global gene expression networks, suggesting that it may ensure a significant biological function to be disclosed in the future.

Keywords: tuberculosis, *Mycobacterium tuberculosis*, *Mycobacterium bovis*, biofilms, *Rv2488c*

RESUMO

O complexo *Mycobacterium tuberculosis* (MTC) inclui várias espécies patogénicas, filogeneticamente próximas e com evolução clonal, que se adaptaram a nichos ecológicos e hospedeiros específicos. Neste grupo de bactérias, incluem-se *M. tuberculosis*, o principal agente etiológico da tuberculose humana (TB), e *M. bovis*, o agente da tuberculose bovina (bTB), sendo esta uma das mais importantes zoonoses a nível mundial. Permanecendo como uma das doenças infeciosas com maior impacto, tendo a Organização Mundial de Saúde estimado em 2013 que um terço da população humana estaria latentemente infetada, a tuberculose traduz-se num flagelo de saúde pública, com elevadas taxas anuais de incidência e de mortalidade, especialmente em indíviduos imunocomprometidos, nomeadamente co-infetados com o vírus da imunodeficiência adquirida (HIV), causando ainda elevados prejuízos monetários.

Pensa-se que a patogenia das espécies do MTC é influenciada por numerosos fatores, tais como fatores genéticos do hospedeiro e do microrganismo, assim como por fatores ambientais. Durante a infeção, e focando apenas a tuberculose pulmonar, uma proporção de bacilos infetantes é capaz de sobreviver e multiplicar-se com sucesso na presença das condições de stresse encontradas no interior do macrófago alveolar. Os atuais modelos de persistência micobacteriana invocam a presença de populações de células dormentes que se reativam e disseminam após imunosupressão do hospedeiro.

Os biofilmes são constituídos por células microbianas imersas numa matriz de exopolissacáridos que conferem proteção significativa aquando da fixação da comunidade a um ambiente desfavorável. Nestas microcomunidades distinguem-se subpopulações heterogéneas de bactérias que demonstram um caráter recalcitrante, como tolerância a xenobióticos e resistência aumentada a condições ambientais inóspitas. A comunicação intercelular (*quorum sensing*, QS) nestas condições, mediada por sinais químicos específicos, permite às populações bacterianas cooperar fisiologicamente e ajustar a expressão génica e os níveis metabólicos em consonância para melhor persistirem no ambiente.

Do ponto de vista clínico, a formação de biofilmes está tipicamente associada a infeções crónicas, conduzindo a um aumento da resistência microbiana a antibióticos e protegendo o agente patogénico da resposta imunitária do hospedeiro, pelo que a agregação das células nestas estruturas pode ser entendida como uma estratégia de sobrevivência, persistência e resistência a stresses ambientais subversivos. Devido à sua natureza hidrofóbica, as micobactérias tendem a formar estes agregados de células *in vitro*, sendo ainda desconhecido se adotam *in vivo* um fenótipo semelhante ao dos biofilmes, estando também por identificar os determinantes genéticos que promovem uma resposta de crescimento em biofilmes. A análise comparativa de genomas de micobactérias saprófitas e bactérias do MTC sugere que o gene *Rv2488c* seja um presumível regulador transcricional pertencente à família LuxR, a

qual inclui genes que codificam proteínas envolvidas em fenómenos de regulação da expressão genética mediados por *quorum sensing*. A análise, por pesquisa de homologia nas bases de dados, dos domínios funcionais subjacentes a esta sequência de DNA codificante sugere ainda a presença de uma região associada à atividade de adenilato ciclase, pelo que se presume que a proteína resultante exercerá atividade regulatória da expressão génica de forma específica em resposta a diferentes estímulos ambientais.

Este trabalho teve como objetivo identificar as condições ambientais que favorecem a formação de biofilmes por bactérias do complexo *M. tuberculosis* e explorar a microdiversidade do gene *Rv2488c* nas estirpes circulantes de *M. tuberculosis* e *M. bovis*, avaliando-se ainda a sua expressão em condições semelhantes às presentes no hospedeiro e/ou que propiciam o estabelecimento e maturação de biofilmes.

A análise bioinformática dos polimorfismos do gene *Rv2488c* em 81 estirpes do complexo *M. tuberculosis* cujos genomas se encontram disponíveis nas bases de dados internacionais demonstrou que a sequência desta região codificante apresenta elevado grau de conservação entre as estirpes de *M. bovis*, enquanto que nas estirpes de *M. tuberculosis* contemporâneas e atualmente em circulação existe grande variação nucleotídica, com acumulação de polimorfismos de nucleótido único (SNPs), inserções e eliminações em regiões específicas. A acumulação de variações nucleotídicas regista particular incidência nos domínios de ligação ao DNA e de adenilato ciclase, o que, à primeira vista, sugere pressão evolutiva nestes domínios, eventualmente para refinar a resposta a diferentes estímulos ambientais, permitindo uma expressão genética mais controlada e específica. A modelação da estrutura tridimensional das proteinas resultantes visou avaliar as presumíveis consequências estruturais e funcionais decorrentes da acumulação dos polimorfismos identificados, verificando-se que as alterações mais frequentes provocam um rearranjo tridimensional distinto, com alteração espacial e posição distal dos domínios de ciclase e LuxR, o que sugere consequências funcionais na atividade da proteína, nomeadamente na sua capacidade de interação com ligandos específicos. Foram também detetadas alterações das sequências nucleotídicas de *Rv2488c* em estirpes clínicas de *M. tuberculosis* que levam à disrupção da grelha de leitura e interrupção precoce da tradução da proteina codificada, afetando drasticamente, e eventualmente inativando, a sua potencial função.

A partir dos genomas disponíveis nas bases de dados de *M. bovis* e *M. tuberculosis* foram construídos cladogramas para avaliação da microevolução de *Rv2488c*, procedendo-se à sua análise filogenética. Os resultados obtidos sugerem diferentes padrões microevolutivos deste gene nos dois ecotipos do MTC que possivelmente refletem a adaptação específica dos dois agentes etiológicos aos respetivos hospedeiros.

Realizou-se o crescimento em descontínuo de *M. bovis* BCG Tokyo, *M. smegmatis* mc^2 155 e duas estirpes de campo de *M. bovis* (I e II) em meio Middlebrook 7H9 modificado com o

objetivo de caraterizar a resposta fisiológica e transcricional destas estirpes *in vitro* a condições ambientais que mimetizam o microambiente do hospedeiro e a maturação do fagossoma, tais como stress ácido, a presença de intermediários reativos de oxigénio, acetato como fonte de carbono, concentrações tóxicas de rifampicina e concentrações residuais de magnésio e ferro. Para determinar e comparar a resposta fisiológica das várias estirpes após exposição às condições impostas, foram determinadas taxas específicas de crescimento, número de células viáveis, percentagens de sobrevivência e densidades ópticas a 590 nm das frações celulares aderidas a superfícies e coradas com violeta de cristal. A exposição a choque ácido (pH 5.5) usando piruvato de sódio como fonte de carbono afetou o crescimento de *M. bovis*, induzindo a desaceleração do metabolismo (taxa específica de crescimento reduzida em duas vezes) e afetando severamente a biomassa final em suspensão, com a redução da contagem de células viáveis em cinco ordens de magnitude e uma percentagem de sobrevivência de 0.00127%, comparando com o controlo. O crescimento de *M. bovis* na presença de 25 µg/ml de rifampicina levou a um decréscimo da percentagem de sobrevivência de 50%, apesar do efeito observado na taxa de crescimento ter sido menos pronunciado em comparação com o induzido pelo choque ácido.

As culturas de micobactérias formaram películas reticuladas na interface entre o meio de cultura-ar e as células fixaram-se na periferia dos poços das microplacas sob stresses específicos. Considerando o efeito da rifampicina na adesão das células aos poços das microplacas, apesar da percentagem de sobrevivência após exposição ao antibiótico ter decrescido significativamente para três estirpes de *M. bovis*, a fração celular aderida foi similar à do controlo, o que, considerando o rácio entre biofilme/sobrevivência, sugere que a rifampicina poderá aumentar a formação de biofilmes por bactérias do MTC como mecanismo para lidar com o stress imposto. Quando as células foram expostas a concentrações residuais de magnésio ou ferro, registaram-se padrões de crescimento e rendimentos de biomassa final formada mais baixos, no entanto visualizaram-se estruturas macroscópicas semelhantes a corda em suspensão, sugerindo que a competição por nutrientes minerais essenciais a atividades celulares fundamentais estimula a agregação celular e possivelmente uma resposta coordenada através da formação de biofilmes.

O perfil transcricional do gene *Rv2488c* em duas estirpes de campo de *M. bovis* nas condições de stress estabelecidas apresentou diferenças, sugerindo diferentes respostas transcricionais aos estímulos ambientais, com o stress ácido (pH 5.5) a exercer uma indução do gene alvo em 131 vezes numa das estirpes. O crescimento com concentrações residuais de nutrientes minerais ou após exposição a rifampicina não afetou a expressão do gene *Rv2488c* em qualquer das estirpes estudadas.

A construção por transdução especializada de um mutante de eliminação para estudar o papel do gene *Rv2488c* na fisiologia micobacteriana foi delineada e avançada, embora não

inteiramente concluida no âmbito do presente trabalho. A comparação do crescimento da estirpe selvagem e do mutante isogénico em descontínuo, em linhas celulares de macrófagos ou num modelo animal de infeção poderia permitir desvendar o possível papel do gene para a sobrevivência do bacilo no interior do hospedeiro, considerando o eixo micobactéria-hospedeiro.

Em suma, os resultados deste trabalho sugerem que, em contraste com *M. bovis*, o gene *Rv2488c* tem acumulado várias mutações em estirpes clínicas atuais de *Mycobacterium tuberculosis*, algumas danificando os seus domínios regulatórios e destruindo o seu potencial de codificação de uma proteína funcional, o que sugere pseudogenização desta região genómica. No entanto, em *M. bovis*, o gene *Rv2488c* é transcrito e resultados preliminares do nosso grupo sugerem que a sua transcrição pode ser articulada com redes globais de expressão génica, sugerindo que este gene poderá assegurar uma função biológica importante a ser desvendada no futuro.

Palavras-chave: tuberculose, *Mycobacterium tuberculosis*, *Mycobacterium bovis*, biofilmes, *Rv2488c*

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

TB – Tuberculosis

bTB – Bovine tuberculosis

MTC – *Mycobacterium tuberculosis* complex

HIV/AIDS – Human immunodeficiency virus/Acquired immune deficiency syndrome

WHO – World Health Organization

G+C – guanine-cytosine content

LM – Lipomannan

PIM – Phosphatidylinositol mannosides

LM – Lipomannan

LAM – Lipoarabinomannan

ManLAM – Mannose-capped lipoarabinomannan

PILAMs – Phosphoinositol-capped lipoarabinomannan

PG – Peptidoglycan

AG – Arabinogalactan

MA – Mycolic acids

MAPc – Mycolic acids plus peptidoglycan core

AFB – Acid-fast bacillus

Mb – Mycobacterium bovis

Mtb – *Mycobacterium tuberculosis*

Msm – *Mycobacterium smegmatis*

RD – Region of Difference

N-RD – New Region of Difference

DCT – Distributive conjugal transfer

rRNA – ribosomal ribonucleic acid

DC – Dendritic cell

PAMPs – Pathogen-Associated Molecular Patterns

PRRs – Pattern Recognition Receptors

MR – Mannose receptor

TLR – Toll-like receptor

NF-kB – Nuclear factor kappa-light-chain-enhancer of activated B cells

NOS – Nitric Oxide Synthase

IL-12 – Interleukin-12

Th1 – T helper 1 cell

CCL22 – Macrophage-derived chemokine

MIP-1α – Macrophage inflammatory protein

GAPDH – Glyceraldehyde 3-phosphate dehydrogenase

Ca – Calcium

Mg – Magnesium

Fe – Iron

H_2O_2 – Hydrogen peroxide

RIF – Rifampicin

NO – Nitric oxide

CO – Carbon monoxide

QS – Quorum-sensing

α – alpha

β – Beta

AI – Autoinducer

AIP – Autoinducing peptide

AHL – acyl-homoserine lactones

SAM – S-adenosylmethionine

NCBI – National Center for Biotechnology Information

BLAST – Basic Local Alignment Search Tool

FASTA – Fast Adaptive Shrinkage Thresholding Algorithm

CDD – Conserved Domain Database

SNP – Single Nucleotide Polymorphism

Phyre2 – Protein Homology/analogy Recognition Engine platform

STITCH – Search Tool for Interactions of Chemicals

STRING – Search Tool for the Retrieval of Interacting Genes/Proteins

INDEL – Insertion (IN)/deletion (DEL)

Rpm – Rotation per minute

g – Gravity

cm – Centimiter

v – Volume

l – Liter

ml – Mililiter

µl – Microliter

mM – Milimolar

µM – Micromolar

g – gram

µg – Microgram

ng – Nanogram

CFU – Colony forming units

OD – Optical density

ADS – Albumin Dextrose Sodium

min – Minutes

s – Seconds

h – Hour

RT – Room temperature

NEB – New England Biolabs

Bp – Base pair

Kb – Kilo base pair

EDTA – Ethylenediamine tetraacetic acid

TM – Melting temperature

pDNA – Plasmidic DNA

gDNA – Genomic DNA

V – Volt

kV – Kilovolt

µF – Microfaraday

o/n – Overnight

hyg – Hygromicin

SAP – Shrimp Alkaline Phosphatase

MCS – Multiple cloning site

AES – Allelic exchange substrate

LB – Luria-Bertani broth

ATP – Adenosine triphosphate

cAMP – Cyclic adenosine monophosphate

7H9 – Middlebrook 7H9 broth

7H10 – Middlebrook 7H10 broth

DMSO – Dimethyl sulfoxide

PCR – Polymerase Chain Reaction

Δ – Delta

Cq – Quantitation cycle

qPCR – Quantitative Polymerase Chain Reaction

RT-qPCR – Reverse Transcriptase Quantitative Polymerase Chain Reaction

E. coli – *Escherichia coli*

Å – Ångström

GOI – Gene Of Interest

Ω – Ohm

Δ – Variation

ORF – Open Reading Frame

TF – Transcription Factor

TR – Transcriptional Regulator

CHD – Cyclase Homology Domain

TCA – Tricarboxylic acid

NMS – Non-modulated site

RNAseq – RNA sequencing

ncRNA – Noncoding RNA

CHAPTER I – INTRODUCTION

1.1 Overview of Tuberculosis

1.1.1 Epidemiology of a Worldwide Infectious Disease

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* complex (MTC) species, remains a severe chronic disease of humans and animals, *M. tuberculosis* (Mtb) being the etiological agent of human TB while *M. bovis* (Mb) is responsible for TB in cattle (bovine tuberculosis, bTB) and in other domestic species, wildlife and, less often, in humans[1]. Human TB is as slow-progressing, chronic disease, which generally affects the lungs (pulmonary tuberculosis), although it can affect other organs, such as the liver or intestines (extra-pulmonary tuberculosis). Transmission is mainly airborne, through the expulsion of viable bacilli in air droplets[2]. Concerning bovine TB, clinical disease onset and progression is similar to human cases. Focusing on human TB, this infectious disease is only surpassed by HIV/AIDS in terms of mortality rate, with 2013 official estimations suggesting that a third of world population is latently infected, which corresponds to approximately 1.96 billion people[2]. Also in 2013, the estimated indicators corresponded to nine million new cases worldwide and 1.5 million deaths[2,3]. Predictions from World Health Organization (WHO) suggest a slight decline of new cases and mortality, mainly attributed to a strict policy on surveillance, prevention and novel treatment routines established by many high-income countries. Despite efforts, incidence and mortality rates remain high, especially in immunocompromised people co-infected with HIV, in whom TB active disease is the number one cause of death[2,3]. Although human cases are mainly due to Mtb, Mb transmitted from animals to humans (and, in lesser extent, *M. caprae*) also causes TB in humans, which is commonly referred as zoonotic tuberculosis. Although the number of human cases due to zoonotic transmission is supposed to be negligible in high-income countries, in the under-developed world zoonotic TB may represent a higher burden due to ingestion of contaminated animal products such as unpasteurized milk and a closer contact with infected animals[4]. In the majority of high-income countries, the incidence of bTB is low and new outbreaks are usually geographically localized. New cases can re-emerge in livestock also due to spill-over in areas of contact with wildlife, as several wild species may act as reservoirs for Mb, depending on the region of the world. Bovine TB thus remains an economical and public health concern[1,4]. Strict surveillance of herd movements, regular testing of livestock and slaughter of positive animals remain effective tools for disease control. In low income countries, the disease is often endemic, surveillance is scarce and strict control policies may be absent or rare, resulting in high incidence and prevalence rates[1].

1.1.2 Characteristics of the *Mycobacterium* Genus

The *Mycobacterium* genus, belonging to the Actinobacteria phylum, is the sole member of the *Mycobacteriaceae* family, and includes more than one hundred species[5]. Species of the

Mycobacterium genus are weakly gram-positive, aerobic, non-sporulating bacteria, with high G+C content genomic DNA, highly homologous 16S ribosomal DNA sequences, and harboring a complex and non-canonic prokaryotic cell wall structure[6]. Classification of mycobacteria can also be made considering growth rate, being considered as fast or slow growers if growth occurs within less or more than 7 days, respectively. Genetic homology-based classifications propose several subgroups or clades, one of which is the *Mycobacterium tuberculosis* complex (MTC)[6]. Focusing on pathogenic mycobacteria and the critical TB agents, Mtb and Mb, the first is a human-adapted pathogen without any known significant environmental reservoir, while the second is a pathogen adapted to ruminants. These hosts act as reservoirs that recharge the bacillus burden in disease dissemination. Ecologically contrasting, other mycobacterial species, such as *M. cheloneae*, *M. fortuitum*, and *M. avium,* are harbored in water and soil[6]. In terms of cell wall structure, an inner and outer fraction irradiate from the cell membrane. The outer fraction is composed of freely-associated lipids and proteins, such as phosphatidylinositol mannosides (PIMs), phthiocerol dimycocerosate, dimycolyl trehalose (cord factor), lipomannan (LM), and lipoarabinomannan (LAM). In pathogenic mycobacteria such as Mtb, LAMs are capped with mannose residues and are referred to as ManLAMs, contrasting with non-pathogenic mycobacteria as *M. smegmatis* (Msm) where LAMs are phosphoinositol-capped and referred as PILAMs[7]. These interspersed components are referred as the soluble cell wall components and have a hypothetic role as signaling effectors associated with the disease process[7]. The inner fraction is referred to as the cell wall core: a covalent bond association between a peptidoglycan layer (PG), arabinogalactans (AG, arabinan and galactan) and mycolic acids (MA). This mycolic acid and peptidoglycan core (MAPc) core is insoluble and suggests a role in keeping cell viability[8]. Due to the existence of a mycolic acid layer that is chemically related to wax, mycobacteria species are dyed by Ziehl–Neelsen, and hence called acid-fast bacillus (AFB). This layer also confers MTC cell colonies their characteristic waxy appearance and a tendency for cells to clump and resist dispersion[6].

1.1.3 The *Mycobacterium tuberculosis* complex (MTC)

Based on cladistics, MTC comprises several closely related mycobacterial species and subspecies that are well studied due to their significance in human and animal infections: *M. cannetti*, the ecotypes adapted to humans Mtb and *M. africanum*, and several animal-adapted species, such as Mb, *M. caprae*, *M. microti* and *M. pinnipedii*. More recently, *M. orygis*, *M. mungi*, *M. suricattae*, the Dassie bacillus, and the chimpanzee bacillus have also been proposed to integrate MTC species[9] although formal taxonomic description has not been fully completed. Members of the MTC share 99.9% similarity at the nucleotide level and identical 16S rRNA gene sequences[10]. Focusing on Mtb and Mb, both have similar genome sizes, respectively 4.3 Mb and 4.4 Mb[11,12], although genomic differences between Mb and Mtb have

been found, specifically deletions of the regions of difference (RD) RD4, RD5, RD7, RD8, RD9, RD10, RD12, RD13 and a novel N-RD25[13]. It is thought that MTC originated from Africa, supposedly in the Horn of Africa (Northeast region)[14]. The different members of the MTC are thought to have diverged by clonal evolution due to large sequence and gene elimination events with horizontal gene transfer being considered negligible, until recent studies suggested a role of distributive conjugal transfer as a mechanism of genetic in Mtb[15,16]. Contemporary species putatively have derived from a common *Mycobacterium prototuberculosis* ancestor and different ecotypes arose with tropism for specific host species[17]. Recently, it has been suggested that the ancestor is related to *Mycobacterium canettii*, a non-clonal evolved mycobacteria depicting recombination events. The switch from non-clonal to clonal evolution has been hypothesized as a result of MTC species becoming strict pathogens, with specific host tropisms and niche adaptation[14].

1.1.4 Mycobacteria-host interaction

Focusing on Mtb infections, lung colonization is initiated by bacilli from air droplets that are expelled through coughing or sneezing of individuals with active pulmonary tuberculosis. The infectious dose is estimated as 10 bacilli[18]. After inhalation, alveolar macrophages will try to phagocyte and destroy the pathogen. Although other cell types interact with mycobacteria in the alveolar environment, such as alveolar epithelial type II pneumocytes and dendritic cells (DC, professional antigen presenters), the best studied case of interaction and the most well understood is with alveolar macrophages[12]. The initial contact is established between PAMPs (Pathogen-Associated Molecular Patterns) and PRRs (Pattern Recognition Receptors). LMs, ManLAMs and PIMs are recognized by mannose receptors (MRs) and toll-like receptors (TLRs), which trigger signal transduction pathways, such as the NF-KB signal pathway to promote Nitric Oxide Synthase (NOS) genes and production of molecular effectors such as interleukin 12 (IL-12) (involved in T cell and dendritic cell recruitment, *Th1* differentiation and gamma interferon production), CCL22 (macrophage-derived chemokine) and macrophage inflammatory protein (MIP-1α)**[19]**. Macrophage-engulfed bacilli will reside inside the vacuolar environment of the phagosome, and while some bacilli will be destroyed, other will evade the phagolysosome hostile acidic environment due to phagosome-lysosome fusion impairment by blockage of Ca^{2+} fluxes[20]. At this stage, the bacilli have evaded annihilation by macrophages and strive in the alveolar environment, but cell recruitment occurs and inactivated monocytes, lymphocytes, and neutrophils are relocated to the infection site due to chemotactic factors secreted by macrophages. Despite the combined effort of the whole immune machinery, bacterial destruction is not efficient, and a focal lesion composed of macrophages, lymphocytes and bacteria, called the granuloma, is formed[12]. These focal lesions restrain the bacteria and impair dissemination and, at a certain point, necrosis events will trigger due to

macrophage-loaded bacteria being killed. Subsequently, a caseous core, full of liquefied tissue, forms in the center of the lesion, which will play a role in infection dissemination during active tuberculosis events. This necrotic core is surrounded by immune cells and fibroblasts[21]. In the caseous core, mycobacteria replication is impaired and a dormant lifestyle is attained due to the anoxic conditions and lower pH, which will enable persistence inside the host through a long period of time (dormant tuberculosis)[12]. At times of immunosuppression, the bacteria will be less negatively pressured by the host and will again multiply, potentially damaging the tissues and bursting this necrotized granuloma, which will then favor dissemination to other parts of the lung, other organs, or even other hosts[22].

1.2 Understanding the Chronic Phenotype of TB infection

1.2.1 Effects of Host Conditions on Mycobacterial Dormancy

Inside the host, mycobacteria naturally encounter different stress conditions, such as nutrient starvation, oxidative stress, or decreased pH, which is expected to alter bacterial metabolism and impair growth. Antibiotic selective pressure also exerts growth impairment to surviving subpopulations. Concerning micronutrients, iron (Fe) is an essential growth factor and a co-factor for virtually all bacterial pathogens in several redox reactions, such as in DNA synthesis and respiration. While in the human body, Fe exists abundantly, it is roughly inaccessible for pathogens because of sequestration by host proteins[23]. To obtain Fe, bacteria employ several strategies such as siderophore production that act as Fe chelators or sphingomyelinases that burst erythrocytes to expose hemoglobin-bound Fe. Regarding Mtb, Fe uptake genetic events are not completely unveiled: siderophore (mycobactins and carboxymycobactins) production is documented and translocation of transferrin (a host Fe transporter) to the phagosome with carboxymycobactins enabling Fe removal from transferrin in a GAPDH-dependent mechanism has also been reported[24]. The Fe uptake mechanism is tightly regulated, even under high iron concentration, which causes toxicity due to hydroxyl radicals, showing an important role of Fe in the physiology of Mtb[23]. The low pH encountered inside the granuloma also influences Mtb infection, since excessive proton concentration affects cellular homeostasis. It seems that phagocytosis of Mtb by macrophages may lead to a variation of the pH level inside the phagosome, depending on whether the macrophage has been immunologically activated or not by gamma-interferon, ranging from pH 6.2 in naive cells to 4.5 in activated cells[25]. Several *in vitro* studies show that Mtb has a limited ability to survive at low pH, the minimum values varying with the model system used and with broth composition. Survival under low pH levels is also dependent on the cell wall permeability to protons, since at pH 5.0 the cytosolic pH of mycobacteria is still approximately 7.0, as reported for Msm and Mtb[25]. Studies by these authors show up-regulation of genes involved in cell wall synthesis after low pH exposure in Mtb CDC1551, such as *pks2*, *pks3* and *papA3*, and survival impairment under low pH[26]. A

clear reason for this observation is still unknown. Mg is another important co-factor for bacterial proteins and has been shown to have a protective effect on Mtb growth under acidic levels. In cultures of Mtb grown in a chemically-defined medium (Sautons), when pH decreased from 7.0 to 6.0, growth (assessed by OD_{580}) was maintained when levels of Mg concentration were increased up to 100 μM, suggesting that Mg may be used by proteins that are recruited for bacterial survival under acidic pH. In addition, higher levels of Mg may be necessary to compensate higher acidic conditions, or even to allow normal Mg uptake, since transport systems may be less efficient under low pH[27]. Oxidative stress is another barrier encountered by host-confined mycobacteria, mainly upon immune cell activation. To protect from the oxidative burst generated by macrophages, mycobacteria deploy scavenging systems, antioxidant enzyme production, cytosolic buffers and DNA/protein repair and protection determinants[28]. Low concentrations of H_2O_2 (0.05 and 0.5 mM) are responsible for up-regulation of several genes of Mtb involved in stress response such as *furA* (gene expression regulator) and *katG* (catalase)[29]. These results suggest that mycobacteria respond to host's defenses against infection, allowing adaptation to the host's environment. Other stimuli, derived from antibiotic presence or differential carbon availability, also impair mycobacterial proliferation under host conditions. Rifampicin, a first line antibiotic against TB, binds to the DNA-dependent RNA polymerase β-subunit inhibiting transcription but not inhibiting translation, being efficient only in early stages of transcription since it physically blocks the growing RNA-chain instead of damaging it. Drug resistance in Mtb is associated with chromosomal mutations of genes encoding drug protein target, such as mutations in the *rpoB* gene (β-subunit), although other mechanisms as efflux pumps may be involved[30].

1.2.2 The Biofilm Phenotype and Its Potential Role in Disease

In nature, bacteria tend to prefer an adherent over planktonic-like existence, thus forming biofilms. These structures, typically harboring multispecies communities, are composed of microbial cells attached to a surface and embedded in an extracellular polymeric matrix[31], enabling cell to cell communication and the coordinated control of gene expression in response to environmental challenges. From the mechanistic point of view, biofilms consist of a several-step gene-controlled program that allows bacterial colonization of a surface[31] and a differentiation suite of events over time and space. Biofilm formation and maturation is modulated by cell density and the accumulation of signaling molecules that trigger quorum-sensing- dependent gene expression[32]. Environmental pressures can vary from nutrient starvation, oxygen depletion or pH variation to include cases of biotic relations as predation, among other. Biofilm formation begins with the attachment of planktonic bacteria to a surface (docking phase), where the close proximity of bacteria to the surface, followed by the sum of attractive and repulsive forces, dictates whether or not docking is successful[31]. The type of

surface (biotic or abiotic), its chemical properties [cell type in case of biotic surfaces, or the material type (polypropylene, polyethylene, metal, etc)], the group of microorganisms (single species or multispecies) and environmental/genetic variables (temperature amplitude, pH fluctuation, nutrient availability or starvation, gene expression, etc) influence the initial adherence phase, whereas the latter docking phase is strongly modulated by molecular interactions between adhesins (such as the flagellum, a class I adhesin) and the target surface[31]. After monolayer biofilm formation, bacteria secrete polymeric substances that consolidate the community to the surface. At this point, the bacterial community will start to maturate until a dynamic equilibrium is attained, with a stable core of replicating bacteria and the outermost cells dying or leaving the biofilm to enter a nomadic state in response to environmental pressure[33,34]. Within the host, the biofilm phenotype has great virulence potential, allowing resistance to environmentally harsh conditions, such as antimicrobial compounds, providing physical protection and the ability to evade and modulate the host's immune system[31].

1.2.3 Mycobacteria Infection - Common aspects to the biofilm lifestyle inside a host

In vivo colonization by pathogenic bacteria associated in biofilms has been reported in *P. aeruginosa* causing respiratory infections in cystic fibrosis patients and in several other infections[35]. Regardless of the causative agent, pathogenic biofilms display enhanced resistance to antibiotics, impairment of host's immune response and are extremely difficult to treat, being a ubiquitous microbial strategy for dissemination of stress-tolerant bacteria[36]. Mtb infections have characteristics similar to microbial biofilms, displaying antibiotic resistance, asymptomatic latency and a persistence phenotype associated to a dormant and non-replicating state of infecting microorganisms that enables evasion from the host's immune system. The clinical phenotype of TB thus raises the question of whether or not Mtb latent infections are associated to the biofilm phenotype *in vivo* and if these multicellular structures may allow persistence of infection and recalcitrance to xenobiotic through time[37]. The *DosR* regulon, associated with Mtb latency, is a set of 48 genes regulated by a two-component signaling system (*DosS/DosT/DosR*) serving as a sensory and transduction mechanism for the bacilli to counteract respiration-impairing NO and CO species[38,39]. Genes of this regulon are thought to be associated with hypoxia adaptation, such as *Rv0570* (ribonuclease reductase, *NrdZ*) and *Rv2007c* (ferredoxin, *FdxA*)[40]. Although persistency through biofilm formation has not been confirmed *in vivo*, it has been reported that Mtb growth *in vitro* in Sauton's medium displays biofilm-like pellicles[35]. The same study hypothesized genetically programmed differences between planktonic and biofilm growth, implying a role of *helY* (encoding DNA helicase), *pks16* (coding for polyketide synthase), *Rv2454c* and *Rv2455c* (the locus encoding 2-oxoglutarate dehydrogenase) in biofilm formation, based on the phenotypes

of transposon mutants in those genes with complementation assays, although interactions and the specific genetic programme is still unknown. On the other hand, host colonization by *M. leprae*, a nontuberculous mycobacteria causing Buruli ulcer, has been suggested to occur *in vivo* as multicellular communities[41,42].

1.2.4 Quorum-sensing in Mycobacteria

Quorum-sensing (QS) mechanisms consist of a series of cell-to-cell communication events through molecular stimulae by signaling molecules (autoinducers - AIs) that allow bacteria to respond to fluctuations of population density and interact with the environment and with each other (intra- and interspecifically) in order to physiologically adjust to challenging external conditions such as nutrient starvation, competition, antibiotic presence and others. In general, when the excreted AI concentrations reach a certain threshold, bacterial receptors in the membrane or cytoplasm (depending on the QS system) are deployed and signal transduction triggers changes in gene expression[32]. Phenomena such as biofilm formation, virulence, sporulation, competence or antibiotic resistance are under the control of such density-dependent mechanisms in many bacterial genera. The QS circuits are usually different between gram-positive and gram-negative bacteria, while hybrid systems have been described in species as *V. harveyi* and *V. cholerae*[32]. Canonical QS machinery in gram-positive deploy an autoinducing peptide (AIP) that is processed and a two-component membrane histidine kinase receptor that detects the mature AIPs at high cellular density and induces autophosphorylation; and subsequently a cytoplasmic response regulator that receives the signal; in gram-negative, the widespread QS system consists of acyl-homoserine lactones (AHL) or SAM (S-adenosylmethionine) derived small molecules as AI (encoded by *luxI* genes) that interact with cytoplasmic receptors, activating their role as transcription factors (encoded by *luxR* genes). Although these are considered prototypical systems, some bacteria possess slight changes in the QS circuits in terms of AI processing or signal sensing and transduction[32]. LuxR family of proteins can be divided according to biochemical properties: class I (AI-dependent folding with tight binding), class II (AI-dependent folding with reversible binding of AI) and class III (non-dependent folding and reversible binding of AI). Target genes also vary in terms of promotor binding affinity (low/high affinity promoters require high/low LuxR concentration). Homologous LuxR proteins that lack the cognate LuxI are designated as homologous LuxR orphans and play a role in regulating the already existing communication network[43]. Contrasting with the well-studied QS circuits in many pathogenic bacterial species, mycobacterial QS circuit remains poorly explored. A recent bioinformatic analysis of Actinobacterial LuxR protein genes encoded in the genomes of 53 species revealed the presence of putative response regulator sequences with different architectural organizations dispersed along the phylum[44]. It is important to highlight the simultaneous association of

adenylate and LuxR domains within the analyzed Actinobacteria genomes, an architecture that is restricted to members of the *Mycobacterium* genus and is thought to be associated with virulence processes[44]. Being coupled with adenylate cyclase activity, LuxR regulators may thus potentially respond to a variety of environmental signals in order to more efficiently respond to stress and deleterious situations. Evidence of QS in the *Mycobacterium* genus is sparse and limited to indirect results, although some studies have shown evidence of LuxR-type genes in virulence, such as *MAP0482* that is hypothetically linked to *M. avium* host adaptation[45]. Mtb H37Rv genome putatively encodes seven LuxR-like genes, specifically *Rv0195*, *Rv0386*, *Rv0491*, *Rv0890c*, *Rv0894*, *Rv1358*, *Rv2488c* and *Rv3133c*[43]. Interestingly, *Rv0491* is conserved across 16 non-tuberculous and tuberculous mycobacterial species (including Msm and *M. avium*), which suggests a possible conserved mechanism of QS within the genus[43].

1.2.5 Rv2488c, a hypothetical LuxR-like transcriptional regulator

In the present work, we focused on the analysis of *Rv2488c* of Mtb (the homolog in Mb is *Mb2515c*) that putatively encodes a LuxR-like transcriptional regulator and whose deduced amino acid sequence contains an unusual combination of protein domains, including domains that are typical of adenylate cyclase activity and P-loop domains. Based on sequence and structure, *Rv2488* thus appears to be involved in transcriptional regulation and response to quorum sensing-like stimuli, although these roles have not been investigated experimentally. Previous studies have shown that a mutant in *Rv0918c* (hypothetic metalloendopeptidase), exhibiting enhanced virulence in a murine model, exhibits 59-fold up-regulation of *Rv2488c*[46]. *Rv2488c* was also found to be two-fold down-regulated in a mutant of Mtb H37Rv for the two-component system *senX3-regX3*; this strain has a growth defect inside macrophages and displays an attenuated virulence phenotype[47]. Indirect studies thus suggest that *Rv2488c* may also be associated with virulence in Mtb infection.

1.2.6 Objectives of the present work

Since in other genera LuxR-like genes have been related with transcriptional regulation triggered by quorum sensing phenomena, and given the fact that biofilm formation may be a strategy employed by mycobacteria to face environmental challenges inside the host, or to go dormant and unnoticed during infection, we sought to explore the conditions that may trigger the biofilm phenotype in *M. smegmatis* mc^2 155, Mb BCG Tokyo and Mb field strains. Simultaneously, we aimed to explore the sequence diversity, evolution and function of *Rv2488c* during *in vitro* planktonic or biofilm-associated growth. We deployed an *in silico* genome analysis approach comparing *Rv2488c* gene and protein sequences across the available genomes of MTC species. We also used an *in vitro* 12-well microtitre plate growth

assay and exposed mycobacterial cells to stress conditions that potentially mimic those found inside the host upon infection: acidic levels, different carbon sources, antibiotic challenge, micronutrient starvation and oxidative stress. Following, we performed the transcriptomic analysis of *Rv2488c* in two field Mb strains under the tested stress conditions using reverse transcriptase real-time PCR and comparing *Rv2488c* expression in relation to reference *16S* ribosomal RNA gene. A phage-mediated mutagenesis approach was also initiated in order to construct an *Rv2488c* mutant, which is still ongoing.

CHAPTER II – MATERIALS AND METHODS

2.1 *In silico* analysis of *Rv2488c* sequences from publicly available genomes

Bibliographic and sequence information on *Rv2488c* was compiled from open-source databases, namely Tuberculist[48], TB Database[49] and MTB Network Portal[50] using as search entry "Rv2488c" (accessed at 25/02/2015). While we were recently aware that a second annotation of Mb AF2122 reference strain genome has been launched in the second semester of 2015, and is currently ongoing, the present dissertation was based on data from the first annotation. Regarding Mtb, there are currently 26 versions of the annotated genome[48].

2.1.1 *Rv2488c* nucleotide and protein sequence alignments and microevolutionary analysis of *Rv2488c* in MTC strains based on cladistics

Rv2488c nucleotide and protein sequences from available genomes deposited in the *National Center for Biotechnology Information* (NCBI) database were retrieved with the *Basic Local Alignment Search Tool* (BLAST) using the Megablast (nucleotide) and blastp (protein) algorithms, deploying as query sequence the nucleotide/protein sequence of *Rv2488c* from Mtb H37Rv (Supplementary Table 1). Among the sequences retrieved from BLAST search, only complete sequences of genome and protein were considered. The *Fast Adaptive Shrinkage Thresholding Algorithm* (FASTA) sequences for nucleotide (25 for Mb and 56 for Mtb) and protein (39 MTC strains) were aligned separately in ClustalX 2.1 against H37Rv reference gene and protein sequences. Total number of alterations in each sequence was registered, clustered and used to infer the most affected domain(s). Domain architecture was obtained inputting *Rv2488c* nucleotide sequence from H37Rv in the *Conserved Domain Database* (CDD) tool available in NCBI, considering displayed architecture as reference. Neighbor joining-based cladograms generated from the alignments were edited using FigTree v1.4.2. and SNPs and INDELs, were registered in colors next to each strain's branch. Interaction of Rv2488c protein with other proteins and molecules was accessed using the *Search Tool for the Retrieval of Interacting Genes/Proteins* (STRING[51]) and *Search Tool for Interactions of Chemicals* (STITCH[52]), with confidence score of 0.4.

2.1.2 SNP analysis and modeling of 3D protein structure

Predicted 3D protein models were built, using Protein Homology/analogy Recognition Engine platform (Phyre2[53]) in intensive modeling mode, for Mtb strains with SNP c.797C>T and Mb strains with SNP c.1605C>G. These SNPs were the most shared among nucleotide sequences, summing up seven strains for Mtb and five for Mb, and 3D protein models were compared against the corresponding Mtb H37Rv (publicly available sequence used) and Mb AF2122-97 (deduced sequence used) 3D protein models.

2.2 Mycobacterial physiology

2.2.1 Bacterial strains, growth conditions and culture maintenance

The -70ºC frozen stocks of Msm mc^2155, Mb BCG Tokyo and two Mb field isolates (I and II) (Supplementary Table 2) were inoculated into Middlebrook 7H9 (BDDifcoTM)+ADS 10% (v/v) (dextrose 20 g/l, sodium chloride 8.5 g/l and bovine albumin fraction five 50 g/l), and 2 ml/l of glycerol 86% (for Msm) or sodium pyruvate 5 g/l (for Mb strains) and incubated at 37ºC with agitation (150 rpm) during one to two months. After entering the exponential growth phase (OD_{600nm} of 0.5), the culture was re-inoculated onto solid media [Löwenstein–Jensen (Biogerm) or Middlebrook 7H10 (BDDifcoTM)+ADS 10% and glycerol or sodium pyruvate]. After growth, a single colony was inoculated into 7H9 broth with sodium pyruvate and left to grow until the exponential phase was reached (two to three weeks). This suspension was subsequently used as inoculum in batch culture assays to define growth curves and in biofilm experiments that were performed on 12-well microtitre plates.

2.2.2 Culture conditions within selected host-like conditions

For pH experiments, 7H9 broth was supplemented with acetic acid 5 M, to final pH of 4.5, 5.0 or 5.5 and sterilized thereafter. When studying the effect of acidic pH and oxidative stress together, H_2O_2 was added to final concentration of 9 mM to 7H9 + pH 5.5. When using sodium acetate as carbon source, 7H9 was prepared omitting glycerol/sodium pyruvate and adding sodium acetate to final concentration of 2 mM, 3 mM or 4 mM. To study RIF effect, a 1 mg/ml stock solution was prepared, filter-sterilized and 7H9 supplemented to final concentration of 25 µg/ml (Mb) or 50 µg/ml (Msm). To determine RIF sub-inhibitory concentration for both bacterial species, several rifampicin concentrations based on prior references were chosen: 0.05, 0.1, 0.5, 1, 10, 50 or 100 µg/ml tested for Msm; 0.5, 5, 10, 25 or 50 µg/ml tested for BCG Tokyo. To study the effect of Mg and Fe starvation, modified 7H9 broth was prepared: glycerol (2ml/l)/sodium pyruvate (5 g/l), monohydrated L-asparagine 0.5 g/l, KH_2PO_4 1 g/l, Na_2HPO_4 2.5 g/l and treated overnight (o/n) with Chelex-100 8 g/l (Biorad) to chelate metal ions. After filtration (0.22 µm pore), essential ions were added back into the base medium: for Mg starvation, $CaCl_2$ 0.5 mg/l, $ZnSO_4.7H_2O$ 0.1 mg/l, ferric ammonium citrate 50 mg/l and

$MgSO_4.7H_2O$ to defined concentration of 5 µM was added; for Fe starvation, $CaCl_2$ 0.5 mg/l, $ZnSO_4.7H_2O$ 0.1 mg/l, $MgSO_4.7H_2O$ 50 mg/l and $FeCl_3.6H_2O$ to defined concentrations of 2 µM, 10 µM or 50 µM was added. Each broth was supplemented with ADS 10%.

2.2.3 Mb II growth under acidic and RIF presence

Growth of Mb II under acidic and antibiotic challenging condition was accompanied during two weeks in cell culture flasks (600 ml, Corning Inc.) containing 100 mL of 7H9+ADS 10% supplemented like previously described. Pre-inoculum was grown until exponential phase and inoculated with OD_{600} of 0.1. Growth was followed during two weeks at 37°C, static. Within regular time-points, absorbance at 600 nm was measured and serial dilutions were plated onto 7H10+sodium pyruvate+ADS 10%, to determine the number of colony forming units (CFU/ml). Growth under stress conditions was compared against the non-stress condition (plain 7H9 broth). Results are means of values from two independent experiments with respective standard deviation.

2.3 Quantification of formed mycobacterial biofilm under challenging conditions

Biofilm growth was assessed in 12 well-microtitre plates (Corning) (one week at 37ºC for Msm and three weeks for Mb strains) using modified 7H9 medium according to specific conditions: Mg (5 µM) or Fe (2 µM, 10 µM or 50 µM) starvation, RIF challenge (50 µg/ml for Msm and 25 µg/ml for Mb strains), acidic pH (pH 4.5, 5.0 or 5.5), joint oxidative and acidic stress (9 mM H_2O_2 and pH 5.5) and sodium acetate as alternative carbon source (2 mM, 3 mM or 4 mM) to glycerol or sodium pyruvate. Mb I and II were manipulated in a biosafety level 3 lab and tested only for pH acidification and RIF presence. After culture growth, two random wells were selected from each conditions, following 1:10 serial dilutions and plating onto 7H10+ADS 10%+glycerol/sodium pyruvate and incubated at 37ºC until colonies were visible (two days for Msm and five days for all Mb strains). The number of CFUs from each condition was compared to the number of CFUs yielded from the initial inoculum to determine survival percentage. Quantification of formed biofilm was based on the method of O' Toole and collaborators (1999): each well was washed three times with 2 ml of sterile water, dried, treated with 2 ml crystal violet (1%) for 15 minutes (min) that subsequently was discarded, ressuspended in 2 ml ethanol for another 15 min and absorbance read at 590 nm. A minimum of three independent experiments were performed for each strain and culture condition, except for Mb I and II, for which only two independent experiments were performed.

2.4 Transcriptional analysis of *Rv2488c* under stress conditions

2.4.1 Bacterial growth conditions, RNA extraction and quantification

Mb BCG Tokyo and Mb I and II were grown until exponential phase and inoculated into 50 ml flasks (Nunclon®) with 10 ml 7H9+ADS 10%+sodium pyruvate with initial OD_{600} 0.05. After reaching OD_{600nm} of 0.3, stresses were added (15 µl of sterile acetic acid 5 M to adjust pH to 5.5 and 250 µl of RIF stock solution of 1 mg/ml to final concentration of 25 µg/ml). After 48 hours (h), cell cultures were centrifuged at 12000 rpm, 10 min, 4ºC. Pellet was ressuspended in 750 µl of TRIzol® (Life technologies) and frozen at -70ºC until use. To extract RNA, pellets were unfrozen five min at room temperature (RT) followed by bead-beating (FastPrep FP120 Bio101, Savant Instruments) (45 s at 6.5 speed, two times). Afterwards, cells were kept 5 min at RT and 200 µl of chloroform was added, vigorously shaked for 15 s, followed by two/three min at RT and centrifugation at 12000xg, 15 min, 4ºC. The upper colorless phase (soluble RNA) was removed to a new 1.5 ml microtube. RNA samples were mixed with 500 µl of isopropanol 100%, followed by 10 min RT incubation and centrifugation at 14000 rpm, 10 min 4ºC. To wash RNA, supernatant was discarded and pellet washed with ethanol 75%, with sample centrifugation at 9500xg, 5 min, 4ºC, discarding the supernatant. Samples were air-dried for 10 min. RNA was ressuspended in 30 µl of RNAse-free water, incubated at 55ºC 10 min and stored at -70ºC. RNA quantification was performed using a Qubit™ fluorometer (Invitrogen), following manufacturer's instructions, and purified using DNase I (NEB) and incubating samples for 2 h at 37ºC. DNase I treatment was ended adding ethylenediamine tetra-acetic acid (EDTA) 5 mM and incubating samples for 10 min at 75ºC.

2.4.2 Quantitative Reverse Transcriptase Reaction Conditions

For *Rv2488c* transcriptional profiling under stress conditions, primers (Supplementary Table 3 targeting the 3' end were designed using Primer3 and diluted into 10 µM work solutions. Initial optimization was performed with conventional PCR using DNA as template, variating Mg levels (3 and 3.5 mM) and melting temperature (TM) (55ºC), using NZYTaq 2x Green Master Mix (Nzytech) and adjusting manufacturer's protocol to expected amplicon size (136 bp). For DNA extraction of H37Rv, colonies on solid media (Lowenstein Jensen pyruvate) were suspended in TE (10 mM TrisHCl, 1 mM EDTA) 1 M pH 8.0 and heated at 99°C, 30 min. After centrifugation (1500xg, 5 min), the suspension was transferred into a new microtube and stored at -20ºC. With visualization gel of amplicons with expected size in a 3% agarose, quantitative PCR was performed using SsoFast™ EvaGreen® Supermix (Biorad) adjusting manufacturer's protocol to expected amplicon size, using H37Rv DNA as template and running amplification products in a 3% agarose gel to confirm amplicon presence. Quantitative reverse transcriptase PCR (RT-qPCR) was performed using iTaq™ Universal SYBR® Green One-step kit. Each reaction was performed in a final volume of 10 µl (5 µl of reaction mix, 0.125 µl iScript Reverse Transcriptase, 0.3 µl of each primer, 25 ng of RNA and water was added to complete volume).

PCR amplification program consisted of one cycle of 30 min, 50ºC, one cycle of 1 min, 95ºC, followed by 50 cycles of 10 s, 95ºC and 30 s, 60ºC with plate read, followed by melting curve analysis of 5 s, from 65ºC until 95ºC with increments of 0.5ºC. Dissociation step performed at the end of the qRT-PCR reaction was meant to determine the TM of the PCR product and to check if there was amplification of a single, specific product. The real-time PCR reactions were performed with negative reverse transcriptase samples to exclude DNA contamination and also with no template controls to exclude cross-contamination of reagents and plates with nucleic acids. All amplifications were done in optical reaction plates (Biorad) in duplicates or triplicates. The expression values of *Rv2488c* were normalized using the ΔΔct method to the *16S* rRNA reference gene and transcript level of *Rv2488c* was expressed as a fold change for the Mb stress-induced sample relative to the un-induced sample. Fold expression was calculated using the formula $2^{-\Delta\Delta ct}$ (ΔΔct - difference between sample Δct and control condition Δct; Δct - Cq of reference gene subtracted from target gene Cq. Equal amplification efficiencies were considered for all reactions.

2.5 Statistical treatment of results

Considering growth assays, for both survival percentage and biofilm quantification, obtained values from tested conditions were compared to controls. When comparing multiple conditions versus the control, a repeated measures ANOVA (Friedman's test, α=0.05) with a Dunn's Multiple Comparison post-test was performed using GraphPad Prism. When comparing one condition versus the control, a two-tailed paired t-student test (Wilcoxon matched pairs test, α=0.05) was performed. For transcriptomic results, a non-parametric ANOVA (Kruskal-Wallis test, α=0.05) with a Dunn's post-test was performed with GraphPad Prism. Results are displayed as means of values of at least three independent experiments with respective standard deviation.

2.6 Phage-mediated mutagenesis

2.6.1 Cosmid construction

Comparing growth rate, biofilm phenotype, survival percentage and gene expression in wild-type versus an *Rv2488c* mutant in Mtb or Mb after challenge with stress conditions, and to test in parallel the induction or repression of target genes, would unveil possible *Rv2488c* role. This work conceptually implied the preparation of a construct in which *Rv2488c* would be disrupted by a hygromycin (hyg) resistance gene (Rv2488c::hyg allele). For complementation, a copy of wild-type *Rv2488c* under its own promoter would be reintroduced into mutant strain. The protocol foreseen for the construction of an *Rv2488c* deletion mutant was based on a specialized transduction strategy and adapted from reference [54] and Paul Golby (unpublished results). Bacterial strains, primers and plasmids are enlisted at Supplementary Tables 2, 3 and

4. Cosmid pYUB854 is a derivative of pYUB572, with a hygromicin cassette flanked by multiple cloning sites (MCS). To construct pYUB854 harboring the allelic exchange substrate (AES), two primer pairs were designed with adaptors for cloning sites matching selected MCS flanking hyg cassete. One pair was used to amplify *Rv2488c* upstream region (877 bp) and cloned into *Spe*I/*Hind*III sites, while the other was used to amplify *Rv2488c* downstream region (962 bp) and cloned into *Xba*I/*Kpn*I sites. Both regions were obtained by PCR, each reaction performed in a final volume of 25 µl (1.25 µl from each primer, 12.5 µl Phusion Mastermix (NEB), 0.75 µl dimethyl sulfoxide and final DNA amount of 50 ng). The PCR amplification program consisted of 1 cycle of 30 s, 98ºC, followed by 30 cycles of 10 s, 98ºC, 30 s, 60ºC, 30 s, 72ºC and a final step of 10 min, 72ºC. PCR products with expected size were observed in a 1% agarose gel and extracted with QIAquick (QIAGEN). Cosmid pYUB854 has a 1.1 kb unknown fragment at *Spe*I/*Hind*III sites, so we extracted the plasmid using QIAprep Spin Miniprep kit (QIAGEN) and performed digestion with *Spe*I/*Hind*III (1 h, 37ºC). Afterwards, digestion was run in a 1% agarose gel and the 3.8 kb band corresponding to purified pYUB854 was extracted using QIAquick. Following, cosmid was dephosphorylated with alkaline phosphatase (NEB), purified with QIAquick and DNA was quantified in NanoDrop2000. We first attempted cloning the *Rv2488c* upstream region, digesting amplicon with *Spe*I/*Hind*III (1 h, 37ºC). Digestion was purified with QIAquick and DNA quantified in NanoDrop2000 (Thermo Scientific). Ligation reaction between dephosphorylated pYUB854 and digested amplicon was performed (1 h, RT) using T4 DNA Ligase following manufacturer's recommendations for vector and insert quantity. Transformation was performed with chemically competent *E. coli* αDH5 strain (NZYStar competent cells, Nzytech) and plated on LB+hyg (50 mg/ml). For confirmation, colonies from transformation plates were replated onto LB+hyg agar and broth. Plasmid DNA was extracted using NZYMiniprep kit (Nzytech) followed by a 1% gel run to confirm cosmid presence. Positive transformant was selected and isolated onto LB+hyg (50 mg/ml) agar and broth, and pYUB854 carrying *Rv2488c* upstream region was extracted with QIAprep Spin Miniprep, to attempt cloning *Rv2488c* downstream region. The followed protocol was identical to cloning attempt of upstream region, using *Xba*I/*Kpn*I restriction enzymes instead.

2.6.2 Construction of shuttle vector

E. coli cultures carrying recombinant pYUB854 and phasmid phAE159 were grown o/n in LB supplemented with hyg and ampicillin (150 µg/ml) to extract pDNA using QIAprep Spin Miniprep, which was digested with *Pac*I. Afterwards, pYUB854 was dephosphorylated with Shrimp Alkaline Phosphatase (*SAP*) for one h, and a 0.8% gel run was performed to extract DNA bands. Extracted pYUB854 DNA was purified using QIAquick and phAE159 purified with QIAEX® II Gel Extraction Kit (QIAGEN). Ligation reaction was performed with purified 1000 ng of phAE59 and 250 ng of pYUB854 (2 h at RT followed by 4°C o/n). At the same time, *E.*

coli HB101 was grown o/n in 5 ml of LB+10 mM $MgSO_4.7H_2$+0.2% maltose, 37°C and 100 rpm. The day after, 25 ml of the same medium was inoculated with the o/n culture with initial OD_{600} of 0.05 until it reached 0.6/0.8 (37°C, 100 rpm). After growth, 1 ml of culture was centrifuged 3 min at 13000 rpm and pellet was ressuspended in 400 µl of MP buffer (50 mM Tris/HCl, 150 mM NaCl, 10 mM $MgCl_2$, 2 mM $CaCl_2$). *In vitro* packaging was performed with 5 µl of o/n ligation reaction placed on top of Gigapack III XL (Agilent Technologies) , incubated 2 h at 20ºC, and 200 µl of SM buffer (5.8g NaCl, 2g $MgSO_4.7H_2O$, 50 ml 1M Tris/HCl, 0.5 ml agar 2%, 94.5ml H_2O) was added. The 1.5 ml microtube was centrifuged for 40 s, placing 200 µl of the supernatant in a new 1.5 mL microtube, remaining 50 µl in the Gigapack 1.5 ml microtube. To both tubes, 200 µl of *E. coli* HB101 were added, following static incubation at 30ºC and addition of 750 µl of LB. The tubes were then incubated at 37ºC, 1 h at 100 rpm and after briefly centrifuged – 900 µl of supernatant and the remaining 100 µl were inoculated into LB+hyg (150 µg/ml) agar and incubated at 37ºC until growth. Any colony was replated and grown in liquid LB+hyg (50 µg/ml), pDNA was extracted, digested with *Pac*I for two h and a 0.8% gel run was performed to check for transformants.

CHAPTER III – RESULTS AND DISCUSSION

3.1 *In silico* Comparative Genomics of *Rv2488c* across MTC bacteria

3.1.1. Predicted organization of the genetic locus and putative function of *Rv2488c*

Mtb *Rv2488c* (*Mb2515c* in Mb) encodes a hypothetical protein with several predicted domains that suggest a multi-branched ontology, from a two-component response regulator, to transcription factor (TF) activity and ATP binding[48]. Encoded in the minus strand (spanning 3414 bp in positions 2797467 and 2800880 of the genome), experimental evidence on the specific regulation that it is under the influence of, or that participates in, is absent, although, as a LuxR-like protein, it is speculated to have a role as a transcriptional regulator (TR) in processes dependent of quorum-sensing phenomena[43]. Assuming that *Rv2488c* is indeed a TR, available information in terms of key regulators and regulation targets, searched across Tuberculist, MTB Network Portal and TB Database, and using as reference Mtb H37Rv genome, is diverse. In terms of regulatory interactions, TB Database predicts that *Rv2488c* is regulated by the products of nine genes, including its own, and has five predicted regulatory targets; searches on MTB Network Portal indicate that Rv2488c upregulates biclusters (set of co-regulated genes) 0141, 0149, 0334, 0435, 0497 and 0538, and is under the regulatory action of products from 11 genes (*Rv1985c*, *Rv3124*, *Rv2884*, *Rv2359*, *Rv2282c*, *Rv1033c*, *Rv0238*, *Rv0042c* and *Rv0260c* positively; *Rv3260c* and *Rv3417c* negatively). Initial analyses in STITCH[52] with confidence score of 0.4 suggest interaction with several genes/proteins and chemicals such as cAMP and glycerol, while on STRING database[51] interaction with nine genes/proteins is predicted. Considering genetic locus localization (Fig. 3.1), MTB Network

Portal indicates a possible operon alongside *Rv2489c*, while TB Database indicates an operon including *Rv2489c* and *Rv2487c*. Of all these, *Rv2487c* and *Rv2489c* have unknown function, although they putatively encode PE_PGRS proteins, a subgroup of PE proteins containing polymorphic GC-Rich Sequences. These proteins, whose nomenclature is given due to abundance of proline and glutamine residues near the N terminal region, account for 10% of the coding capacity of Mtb. Although their functional roles remain enigmatic, these proteins have been associated with antigenic variation and immunological evasion[55]. Other genes nearby *Rv2488c* encode for: transfer RNA (*MTB000032*), an enoyl hydrase for fatty acid oxidation (*Rv2486*) and a carboxylesterase (*Rv2485c*). *Rv2488c* has been identified in the cytosol and cell membrane fraction of H37Rv using bi-dimensional LC/MS[56]. It has also been identified by mass spectrometry in Mtb H37Rv-infected guinea pig lungs at 90 days[57], suggesting a role in TB persistence. Transposon mutagenesis in H37Rv and CDC1551 Mtb strains has evidenced that it is a non-essential gene[58,59]. TB Database suggests that *Rv2488c* may be part of a regulon controlled by *senX3-regX3*, a two-component sensor and regulator system[47] suggested to alter gene expression in aerobic respiration and phosphate depletion[60].

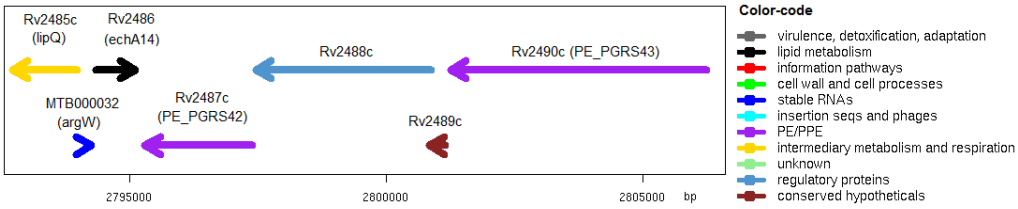


Fig. 3.1 – Genetic locus of *Rv2488c*. Gene function is color-coded. Adapted from Tuberculist.

3.1.2. Predicted domain structure and 3D model of Rv2488c protein

Rv2488c protein encompasses 1137 amino acids. We assessed its domain structure using the Conserved Domains Database (CDD) at NCBI. A Cyclase Homology Domain (CHD) is present in the N-terminal region. In terms of catalytic activity, mycobacterial cyclases are similar to mammalian cyclases, requiring residues for metallic cofactor binding, ATP binding and transition-state stabilizing residues[62]. Mtb H37Rv genome encodes 16 class III cyclases, while CDC1551 strain genome has 17. The CHD possesses a glutamine over arginine residue at position 52 (Fig. 3.2A) suggesting loss of cyclase activity and substrate selectivity, which may suggest a protein more suited for attachment to DNA stretches, consistent with a TR function and in agreement with ontology analysis. Two aspartic acid residues have been identified as possibly responsible for metallic cofactor coupling and an arginine and aspartic acid responsible for transition-state stabilization (once again noting that Rv2488c cyclase domain is suggested as catalytically inactive). In the central region, the protein has a NB-ARC domain (nucleotide binding domain of the AAA^+ superfamily - ATPases Associated with several cellular Activities superfamily, a group of chaperone-like ATPases), although, depending on the

domain prediction algorithm used, an ATPase predicted domain hit can be obtained. To simplify terminology, we assume this domain as NB_ARC. At a subdomain level, this region has other specific sequences that may allow nucleotide binding specificity such as Walker motifs/P-loops A and B[63,64]. At the C-terminal region, a LuxR domain with Helix-Turn-Helix motifs is present and possibly indicates a DNA-binding protein that regulates gene expression[44,63]. No 3D model of the protein was previously available, so in the context of this work the 3D structure of Rv2488c was predicted and modeled for Mtb H37Rv and Mb AF2122-97 using Phyre2 intensive mode. Modeling results for H37Rv and AF2122-97 deduced sequences are shown in Fig. 3.3. The majority of the amino acid sequence (85%) was modeled

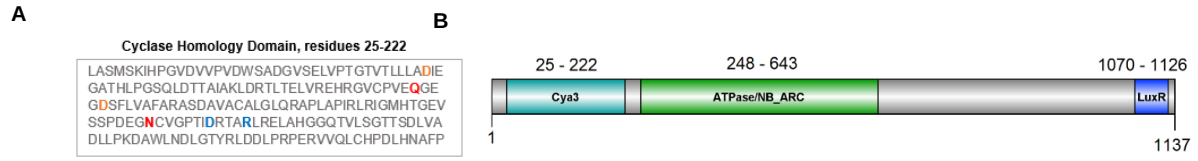


Fig. 3.2 – (A) Cyclase domain of Mtb H37Rv Rv2488c. Key residues are color-coded: metal binding (orange), substrate specifying (red) and transition-state stabilizing (blue); **(B)** Rv2488c domain architecture obtained by CDD.

with a confidence level of over 90%, except for residues 1 to 52, 229 to 233, 601 to 699, 1031 to 1058 and 1133 to 1137, for both modeled reference strains. We thus assume that the computed model may be a good approach for the actual *in vivo* structure of Rv2488c.

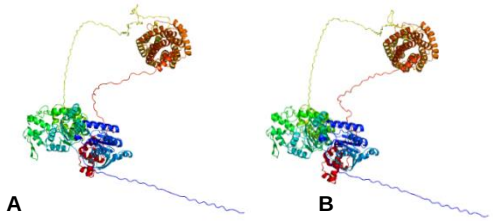


Fig. 3.3 - Predicted 3D models of Rv2488c for *M. tuberculosis* H37Rv (**A** - Model dimensions (Å) X: 103.126 Y: 219.865 Z: 190.547) and *M. bovis* AF2122-97 (**B** - Model dimensions (Å) X: 100.666 Y: 220.192 Z: 192.594). Color-code: Cyclase domain (green/blue), NB_ARC (yellow/brown), LuxR-like (red).

3.1.3 Comparative sequence analyses of *Rv2488c* across MTB genomes

Following, we analyzed if nucleotide and amino acid sequences are conserved across MTB sequenced genomes. From the 81 nucleotide sequences retrieved (25 from Mb and 56 from Mtb), we annotated and compiled every single difference in terms of position, single nucleotide polymorphisms (SNP) or insertions (IN) or deletions (DEL) in comparison with Mtb H37Rv reference genome, and clustered the differences by domain and non-modulated DNA stretches between domains (referred as Non-Modulated Stretches, NMS, from 1 to 4). Results from comparative nucleotide analyses (Fig. 3.4A) are summarized in Supplementary Table 5, while results from amino acid sequence analysis (39 sequences curated with the same approach and clustered by domain) are summed up in Supplementary Table 6. From whole nucleotide sequence alignments, we only detected seven SNPs and no IN or DEL among Mb

strains, while 114 alterations were detected across Mtb (15 IN, 12 DEL and 87 SNPs). Total alterations in both species consisted of 121: 22.3% INDEL [15 IN (12.4%) and 12 DEL (9.9%)] and 94 SNPs (77.7%). Mtb CAS-NITR204 and Haarlem-NITR202 genomes exhibited the highest accumulation of alterations (in NMS3 and NB_ARC domains, respectively), suggesting high selective pressure towards this genomic region in these two world circulating clinical strains. SNPs c.647A>G, c.797C>T, c.1400T>G, c.1605C>G and DEL c.1404_1408delGGTGC were common to two or more sequences, while remaining alterations were orphan. In order to understand SNP effect on the predicted 3D protein structure, two of these SNPs were selected (c.797C>T for Mtb and c.1605C>G for Mb) to remodel protein sequence based on higher frequency of these alterations among analyzed strains. Corresponding deduced amino acid sequences were modeled by Phyre2 in intense mode with

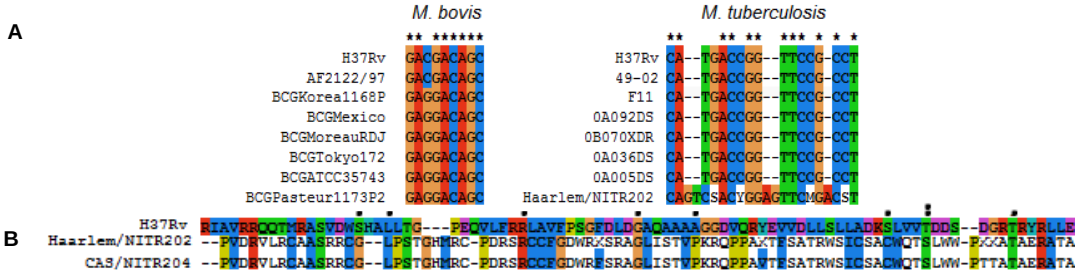


Fig. 3.4 – (A) Alignment of *Rv2488c* nucleotide sequences from Mb and Mtb strains, depicting two separate regions with specific alterations (1603-1611 bp). **(B)** Translated Mtb amino acid sequences of H37Rv, Haarlem/NITR202 and CAS/NITR204 genomes, showing a very different protein sequence in these clinical strains (residues 480 to 570). Symbols: * (equal nucleotides), **.** (semi-conservative), **:** (conservative), blank space (non-conservative), - (gaps).

a confidence level of over 80% for both MTC species (Mtb BT1, Mtb BT2, Mtb CCDC5180, Mtb HKBS1, Mtb Beijing-NITR203, Mtb CCDC5079, Mtb 96075 – 91%; BCG Korea 1168P, BCG ATCC 35743, BCG Tokyo 172, BCG Mexico, BCG Moreau RDJ – 85%). All models were identical among Mtb and Mb strains. The SNP c.797C>T in Mtb created a non-conservative alteration, leading to an amino acid transition, from a threonine to an isoleucine which harbors a hydrophobic side chain. SNP c.1605C>G in Mb leads to an amino acid substitution from charged aspartic acid to also charged glutamic acid. Both SNPs create amino acid alterations in the NB_ARC domain and alter protein structure from a closed structure to a more open state, where LuxR and Cya3 domains are distally located in comparison with the predicted models obtained for reference strain (Fig. 3.5). Publicly available nucleotide sequences were translated into amino acid sequences to determine SNP/INDEL effects, such as frameshifts and/or stop codons. Considering Mtb strains, translated nucleotide sequences resulted in several apparent disruptions of protein sequences (as compared to H37Rv reference sequence, Fig. 3.4B). Focusing on the central DNA-binding domain (with more than 50% of detected alterations considering nucleotide analysis), strains Mtb NITR202 and Mtb NITR204 had a very different amino acid sequence compared to H37Rv, suggesting functionality

impairment. Mtb RGTB327 also displayed the same disruptive pattern, equal to findings of reference [65], a study that was published during our already ongoing analyses, in which the authors suggest a pattern of pseudogenization of *Rv2488c* (and other Lux-R genes) and loss of functionality due to frameshift-causing INDELs. Considering protein sequence alignment retrieved from NCBI, we found 41 amino acid substitutions: 14 conservative (34.1%), 22 non-conservative (53.7%) and five semi-conservative (12.2%) that denote similar properties. Substitutions 216_G>W (sc), 265_T>I (nc), 292_W>G (nc) and 1051_Y>C (nc) were common among 5.1%, 5.1%, 12.8% and 5.1% strains respectively, while remaining alterations were orphan. After clustering of alterations, Fig. 3.6 was constructed rationalizing total alterations in

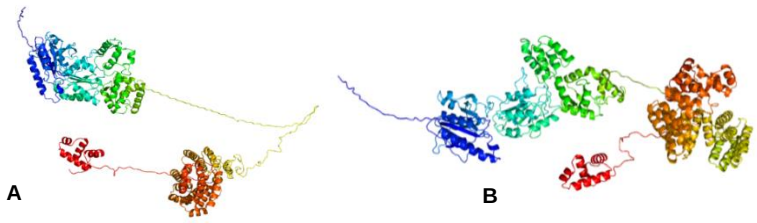


Fig. 3.5 - Predicted 3D models of Rv2488c for *M. tuberculosis HKBS1* harboring SNP c.797C>T (**A** - Model dimensions (Å) X: 121.420 Y: 115.848 Z: 220.213) and *M. bovis* BCG Tokyo 172 harboring SNP c.1605C>G (**B** - Model dimensions (Å) X: 131.881 Y: 149.162 Z: 309.778*).* Color-code: Cyclase domain (green/blue), NB_ARC (yellow/brown), LuxR-like (red).

each domain per total number of alterations detected. Interestingly, for both Rv2488c nucleotide and protein sequences, NB_ARC domain was the most altered domain (50.4% at nucleotide level and 43.9% at protein level), followed by the DNA region between NB_ARC and LuxR domains (28.9% at nucleotide and 24.3% at protein levels), which suggest selective recognition of specific DNA sequences, possibly under specific environmental conditions or host niches. The cyclase domain alterations in protein sequences accounted for almost 22%, while in terms of nucleotide sequences it accounted for slightly half (10%).

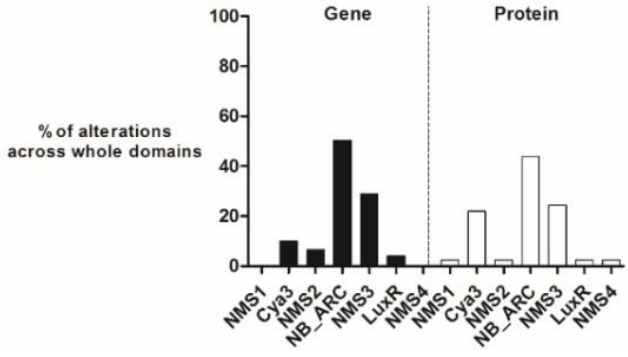


Fig. 3.6 - Percentage of *Rv2488c* gene and protein alterations per domain, in relation to reference H37Rv genome. NMS refers to alterations in the nucleotide sequences in-between domains, ordered from 1 to 4 from N to C-terminal ends.

3.1.4 Gaining insights into the microevolutionary pattern of *RV2488c*

To gain insights into the microevolutionary pattern of *Rv2488c*, two neighbor-joining cladograms were generated using previous nucleotide sequence alignments produced in

ClustalX and edited in FigTree, discriminating a polar tree layout, with midpoint root and aligned labels (Fig. 3.7). SNPs and/or INDELs were represented in each branch or node by color. Concerning Mb, *Rv2488c* (*Mb2515c*) nucleotide sequence is highly conserved among Mb strains, in agreement with results from multiple sequence alignments. In contrast, Mtb reveals higher degree of variation across Mtb strains, with several alterations detected. These results suggest selective pressure to conserve *Rv2488c* among Mb strains and possibly an *in vivo* protein role. On the other hand, Mtb cladogram depicting dispersed SNPS and INDELs across branches suggests differential microevolution and adaptation pattern to host and thus one cannot discard the possibility of a crucial *Rv2488c* role in the human bacilli pathogenesis. Altogether, these findings suggest different microevolutionary histories of *Rv2488c* when comparing Mtb and Mb strains.

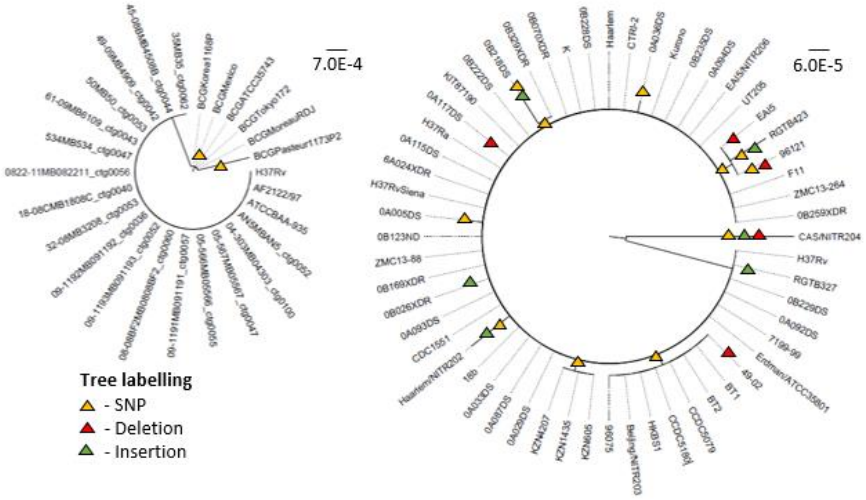


Fig. 3.7 - Neighbor joining-based cladograms depicting distribution of genomic alteration events. Cladograms constructed using ClustalX 2.1 and FigTree v1.4.2. Alterations, namely SNP, insertions and deletions across Mb and Mtb strains were color-mapped.

3.2 Mycobacterial physiology under host-like conditions

3.2.1 *M. bovis* II growth response after exposure to RIF or acidic environment

In vitro stress conditions as acid or antibiotic presence, and oxidative stress, may mimic host environment and were selected to assess physiological and transcriptional response of mycobacterial isolates. Conditions screened in this work were also based on microarray data and gene expression profiles published on TB Database that depict higher or lower expression of *Rv2488c* under different growth conditions. Mg and Fe depletion in growth media were selected due to the importance of these cofactors in mycobacteria physiology and possible roles in host infection-pathway. To study mycobacterial susceptibility to RIF, several concentrations, ranging from 0.05 to 100 µg/ml for Msm and 0.5 to 50 µg/ml for Mb, were tested in petri dishes to determine the sub-inhibitory concentration. Based on experimental

data (Fig. 3.8A), 50 µg/ml was chosen for Msm and 25 µg/ml for all Mb strains. Growth of Mb II under pH 5.5 and in the presence of RIF 25 µg/ml was followed during two weeks. As expected, exposure to acidic medium using sodium pyruvate as carbon source heavily affected Mb growth and diminished CFUs in five orders of magnitude ($p<0.0001$) to approximately 0.00127% of survival percentage compared to control (Fig. 3.8B). The amount of final biomass in suspension (as indicated by OD) significantly decreased in comparison with control (pH approximately 6.8), indicating that growth was impaired under acidic conditions. The specific growth rate (µmax) at pH 5.5 was considerably lower (reduction by more than 2-fold) than that of control, indicating slower bacterial metabolism and low acid tolerance. In agreement with these results, previous observations of Mtb growth at pH 5.5 or lower have described almost no duplication and/or low tolerance to acid shifts in medium[25], although depending on initial bacterial density and carbon source, greater tolerances could be observed. When Tyloxapol (a detergent) was used instead of Tween80, killing was reduced but not eliminated for inocula densities of $1x10^6$ to $2.5x10^8$ CFU. In our case, we did not use any detergent, and initial inoculum had OD_{600} of 0.1 (approximately $2.5x10^7$ CFU). Growth under the presence of RIF 25 µg/ml (ns) also led to reduction of viable cell count (50% survival percentage) and the specific

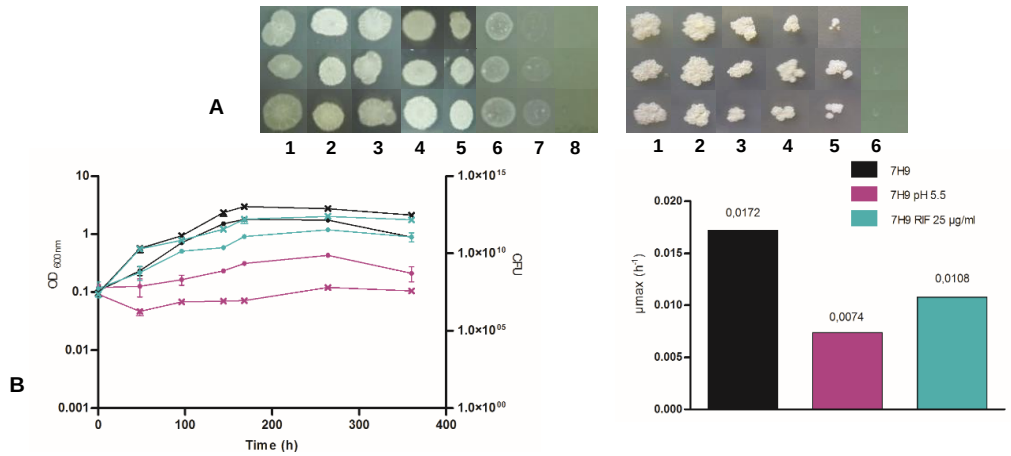


Fig. 3.8 – (A) Rifampicin sub-MIC determination for *M. smegmatis* mc^2 155 (left panel) and *M. bovis* BCG Tokyo (right panel). Each column (1-8/1-6) represents specific RIF concentrations. Left panel: control (1), 0.05 µg/ml (2), 0.1 µg/ml (3), 0.5 µg/ml (4), 1 µg/ml (5), 10 µg/ml (6), 50 µg/ml (7) and 100 µg/ml (8). Right panel: control (1), 0.5 µg/ml (2), 5 µg/ml (3), 10 µg/ml (4), 25 µg/ml (5) and 50 µg/ml (6). **(B)** Growth curve and growth rate (µmax) *of M. bovis* II in modified 7H9 (pH 5.5 and RIF 25 µg/ml). Symbols indicate CFU (x) and OD_{600nm} (o). Middlebrook 7H9 (100 ml) was inoculated with initial OD_{600nm} of 0.1 and incubated two weeks at 37°C without agitation. Data displayed as means of values with respective standard deviation of two independent experiments (in some cases, error bars are too small to visualize), analyzed with Friedman's test (p-value<0.05) with a Dunn's Multiple Comparison post-test. For CFU, extremely significant data ($p<0.0001$, ***); for OD, non-significant data (ns) with a significant comparison ($p<0.05$, *) between pH 5.5 and 7H9.

growth rate as compared to control, although the effect was less pronounced when considering the pH effect. Saturation point was however similar to control. The physiological behavior of Mb under these conditions may reflect (i) decreased susceptibility of this Mb II strain to RIF, being less efficient on blocking RNA synthesis near RNA polymerase active center and hence

transcription inhibition or (ii) the short lifespan of RIF at temperatures of 37°C, which may decrease antibiotic activity during incubation and thus enable pressure relief as time progressed.

3.2.2 *In vitro* biofilm formation under host-like conditions

Planktonic and attached growth of selected mycobacteria under challenging environmental conditions was assessed in 12 well-microtitre plates, after three weeks of incubation for Mb strains and one week for Msm, using modified 7H9 medium according to each specific condition tested. Survival percentages and OD_{590} of crystal violet-stained attached cells were plotted to disclose and compare mycobacterial physiological responses to each stress condition (Fig. 3.9). Mycobacteria formed reticulated pellicles at the air-medium interface of microtitre plates under different conditions and cells adherent to the periphery of each well were visible and quantified after removal of bacterial suspensions and staining with crystal violet. Growth under Fe starvation was compared to growth in 7H9 as control (1.03 to 1.15 mM Fe). For Msm (saprophyte) progressively higher survival percentages were achieved as Fe concentration incremented from 2 µM to control conditions. Under these conditions, concentration of stained attached cells did not suffer expressive variation in relation to control cells, except for an Fe concentration of 10 µM, where biofilm formation was slightly higher, only significant in comparison to residual Fe concentration (Fig 3.9A). BCG Tokyo, a virulence attenuated Mb strain, demonstrated the same behaviour although the gross values of OD_{590} were lower (Fig. 3.9A). Overall, Fe availability in growth media seems to influence both the pattern of growth and the amount of biofilm formed. Residual levels of Fe were previously described as sufficient for biofilm formation in Msm, although not affecting planktonic growth. Different levels of OD_{590}, corresponding to differences in the quantity of attached Mb and Msm cells, suggest differences in the ability of these two species to form biofilms under limiting Fe conditions, since Msm is (i) a saprophyte strain, able to sustain higher amplitudes of inhospitable conditions given the diversity of its ecological niches; (ii) Fe uptake in Msm is based on exochelins siderophores, contrasting with Mb (as with Mtb) which use mycobactins instead (since exochelin-coding genes are absent)[73]. Effect of acid exposure on cell attachment was also tested (Fig. 3.9F), choosing pH values of 5.5, 5.0, and 4.5 (encountered by mycobacteria-containing phagosomes, although pH 5.0 and 4.5 are specific to post-phagosomal acidification). In these studies, we also included two Mb field strains. As expected, Msm and BCG Tokyo were highly susceptible to pH 4.5 and 5.0, as were the field strains. BCG Tokyo evidenced slightly higher survival percentage at pH 5.5 (when compared with virulent strains). In addition, the amount of BCG Tokyo attached cells was higher when compared with Mb I and II, although the absolute values of OD_{590} are low for all three strains, since survival percentage was also affected under these conditions. Considering these results, we chose to

expose bacteria to oxidative stress in combination with acidic stress, specifically pH 5.5, in which growth was less affected (Fig. 3.9B). Combination of both stress conditions affected Msm growth as compared to control, however the exposure to a combination of acidic and oxidative stress in simultaneous resulted in severe survival reduction, from approximately 85% to almost 10%. Unexpectedly, the amount of attached cells across these conditions remained fairly unaltered, with minor non-statistically significant differences. In contrast, BCG Tokyo growth was significantly impaired under oxidative or acidic stress, with survival percentage nearing zero. Biofilm levels also decreased significantly with joint acidic and oxidative stresses.

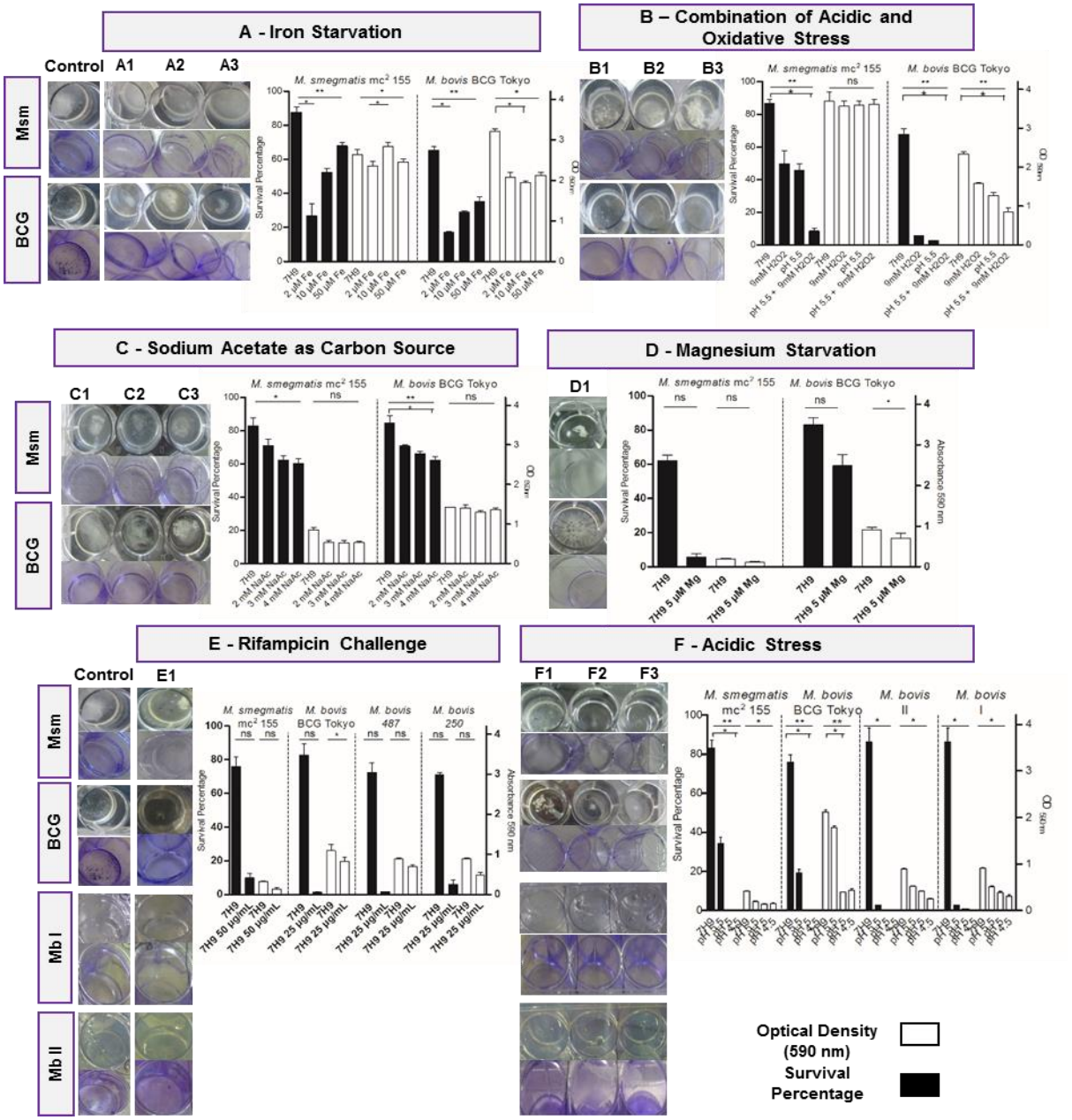
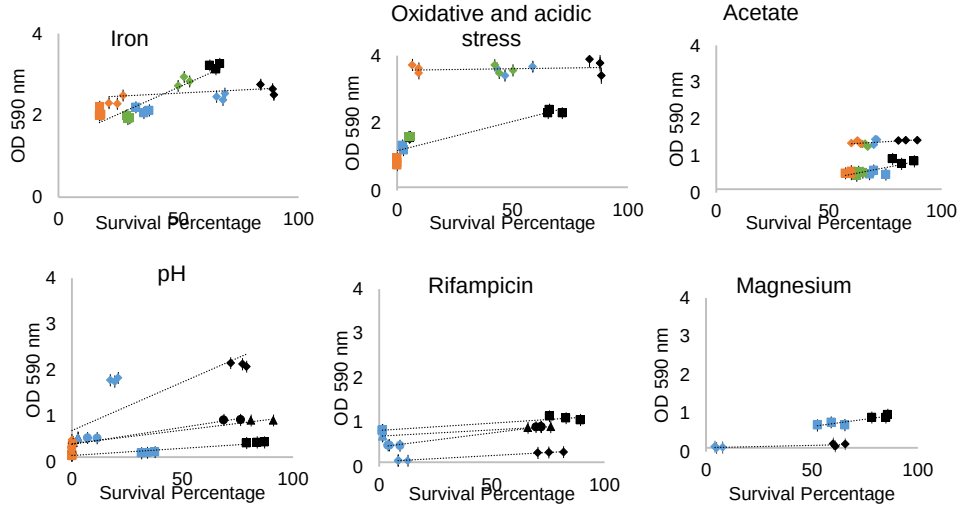


Fig. 3.9 - Quantification of biofilm formed and survival percentage in *M. smegmatis* mc^2 155 (after one week), *M.* bovis BCG Tokyo and *M. bovis* I and II (after three weeks) in different conditions: iron starvation (A1-2 µM, A2-10 µM, A3-50 µM), acidic and oxidative stress (B1-pH 5.5, B2-9 mM H_2O_2, B3- pH 5.5 + 9 mM H_2O_2), sodium acetate as carbon source (C1-2 mM, C2-3mM, C3-4 Mm), magnesium starvation (D1-5 µM), rifampicin challenge (E1-25 µg/ml for Mb strains and 50 µg/ml for Msm) and acidic stress (F1-pH 5.5, F2-pH 5.0, F3-pH 4.5). Strains were grown in 12-well microtitre plates with final volume of 2 mL and incubated at 37 °C.

Data displayed as comparison between means of at least 3 independent experiments, considering p-value≤0.05 in Friedman's test (upper statistics) with a Dunn's Multiple Comparison post-test (lower statistics) to compare all conditions in pairs, for multiple conditions versus control); and considering p-value≤0.05 Wilcoxon matched pairs' test, for single condition versus control; ns – nonsignificant, * (p-value≤0.05), ** (p-value≤0.01).

Previous studies suggest a connection between acid tolerance and increase in Mg availability for Mtb[27], suggesting a role for Mg in bacterial physiology, not only in normal but also under challenging conditions. Based on these studies, we chose to analyse the effect of Mg starvation

A



B

Condition	Iron		Oxidative and pH		Acetate		Magnesium		Rifampicin				pH			
Strain	Msm	BCG	Msm	BCG	Msm	BCG	Msm	BCG	Msm	BCG	Mb i	Mb II	Msm	BCG	Mb i	Mb II
R^2	0,8377	0,1207	0,8074	0,0321	0,6037	0,2418	0,8395	0,6664	0,9272	0,9927	0,9961	0,9995	0,9385	0,7229	0,9497	0,8517

Fig. 3.10 – (A) Correlation between survival percentage and optical density at 590 nm (indicating crystal-violet-stained adherent cells) of different mycobacterial strains under host-like conditions. Symbols: ◇ - Mb BCG; □ - Msm; ○ - Mb I; Δ – Mb II. Some error bars are too small to visualize. Color-code: black – control; blue – pH 5.5, 50 µM iron, 5 µM Mg, 2 mM acetate and 25/50 µg/ml; green – 9 mM H_2O_2, 10 µM iron, 3 mM acetate, pH 5; orange – pH 5.5 and 9 mM H_2O_2, 2 µM iron, 4 mM acetate and pH 4.5. **(B)** Coefficient of determination of each trend line.

(Fig. 3.9D). Decrease of Mg to residual concentration of 5 µM significantly affected Msm survival percentage, resulting in a 56% decrease compared with control level. The amount of biofilm formed was similar to reference medium, although the presence of a cord-like structure within broth was visible and the typical reticulated pellicle at the air-medium interface was not observed. BCG Tokyo retained some growth capability as seen by the slight decrease of survival percentage. In terms of OD_{590} levels, these were slightly below the control level and a clear reticulated pellicle was absent, also depicting a cord-like structure. These cord-like structures were absent in control condition, suggesting that Mg starvation stimulates cell aggregation and thus may alter the physiology of mycobacteria. We also tested the influence of a different carbon source, specifically three concentrations of acetate (Fig. 3.9C), since it feeds central metabolism at the glycolysis-TCA cycle intersection generating acetyl-coenzyme[66]. For Msm, growth with sodium acetate as carbon source led to a slight decrease

of survival percentage (as compared to glycerol as carbon source) that was more evident with higher concentrations. Regarding the amount of attached cells, no differences were registered across the tested acetate concentrations, although they were slightly below the control levels (7H9). Considering BCG Tokyo, biofilm formation is similar to control levels in all acetate concentrations. Survival percentage also decreased moderately with increase of acetate concentration (p=0.0017). Considering the effect of rifampicin challenge on attached growth, survival percentage under RIF exposure decreased significantly for the three Mb strains (Fig. 3.9E), although biofilm formation was similar to control conditions, which, considering the A_{590}/survival ratio, suggests that exposure to RIF may stimulate the biofilm response as a mean to cope with imposed stress. Correlation trend between survival percentage and cell attachment to microplate surface (related to the amount of crystal-violet stained biofilm) under stress was also established (Fig. 3.10). Analyzing the covariation of survival percentage and A_{590} we can see that, in general, the relationship between both variables was linear and the higher the survival percentage, the higher the biofilm; we obtained high coefficients of determination (R^2) for every strain and condition tested, except for BCG growth under Fe starvation, presence of acetate and a combination of oxidative and acidic stress. In these cases, R^2 was 0.120, 0.032 and 0.241, respectively, indicating that applied regression analysis is unable to explain with high confidence the covariation of survival percentage and A590, being data distribution nonlinear and with possible outliers.

3.3 Transcriptomic analysis of *Rv2488c*

In order to assess whether the expression level of *Rv2488c* may be affected after mycobacterial growth under host-like conditions, we performed an RT-qPCR assay to indirectly

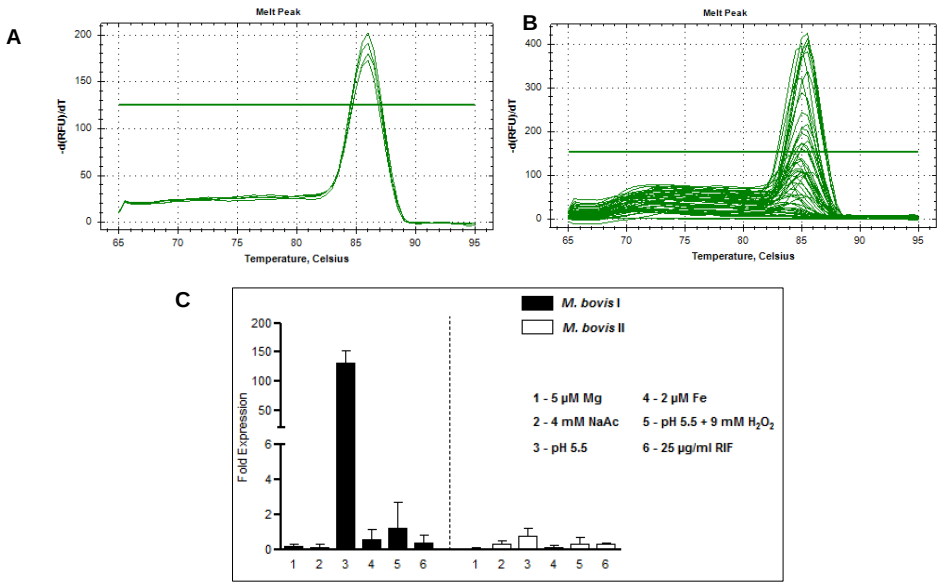


Fig. 3.11 – (A) Melting peaks for the qPCR optimization reaction performed for *16S* reference gene. **(B)** Melting peaks for reverse transcriptase quantitative PCR performed for *Rv2488c* under studied conditions. **(C)** Expression of *Rv2488c* under studied conditions measured by RT-qPCR. The fold-expression indicates the ratio of *16S rRNA* -normalized gene expression levels in *M.*

bovis I and II cultures exposed to stress as compared to growth in 7H9 medium. Data represented as average of three independent experiments and the mean values with standard deviations, analyzed with a non-parametric ANOVA (Kruskal-Wallis test, α=0.05).

measure the mRNA levels of *Rv2488c* in Mb field strains I and II grown in 7H9 or in modified 7H9. Optimization of the reactions was first performed using conventional PCR and qPCR (Chapter 2.4.2), followed by RT-qPCR assay in a single step (Fig. 3.11B). The melting temperature of amplified *16S* rRNA gene and *Rv2488c* corresponded to 85ºC and 84.5ºC respectively. The expression level of *Rv2488c* under each stress condition was obtained after normalization with the expression level of reference gene *16S* rRNA gene, which was also determined simultaneously for each control and stress condition. Expression levels of *Rv2488c* were higher (in relation to control) for Mb I after growth in acidified 7H9 (pH 5.0) (131.1±21.1-fold induction) and combined presence of acidic and oxidative stress1.24±1.46- fold induction) (Fig. 3.11C). Growth under Fe starvation, RIF challenge, Mg starvation or acetate as sole carbon source did not affect *Rv2488c* expression (corresponding to, 0.57±0.55, 0.35±0.49, 0.15±0.16 and 0.14±0.19 fold changes, respectively). In the case of Mb II, *Rv2488c* expression levels remained fairly unaltered.

3.4 Construction of Mb *Rv2488c* knock-out strain by phage-mutagenesis

Construction of a knock-out *Rv2488c* mutant through phage-mediated mutagenesis was initiated in this work but was not fully completed in due time. The strategy underneath was based on a specialized transduction protocol that allows directed mutagenesis using an AES (Allelic Exchange Substrate) to disrupt target GOI (Gene of Interest) and create isogenic strains[67]. This work conceptually implied the preparation of an Rv2488c::hyg allele. For complementation, a copy of wild-type *Rv2488c* under its own promoter would be reintroduced into the respective mutant. This has been reported as an efficient method for generating marked and unmarked targeted gene disruptions in MTC. The protocol involved cloning *Rv2488c* into cosmid pYUB854 and then into the conditionally replicating (temperature-sensitive) shuttle vector phAE159 to generate a specialized transducing mycobacteriophage. The temperature sensitive mutations in the mycobacteriophage genome permit replication at the permissive temperature of 30ºC but prevent replication at the nonpermissive temperature of 37 ºC. Transduction at a non-permissive temperature supposedly results in highly efficient delivery of the recombination substrate to cells in the recipient population. The deletion mutations in the targeted genes are marked with antibiotic-resistance genes that are flanked by resolvase recognition sites. Transductants with the AES can be selected on antibiotic-containing media. Using a plasmid-encoding *tnpR* (revolvase), resistance genes could be removed, generating unmarked mutants. In this work, we were able to clone the AES (upstream and downstream regions of *Rv2488c*) into the pYUB854 vector (Fig. 3.13A and B).

After successful confirmation of pYUB854_Rv2488c construction, we attempted to ligate cosmid with phasmid phAE159 (46.95 Kb), in order to obtain shuttle vector for electroporation into *E. coli* HB101 and Msm mc^2 155. After ligation reaction and attempted transformation of recipient strain, isolated colonies were grown in supplemented medium and pDNA was extracted and digested with *Pac*I to confirm hypothetic transformation. No transformants carrying the expected 5.6 Kb DNA band (corresponding to pYUB854_Rv2488c) were obtained that would confirm successful cosmid and phasmid ligation. A single DNA band (over 10 Kb, putatively corresponding to phAE159) was observed. Efforts to complete this genetic construction will be continued in the future.

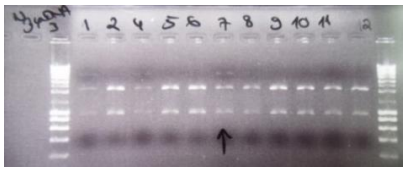
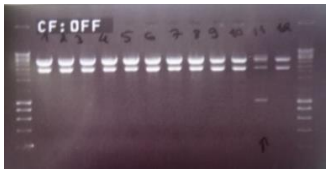
A  **B** 

Fig. 3.12 – Confirmation of *E. coli* αDH5 transformation with constructed pYUB854. **(A)** pYUB854 with 5' end of *Rv2488c* (3.8 Kb plus 877 bp). **(B)** Confirmation of constructed pYUB854_Rv2488c.

CHAPTER IV – CONCLUDING REMARKS AND PERSPECTIVES

Mtb virulence is linked to its ability to survive inside macrophages. During infection, lung macrophages engulf bacillus and provide an intracellular niche for bacteria replication. Mtb and macrophage interaction is a critical and complicated process, key feature in the arms race between pathogen and host. LuxR-family proteins have also been associated with virulent Actinobacteria members[44]. Since LuxR proteins and biofilm formation are related to QS in other pathogenic organisms, we sought to explore a hypothetic connection between these, considering the absence of information on biofilm and QS processes in pathogenic mycobacteria. Previous studies[46,47] hypothesized *Rv2488c* as a virulence-associated gene although experimental evidence is still very scarce. *Rv2488c* presents a novel protein domain organization, showing a CHD+NB_ARC+LuxR architecture. From a survey on actinobacterial LuxR genes, CHD associated with LuxR domain is restricted to *Mycobacterium* spp. and specifically found in pathogenic species[44]. Our *in silico* analysis of *Rv2488c* nucleotide and amino acid sequences across MTC bacteria disclosed several polymorphisms in the NB_ARC domain. This variability suggests evolutionary pressure on this region to modulate DNA-binding activity and specificity, possibly in response to different input signals, which is also in agreement with flexible CHD activity, another domain in which we also found a few polymorphisms[61]. Despite these intriguing clues, it is still unknown how Rv2488c protein generates an in vitro and *in vivo* response to environmental stimuli, since activity as an adenylate cyclase is predicted as inactive when considering amino acid residue alteration in relation to other cyclase genes[62]. According to our modelling effort, the SNPs c.797C>T and

c.1605C>G that are present in several clinical strains of Mtb and Mb, respectively, apparently lead to major modifications of protein structure and possibly impose functional consequences. A recent regression analysis-based study overlapping ours suggests that the striking genomic variation of *Rv2488c* (and other related CDH+LuxR LuxR genes) across clinical MTC strains indicate pseudogenization. The same authors found a statistically significant association of *Rv2488c* frameshift-causing INDELs with the establishment of pulmonary tuberculosis and progression to extrapulmonary disease[65]. Although clustered alterations depicted high variation in DNA-binding domain, phylogenetic insights through our multifaceted bioinformatic analyses suggest a differential *Rv2488c* microevolutionary history between Mb and Mtb strains, with almost complete conservation of *Rv2488c* nucleotide sequence across analyzed Mb strains; in contrast, Mtb clinical strains show remarkable variation among each other and in relation to reference H37Rv with SNPs and INDELs scattered across genomes in an apparently disperse pattern. On one hand, these findings suggest a selection scenario of *Rv2488c* among Mb strains, possibly indicating an essential function of this genomic region in Mb ecology. However, the underlying variation across Mtb suggest two antagonist scenarios along with Mtb coevolution history with human population: (i) a differential microevolution pattern and adaptation to human host by accumulation/selection of a higher degree of non-synonymous mutations and/or INDELs at this genome locus which may result in improved functionality and, thus, one cannot discard the possibility of a crucial role of *Rv2488c* in Mtb pathogenesis inside the human host; (ii) accumulation of random mutations and pseudogenization leading to progressive or abrupt loss function of this ORF. Consistently, Mtb genomes putatively encode other LuxR-like+CHD genes. Bioinformatics analysis of the upstream region of the transcription unit enclosing *Rv2487c* and *Rv2488c* from seven Mb and 56 Mtb genome sequences, encompassing 500 bp before the ORF of *Rv2487c*, was also carried out to detect potential polymorphisms that could affect operon expression upon different stimuli. Alignment of Mb and Mtb sequences (H37Rv as reference) also evidenced remarkable nucleotide variation within the hypothetic promoter region, which suggests that this region is highly unstable and that its notorious variability across MTC may arise from selection of environmental and molecular cues that exert control at several levels. By elucidating the physiological response and molecular basis of gene regulation of MTC bacteria under diverse stimuli, we hope to shed light on the function of several genes, including *Rv2488c*. If normal phagosome-lysosome fusion process inside the host is successful upon infection, bacilli are exposed to an environment that includes low pH, reactive oxygen intermediates (ROI), and toxic peptides produced after phagosomal maturation[20]. In this context, *in vitro* stress conditions as acid or antibiotic presence and sodium acetate may mimic host environment and thus were selected to assess the physiological and transcriptional response of selected mycobacterial isolates. In terms of acidic stress, Msm displayed more sustainability in pH 5.5,

concurring with existence of saprophyte mycobacteria in acidic soil and water[25]. Mb strains responded poorly to pH 5.5 and to lower pH values, with severely impaired growth. This response was expected since these pH values are common in the phagosome of IFN-γ immunological activated macrophages[20]. During *in* vivo growth, mycobacteria exhibit exclusion of vacuolar ATPase and phagosome-lysosome fusion to evade lethal acid conditions, while some non-pathogenic mycobacteria produce ammonia to raise vacuolar pH[68]. Batch growth curves in polystyrene flasks for Mb II showed differences compared to microtitre growth, although in this case growth was also impaired, as expected. This result may be due to growth conditions, since flasks allow greater culture aeration and harbour greater culture volume, leading to lower dehydration under extended incubation periods. Addition of oxidative stress to acidic medium imposed an additional threat to physiological homeostasis, so that mycobacteria deploy redox-sensing mechanisms, such as the *DosR/DosS/DosT* two-component system and *whiB3*, the two regulons related to Mtb redoxome[40]. Not surprisingly, addition of oxidative stress to the already acidic broth further impaired cell growth, as evidenced by low survival percentages. Key nutrients may dictate infection outcome upon pathogen establishment in host environment. Mg is an abundant and essential element in all organisms, essential in many fundamental cellular activities. For instance, many ATP-dependent enzymatic reactions in the cell require bound Mg[69], also stabilizing cell membranes by cross-linking carboxylated/phosphorylated lipids with cations[70]. Mg role as an enzymatic cofactor is acknowledged, although role as an adjuvant to cope with acidic conditions has been suggested, based on increased growth of Mtb at pH 6 when Mg availability is higher[27]. It is also possible that the bacilli encounter a sudden Mg depletion when moving from the extracellular environment into the phagosome. While Mg concentration in oral and lung mucus has been shown as over 0.5 mM, the concentration in macrophages may be lower[71,72]. One study pointed out that the concentration in epithelial cell phagosomes containing *Salmonella* is 10-50 Mm[73], although these values have not been confirmed for Mtb. Intracellular concentration of Mg must be controlled by dedicated molecular mechanisms to achieve Mg homeostasis, which has still been largely unexplored in Mtb. Interestingly, although in our study Mb growth was impaired under Mg starvation, the formation of cord-like structures in microtitre plates was stimulated, suggesting that Mg starvation promotes cell aggregation. On the other hand, Fe has been indicated as a requirement for biofilm formation in Msm, even at residual levels of 2 μM, and authors suggested possible role in biofilm formation of Mtb[74]. *In vivo* mycobacteria species capture Fe by means of siderophore expression (exochelins and mycobactins), chelators with high affinity for Fe, whose expression is controlled by extracellular Fe abundance[75]. Pathogenic mycobacteria do not express exochelins as saprophyte strains, rather expressing two mycobactins (one secreted and the other one cell-associated)[75]. In our studies, Fe availability in growth media seems to influence both the pattern of growth and the amount of

biofilm formed. Observations between Msm and BCG Tokyo are in agreement with reviewed literature. Mycobacteria exposure to acetate carbon source (4 mM) led to lower survival percentages of both Msm and BCG strains. Acetate is a substrate for acetyl-CoA production, feeding tricarboxylic acid cycle and generating intermediary metabolites to generate ATP. It has been shown that specific carbon sources alter growth regiment at acidic conditions for Mtb[76], so we can infer that successful host colonization is carbon-source dependent. Metabolic profiling in guinea pigs' lungs has revealed presence of several metabolites, one of which is acetate[77]. Still, one might consider usage of several carbon sources, including lipids, by bacilli depending on environmental challenges faced at specific stages of infection. RIF-impaired growth, imposing transcription blockage, in all Mb strains and Msm was confirmed and expected, although higher OD suggests that biofilm formation may be as a survival strategy employed by mycobacteria against stress. Transcriptional profiling of *Rv2488c* revealed differential gene expression responses to different growth conditions. Highest fold induction was obtained for pH 5.5 growth of Mb I (131 times higher than control condition) and joint acidified medium with oxidative stress resulted in opposite findings to those in TB Database microarray profiles with lower fold-induction for acidified medium. These may be explained by protocol followed, since cells were grown in acidified medium from the start and weren't acid-treated during a fixed amount of time. Transcriptional profiles were different among clinical strains, suggesting that *Rv2488c* is transcribed in Mb but prone to variation among strains and its regulation appears to be dependent on environmental cues. Recently, RNA-sequencing has emerged as a tool for analysing transcriptomes, being able to distinguish between primary and processed transcripts (differential RNA-sequencing)[78]. Non-coding transcripts (ncRNA) play a role in regulation of gene expression, allowing bacteria to adjust their metabolism and express genes necessary for subversion of host's defences[79]. It would be interesting to refine the transcription profile of *Rv2488c* by RNAseq and to study the non-coding transcriptome under tested growth conditions. Concerning future perspectives, we envisage completion of *Rv2488c* knock-out mutant construction since gene inactivation is a strategic tool for functional studies and allows discovery of key genes in infection onset/establishment and/or reactivation. *In vitro* macrophage infection comparing wild-type and isogenic knock-out mutant would also allow ascertainment of this gene's role to bacilli survival in host. Heterologous expression and purification of Rv2488c protein, to test CHD and DNA-binding activities under different conditions *in vitro*, could also be carried out to give further insights into its function and to ascertain its contribution to the chronic phenotype, a hallmark of TB.

CHAPTER V – BIBLIOGRAPHY

1 - Michel *et al.*, 2010. *Mycobacterium bovis* at the animal–human interface: A problem, or not? Veterinary Microbiology 140: 371-381

2 - WHO, 2014. Global Tuberculosis Report

3 - Fogel, 2015. Tuberculosis: A disease without boundaries. Tuberculosis 95(5): 527-531

4 - Müller *et al.*, 2013. Zoonotic *Mycobacterium bovis*–induced Tuberculosis in Humans. Emerging Infectious Diseases 19(6). DOI: http://dx.doi.org/10.3201/eid1906.120543

5 - Smith *et al.*, 2009. Myths and misconceptions: the origin and evolution of *Mycobacterium tuberculosis*. Nature Reviews Microbiology 7(7):537-44

6 - Saviola & Bishai, 2006. The Genus Mycobacterium - Medical. Prokaryotes 3: 919-933

7 - Hett & Rubin, 2008. Bacterial Growth and Cell Division: a Mycobacterial Perspective. Microbiology and Molecular Biology Reviews 72(1): 126-156

8 - Brennan, 2003. Structure, function, and biogenesis of the cell wall. Tuberculosis 83(1-3): 91-97

9 - Coscolla & Gagneux, 2014. Consequences of genomic diversity in *Mycobacterium tuberculosis*. Seminars in Immunology 26: 431-444

10 - Brosch *et al.*, 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. PNAS 99(6): 3684-3689

11 - Garnier *et al.*, 2003. The complete genome sequence of *Mycobacterium bovis*. PNAS 100(13):7877-82

12 - Smith, 2003. *Mycobacterium tuberculosis* Pathogenesis and Molecular Virulence. Clinical Microbiology Reviews 16(3): 463-496

13 - Mostowy *et al.*, 2005. Revisiting the Evolution of *Mycobacterium bovis*. Journal of Bacteriology 187(18): 6386–6395

14 - Blouin *et al.*, 2012. Significance of the Identification in the Horn of Africa of an Exceptionally Deep Branching *Mycobacterium tuberculosis* Clade. PLoS ONE 7(12): e52841. doi: 10.1371/journal.pone.0052841

15 - Mortimer & Pepperell, 2014. Genomic Signatures of Distributive Conjugal Transfer among Mycobacteria. Genome Biol. Evol. 6(9):2489–2500. doi:10.1093/gbe/evu175

16 - Derbyshire & Gray, 2014. Distributive Conjugal Transfer: New Insights into Horizontal Gene Transfer and Genetic Exchange in Mycobacteria. Microbiol Spectrum 2(1):MGM2-0022- 2013. doi:10.1128/microbiolspec.MGM2-0022-2013.

17 - Wirth *et al.*, 2008. Origin, Spread and Demography of the *Mycobacterium tuberculosis* Complex. PLoS Pathogens 4(9): e1000160

18 - Nicas *et al.*, 2005. Toward Understanding the Risk of Secondary Airborne Infection: Emission of Respirable Pathogens. Journal of Occupational and Environmental Hygiene 2(3):143-54

19 - Sia *et al.*, 2015. Innate Immune Defenses in Human Tuberculosis: An Overview of the Interactions between Mycobacterium tuberculosis and Innate Immune Cells. Journal of Immunology Research 2015, Article ID 747543 (12 pages)

20 - Vergne *et al.*, 2004. Cell Biology of *Mycobacterium tuberculosis* phagosome. Annual Review Cell Developmental Biology 20: 367-394

21 - Huynh *et al.*, 2011. A delicate dance: host response to mycobacteria. Current Opinion in Immunology 23(4): 464-472

22 - Crevel *et al.*, 2002. Innate Immunity to tuberculosis. Clinical Microbiology Reviews 15(2): 294-309

23 - Neyrolles *et al.*, 2015. Mycobacteria, metals, and the Macrophage. Immunological Reviews 264: 249-263

24 - Boradia *et al.*, 2014. *Mycobacterium tuberculosis* acquires iron by cell-surface sequestration and internalization of human holo-transferrin. Nature Communications 5: 4730

25 - Vandal *et al.*, 2009. Acid Resistance in *Mycobacterium tuberculosis*. Journal of Bacteriology, 191(15): 4714-4721

26 - Rohde *et al.*, 2007. *Mycobacterium tuberculosis* invasion of macrophages: linking bacterial gene expression to environmental cues. Cell Host Microbe 2(5):352-64

27 - Piddington *et al.*, 2000. Growth of *Mycobacterium tuberculosis* in a Defined Medium Is Very Restricted by Acid pH and Mg^{2+} Levels. Infection and Immunity 68(8): 4518-4522

28 - Nambi *et al.*, 2015. The Oxidative Stress Network of *Mycobacterium tuberculosis* Reveals Coordination between Radical Detoxification Systems. Cell Host & Microbe 17: 829-837

29 - Voskuil *et al.*, 2011. The response of *Mycobacterium tuberculosis* to reactive oxygen and nitrogen species. Frontiers in Microbiology 2:e105

30 - Koch *et al.*, 2014. The impact of drug resistance on *Mycobacterium tuberculosis* physiology: what can we learn from rifampicin? Emerging Microbes and Infections, e17 doi: 10.1038/emi.2014.17

31 - Dunne, 2002. Bacterial Adhesion: Seen Any Good Biofilms Lately? Clinical Microbiology Reviews 15(2): 155-166

32 - Rutherford & Bassler, 2012. Bacterial Quorum Sensing: Its Role in Virulence and Possibilities for Its Control. Cold Spring Harbor Perspectives in Medicine 2:a012427

33 - Branda *et al.*, 2005. Biofilms: the matrix revisited. Trends in Microbiology 13(1): 20-26

34 - Karatan & Watnick 2009. Signals, Regulatory Networks, and Materials That Build and Break Bacterial Biofilms. Microbiology and Molecular Biology Reviews 73(2): 310-347

35 - Ojha *et al.*, 2008. Growth of *Mycobacterium tuberculosis* biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. Molecular Microbiology 69(1): 164-174

36 - Høiby *et al.*, 2010. Antibiotic resistance of bacterial biofilms. International Journal of Antimicrobial Agents 35(4):322-32

37 - Islam *et al.*, 2012. Targeting drug tolerance in mycobacteria: a perspective from mycobacterial biofilms. Expert Review of Anti-infective Therapy 10(9): 1055-1066

38 - Boon & Dick, 2012. How *Mycobacterium tuberculosis* goes to sleep: the dormancy survival regulator DosR a decade later. Future Microbiology 7(4):513-8

39 - Leistikow *et al.*, 2010. The *Mycobacterium tuberculosis* DosR Regulon Assists in Metabolic Homeostasis and Enables Rapid Recovery from Nonrespiring Dormancy. Journal of Bacteriology 192(6): 1662-1670

40 - Kumar *et al.*, 2011. Redox homeostasis in mycobacteria: the key to tuberculosis control? Expert Reviews in Molecular Medicine 13: e39 (25 pages)

41 - Marsollier *et al.*, 2005. Colonization of the salivary glands of *Naucoris cimicoides* by *Mycobacterium ulcerans* requires host plasmatocytes and a macrolide toxin, mycolactone. Cell Microbiology 7(7):935-43

42 - Marsollier *et al.*, 2007. Impact of *Mycobacterium ulcerans* Biofilm on Transmissibility to Ecological Niches and Buruli Ulcer Pathogenesis. PLoS Pathog 3(5): e62 doi:10.1371/journal.ppat.0030062

43 - Chen & Xie, 2011. Role and regulation of bacterial LuxR-like regulators. Journal of Cellular Biochemistry 112: 2694-2702

44 - Santos *et al.*, 2012. A Walk into the LuxR Regulators of Actinobacteria: Phylogenomic Distribution and Functional Diversity. PLoS ONE 7(10): e46758

45 - Alonso-Hearn *et al.*, 2010. A *Mycobacterium avium* subsp. paratuberculosis LuxR regulates cell envelope and virulence. Innate Immunity 16(4):235-47

46 - Muttucumaru *et al.*, 2010. *Mycobacterium tuberculosis Rv0198c*, a putative matrix metalloprotease is involved in pathogenicity. Tuberculosis 91: 111-116

47 - Parish *et al.*, 2003. The *senX3–regX3* two-component regulatory system of *Mycobacterium tuberculosis* is required for virulence. Microbiology 149: 1423-1435

48 - Lew *et al.*, 2011. TubercuList-10 years after. Tuberculosis 91(1): 1-7

49 - Reedy *et al.*, 2009. TB database: an integrated platform for tuberculosis research. Nucleic Acids Research 37(Database issue):D499-508

50 - Turkarslan *et al.*, 2015. A comprehensive map of genome-wide gene regulation in *Mycobacterium tuberculosis*. Scientific Data 2:150010. doi: 10.1038/sdata.2015.10

51 - Szklarczyk *et al.*, 2015.STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Research 43(Database issue):D447-52

52 - Kuhn *et al.*, 2014. STITCH 4: integration of protein–chemical interactions with user data. Nucleic Acids Research 42(Database issue): D401–D407

53 - Kelley *et al.*, 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols 10(6):845-58

54 - Bardarov *et al.*, 1997. Conditionally replicating mycobacteriophages: a system for transposon delivery to *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A 94, 10961–10966.

55 - Sampson, 2010. Mycobacterial PE/PPE proteins at the host-pathogen interface. Clinical and Developmental Immunology 2011:497203

56 - Mawuenyega *et al.*, 2015. *Mycobacterium tuberculosis* functional network analysis by global subcellular protein profiling. Mol Biol Cell 16(1):396-404

57 - Kruh *et al.*, 2010. *Portrait of a pathogen: the Mycobacterium tuberculosis* proteome *in vivo.* PLoS One 5(11):e13938

58 - Sassetti *et al.*, 2003. Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol 48(1):77-84

59 - Lamichhane *et al.*, 2003. A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: application to *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A 100(12):7213-8

60 - Roberts *et al.*, 2011. Control of *CydB* and *GltA1* expression by the *SenX3 RegX3* two component regulatory system of *Mycobacterium tuberculosis*. PLoS ONE 6(6):e21090

61 - Linder & Schultz, 2003. The class III adenylyl cyclases: multi-purpose signalling modules. Cellular Signalling 15: 1081-1089

62 - Shenoy & Visweswariah, 2006. Mycobacterial adenylyl cyclases: Biochemical diversity and structural plasticity. FBS Letters 580: 3344-3352

63 - Shenoy *et al.*, 2004. A survey of nucleotide cyclases in actinobacteria: unique domain organization and expansion of the class III cyclase family in *Mycobacterium tuberculosis*. Genomics, Comparative and Functional 5: 17-38

64 - Neuwald *et al.*, 1999. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. Genome Research 9: 27-43

65 - Santos *et al.*, 2015. To Be or Not to Be a Pseudogene: A Molecular Epidemiological Approach to the mclx Genes and Its Impact in Tuberculosis. PLoS ONE 10(6): e0128983

66 - Li *et al.*, 2011. Purification and characterization of the acetyl-CoA synthetase from *Mycobacterium tuberculosis*. Acta Biochim Biophys Sin (Shanghai) 43(11):891-9

67 - Tufariello *et al.*, 2014. Enhanced Specialized Transduction Using Recombineering in *Mycobacterium tuberculosis*. mBio 5(3):e01179-14

68 - Cotter & Hill, 2003. Surviving the Acid Test: Responses of Gram-Positive Bacteria to Low pH. Microbiology and Molecular Biology Reviews 67:429-453

69 - Wester, 1987. Magnesium. The American Journal of Clinical Nutrition. 45:1305-1312

70 - Shine & Douglas, 1974. Magnesium effects on ionic exchange and mechanical function in rat ventricle. Am. J. Physiol., 227, 317-323.

71 - Effros *et al.*, 2005. Epithelial lining fluid solute concentrations in chronic obstructive lung disease patients and normal subjects. Journal Appl Physiology 99(4):1286-92

72 - Shpitzer *et al.*, 2007. A comprehensive salivary analysis for oral cancer diagnosis. Journal Cancer Research Clinical Oncology 133(9):613-7

73 - Garcia-del Portillo *et al.*, 1992. Characterization of the micro-environment of *Salmonella typhimurium*-containing vacuoles within MDCK epithelial cells. Molecular Microbiology 6(22):3289-97

74 - Ojha & Hatfull, 2007. The role of iron in *Mycobacterium smegmatis* biofilm formation: the exochelin siderophore is essential in limiting iron conditions for biofilm formation but not for planktonyc growth. Molecular Microbiology 66(2): 468-483

75 - De Voss *et al.*, 1999. Iron Acquisition and Metabolism by Mycobacteria. Journal of Bacteriology 181(15): 4443-4451

76 - Baker *et al.*, 2014. Slow growth of *Mycobacterium tuberculosis* at acidic pH is regulated by phoPR and host-associated carbon sources. Molecular Microbiology 94(1): 56-69

77 - Somashekar *et al.*, 2011. Metabolic profiling of lung granuloma in *Mycobacterium tuberculosis* infected guinea pigs: ex vivo 1H magic angle spinning NMR studies. Journal of Proteome Research 10(9):4186-95

78 - Arnvig & Young, 2012. Non-coding RNA and its potential role in Mycobacterium tuberculosis pathogenesis. RNA Biology 9(4): 427–436

79 - Gripenland *et al.*, 2010. RNAs: regulators of bacterial virulence. Nature Reviews Microbiology 8, 857-866

80 - Snapper *et al.*, 1990. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. Mol Microbiol;4(11):1911-9

CHAPTER VI - APPENDIXES

Supplementary Table 1 - List of bacterial strains obtained from NCBI used in the current work with information referring type of sequence (gene or protein), accession number and entry date. *M. africanum* (Ma), *M. bovis* (Mb), *M. tuberculosis* (Mtb).

Strain		Accession Number	Entry date
Mb 04-303 MB04303_ctg0100	Nucleotide	AVSW01000100.1	25/02/15
Mb AF2122-97	Nucleotide	BX248342.1	25/02/15
Mb ATCC BAA-935	Nucleotide	CP009449.1	25/02/15
Mb 0822-11 MB082211_ctg0056	Nucleotide	JQEW01000056.1	25/02/15
Mb 534 MB534_ctg0047	Nucleotide	JQEM01000047.1	25/02/15
Mb 61-09 MB6109_ctg0043	Nucleotide	JQEX01000043.1	25/02/15
Mb 50 MB50_ctg0053	Nucleotide	JQEU01000053.1	25/02/15
Mb 49-09 MB4909_ctg0042	Nucleotide	JQES01000042.1	25/02/15
Mb 45-08B MB4508B_ctg0044	Nucleotide	JQEP01000044.1	25/02/15
Mb 35 MB35_ctg0062	Nucleotide	JQEV01000062.1	25/02/15
Mb 32-08 MB3208_ctg0053	Nucleotide	JQER01000053.1	25/02/15
Mb 18-08C MB1808C_ctg0040	Nucleotide	JQEQ01000040.1	25/02/15
Mb 09-1193 MB091193_ctg0052	Nucleotide	JQER01000053.1	25/02/15
Mb 09-1192 MB091192_ctg0036	Nucleotide	JQEO01000036.1	25/02/15
Mb 09-1191 MB091191_ctg0057	Nucleotide	JPFP01000057.1	25/02/15
Mb 08-08BF2 MB0808BF2_ctg0060	Nucleotide	JQET01000060.1	25/02/15
Mb 05-567 MB05567_ctg0047	Nucleotide	JPFQ01000047.1	25/02/15
Mb 05-566 MB05566_ctg0055	Nucleotide	JPFR01000055.1	25/02/15
Mb BCG Tokyo 172	Nucleotide	AP010918.1	25/02/15
Mb BCG Moreau RDJ	Nucleotide	AM412059.2	25/02/15
Mb BCG Mexico	Nucleotide	CP002095.1	25/02/15
Mb BCG Korea 1168P	Nucleotide	CP003900.2	25/02/15
Mb BCG ATCC 35743	Nucleotide	CP003494.1	25/02/15
Mb BCG Pasteur 1173P2	Nucleotide	AM408590.1	25/02/15
Mb AN5 MBAN5_ctg0052	Nucleotide	AWPL01000052.1	25/02/15
Mtb H37Rv	Nucleotide	CP009480.1	25/02/15
Mtb 18b	Nucleotide	CP007299.1	25/02/15
Mtb 49-02	Nucleotide	HG813240.1	25/02/15
Mtb 7199-99	Nucleotide	HE663067.1	25/02/15
Mtb BT1	Nucleotide	CP002883.1	25/02/15
Mtb BT2	Nucleotide	CP002882.1	25/02/15
Mtb CAS-NITR204	Nucleotide	CP005386.1	25/02/15
Mtb CCDC5079	Nucleotide	CP002884.1	25/02/15
Mtb CCDC5180	Nucleotide	CP002885.1	25/02/15
Mtb CDC1551	Nucleotide	AE000516.2	25/02/15
Mtb CTRI-2	Nucleotide	CP002992.1	25/02/15
Mtb EAI5	Nucleotide	CP006578.1	25/02/15
Mtb EAI5-NITR206	Nucleotide	CP005387.1	25/02/15
Mtb F11	Nucleotide	CP000717.1	25/02/15
Mtb H37Ra	Nucleotide	CP000611.1	25/02/15
Mtb H37RvSiena	Nucleotide	CP007027.1	25/02/15
Mtb HKBS1	Nucleotide	CP002871.1	25/02/15
Mtb K	Nucleotide	CP007803.1	25/02/15
Mtb KZN 605	Nucleotide	CP001976.1	25/02/15
Mtb KZN 1435	Nucleotide	CP001658.1	25/02/15

Mtb KZN 4207	Nucleotide	CP001662.1	25/02/15
Mtb RGTB327	Nucleotide	CP003233.1	25/02/15
Mtb RGTB423	Nucleotide	CP003234.1	25/02/15
Mtb Beijing-NITR203	Nucleotide	CP005082.1	25/02/15
Mtb Erdman/ATCC 35801	Nucleotide	AP012340.1	25/02/15
Mtb Haarlem	Nucleotide	CP001664.1	25/02/15
Mtb Haarlem-NITR202	Nucleotide	CP004886.1	25/02/15
Mtb Kurono	Nucleotide	AP014573.1	25/02/15
Mtb 0A005DS	Nucleotide	CP008983.1	25/02/15
Mtb 0A029DS	Nucleotide	CP008981.1	25/02/15
Mtb 0A033DS	Nucleotide	CP008980.1	25/02/15
Mtb 0A036DS	Nucleotide	CP008979.1	25/02/15
Mtb 0A087DS	Nucleotide	CP008978.1	25/02/15
Mtb 0A092DS	Nucleotide	CP008977.1	25/02/15
Mtb 0A093DS	Nucleotide	CP008976.1	25/02/15
Mtb 0A094DS	Nucleotide	CP008975.1	25/02/15
Mtb 0A115DS	Nucleotide	CP008974.1	25/02/15
Mtb 0A117DS	Nucleotide	CP008973.1	25/02/15
Mtb 0B026XDR	Nucleotide	CP008972.1	25/02/15
Mtb 0B070XDR	Nucleotide	CP008970.1	25/02/15
Mtb 0B123ND	Nucleotide	CP008968.1	25/02/15
Mtb 0B169XDR	Nucleotide	CP008967.1	25/02/15
Mtb 0B218DS	Nucleotide	CP008966.1	25/02/15
Mtb 0B222DS	Nucleotide	CP008965.1	25/02/15
Mtb 0B228DS	Nucleotide	CP008964.1	25/02/15
Mtb 0B229DS	Nucleotide	CP008963.1	25/02/15
Mtb 0B235DS	Nucleotide	CP008962.1	25/02/15
Mtb 0B259XDR	Nucleotide	CP008961.1	25/02/15
Mtb 0B329XDR	Nucleotide	CP008960.1	25/02/15
Mtb 6A024XDR	Nucleotide	CP008959.1	25/02/15
Mtb 96075	Nucleotide	CP009426.1	25/02/15
Mtb 96121	Nucleotide	CP009427.1	25/02/15
Mtb KIT87190	Nucleotide	CP007809.1	25/02/15
Mtb ZMC13-88	Nucleotide	CP009101.1	25/02/15
Mtb ZMC13-264	Nucleotide	CP009100.1	25/02/15
Mtb UT205	Nucleotide	HE608151.1	25/02/15
Ma GM041182	Protein	YP_004724156.1	25/02/15
Ma MAL010131	Protein	KBH27038.1	25/02/15
Mb BCG Pasteur 1173P2	Protein	YP_978593.1	25/02/15
Mb Bz 31150	Protein	KAN82341.1	25/02/15
Mtb 2089HD	Protein	KAN52612.1	25/02/15
Mtb 3499MM	Protein	KAN56044.1	25/02/15
Mtb BTB05-013	Protein	KCN20610.1	25/02/15
Mtb BTB06-001	Protein	KCN49036.1	25/02/15
Mtb BTB08-001	Protein	KCO43098.1	25/02/15
Mtb BTB11-133	Protein	KCQ38774.1	25/02/15
Mtb CWCFVRF MDRTB 670	Protein	KAK26294.1	25/02/15
Mtb H37Rv	Protein	Tuberculist	25/02/15
Mtb H2581	Protein	KBF12393.1	25/02/15
Mtb KT-0011	Protein	KCF55192.1	25/02/15
Mtb KT-0048	Protein	KCG21501.1	25/02/15
Mtb KT-0084	Protein	KCG81358.1	25/02/15
Mtb KT-0085	Protein	KCG89018.1	25/02/15

Mtb KT-0109	Protein	KCH36279.1	25/02/15
Mtb KZN 1435	Protein	ACT24540.1	25/02/15
Mtb M1374	Protein	KAZ07546.1	25/02/15
Mtb M1397	Protein	KAZ36735.1	25/02/15
Mtb MAL010087	Protein	KBG42689.1	25/02/15
Mtb MAL020179	Protein	KBI49506.1	25/02/15
Mtb MAL020197	Protein	KBI94221.1	25/02/15
Mtb OFXR-6	Protein	KBJ48907.1	25/02/15
Mtb SUMu011	Protein	EFP50523.1	25/02/15
Mtb TB_RSA08	Protein	KBN67555.1	25/02/15
Mtb TKK02_0069	Protein	KCC97803.1	25/02/15
Mtb TKK03_0037	Protein	KAR48543.1	25/02/15
Mtb TKK03_0089	Protein	KAS14940.1	25/02/15
Mtb TKK03_0094	Protein	KAS26373.1	25/02/15
Mtb TKK03_0099	Protein	KAS32923.1	25/02/15
Mtb TKK05MA0044	Protein	KAU92409.1	25/02/15
Mtb TKK05SA0055	Protein	KAW10614.1	25/02/15
Mtb TKK-01-0027	Protein	KBY74087.1	25/02/15
Mtb UT0014	Protein	KBT01310.1	25/02/15
Mtb UT0119	Protein	KBT36390.1	25/02/15
Mtb Wt 21231	Protein	KAN77215.1	25/02/15
Mtb XTB13-082	Protein	KCS04290.1	25/02/15

Supplementary Table 2 - List of bacterial strains used in the current work, with relevant phenotype and reference/origin.

Strain	Relevant Phenotype	Reference or Origin
Mycobacterium smegmatis mc^2 155	Ept^+, Kan^S	[80]
Mycobacterium bovis BCG Tokyo	SB0120	INIAV
Mycobacterium bovis I	SB0121	INIAV, isolated in 2004 from a bovine
Mycobacterium bovis II	SB0121	INIAV, isolated in 2007 from a bovine in Lisbon
E. coli αDH5	endA1 hsdR17(rk^-, mk^+) supE44 thi -1 recA1 gyrA96 relA1 lac [F´ $proA^+B^+$ $lacI^q$ ZΔM15:Tn10(Tc^R)]	nzytech
E. coli HB101	F^-(gpt-proA)62, leuB6, glnV44, ara-14, galK2, lacY1, ($mcrC^-$mrr), rpsL20(Str^r), xyl-5, mtl-1, recA13, thi-1	INIAV

Supplementary Table 3 - List of primers used in the current work, with primer name, sequence, target, genome location, features (if any), hybridization temperature and reference (if any). Genome location from reference strain *Mycobacterium tuberculosis* H37Rv.

Name	Primer	Target	Genome Location (bp)	Features	Tm (°C)	Reference
Rv2488cF_RT	TTGAGGTGGCTCTACTGGTC	*Rv2488c* (136 bp at 3')	2792065	-	70	This work
Rv2488cR_RT	AGTTGAAGACGGGAGGACAG		2791949	-	70	This work
Rv2488_Xba	ATCGTCTAGACAAGCCAGCATCGACACC	962 pb at *Rv2488c* 5' region	2792761	ATCG with adaptor *Xba*I site	65.6	This work
Rv2488_Kpn	ATCGGGTACCCACCTCCCGATACCCCATAT		2791817	ATCG with adaptor *Kpn*I site	64.8	This work
Rv2488_Spe	ATCGACTAGTGTCACCCTGATCTGCACAAC	877 pb at *Rv2488c* 3' region	2794651	ATCG with adaptor *Spe*I site	64.7	This work
Rv2488_Hind	ATCGAAGCTTGCACCGTCGAGATCAAACC		2793794	ATCG with adaptor *Hind*III site	66.4	This work
rrsRT_F	AAGAAGCACCGGCCAACTAC	16S (272 bp)	1468929	-	61.66	This work
rrsRT_R	TCGCTCCTCAGCGTCAGTTA		1469181	-	62.21	This work

Supplementary Table 4 - List of plasmids used in the current work, with size, general phenotype and reference.

Plasmid	Size (Kbp)	Phenotype	Work used	Reference
pYUB854	3.893	Cosmid vector, with res sites flanking the Hyg^r gene	Phage-mediated mutagenesis	[54]
pHAE159	50.725	Mycobacterial phage, Amp^R	Phage-mediated mutagenesis	J. I. Kriakov & W. R. Jacobs, Jr, unpublished results
pYUB854_Rv2488c	5.747	Derived from pYUB854 by insertion of 2 regions of *Rv2488c* flanking the Hyg^r gene	Phage-mediated mutagenesis	This work

Supplementary Table 5 - Comparison of *Rv2488c* nucleotide sequences for 25 Mb and 56 Mtb strains. Annotated differences in relation to reference H37Rv are depicted per domain and comprise SNPs and INDELs in six Mb and 23 Mtb strains. Strains not shown, such as *M. bovis* AF2122-97, were identical to reference genome.

		DOMAIN					
Strain	No. isolates with identical profile	Cya3	NMS2	NB_ARC	NMS3	LuxR	Total Δs
Mtb H37Rv	26	Reference genome					
Mb BCG Tokyo 172	4			1			1
Mb BCG Moreau RDJ	4			1			1
Mb BCG Mexico	4			1			1
Mb BCG Korea 1168P	4			1			1
Mb BCG ATCC 35743	4			1			1
Mb BCG Pasteur 1173P2		1		1			2
Mtb 96121	1			2			2
Mtb 96075	2			1			1
Mtb 0B329XDR					1		1
Mtb 0B218DS					5		5
Mtb 0B026XDR						1	1
Mtb 0A117DS				1			1
Mtb 0A036DS					2		2
Mtb 0A005DS				2			2
Mtb Haarlem-NITR202		4	5	33		4	46
Mtb Beijing-NITR203	2			1			1
Mtb RTB423				4	4		8
Mtb RGTB327				1	1		2
Mtb KZN 4207	2		1				1
Mtb KZN 1435	2		1				1
Mtb KZN 605	2		1				1
Mtb HKBS1	2			1			1
Mtb EAI5	1			2			2
Mtb CCDC5180	3			1			1
Mtb CCDC5079	3			1			1
Mtb CAS-NITR204				2	22		24
Mtb BT1	3			1			1
Mtb BT2	3			1			1
Mtb 49-02		7		1			8
Total Δs		12	8	61	35	5	121

Supplementary Table 6 - Comparison of Rv2488c protein sequences for 39 MTC bacteria. Alterations detected in relation to reference H37Rv are depicted per domain and comprise amino acid substitutions in two *M. africanum* (Ma), two Mb and 34 Mtb strains. Aligned protein sequences correspond to sequences directly retrieved from NCBI (i.e. not translated from nucleotide sequences). Mb BCG Pasteur 1173P2 (Mb BCG Pasteur), Mtb

CWCFVRFMDRTB670 (Mtb CW70), Mtb TKK05MA0044 (Mtb TKK05MA) and Mtb TKK05SA0055 (Mtb TKK05SA).

		DOMAIN							
Strain	No. isolates with identical profile	NMS1	Cya3	NMS2	NB_ARC	NMS3	LuxR	NMS4	Total Δs
Mtb H37Rv	3	Reference genome							
Ma GM041182					1				1
Ma MAL010131					1				1
Mb BCG Pasteur					1				1
Mb Bz 31150								1	1
Mtb 2089HD						1			1
Mtb 3499MM					1				1
Mtb BTB05-013			1						1
Mtb BTB06-001				1	1				2
Mtb BTB08-001	3								0
Mtb BTB11-133			1						1
Mtb CWC70					1				1
Mtb H2581			1		1				2
Mtb KT-0011			1						1
Mtb KT-0048						1			1
Mtb KT-0084						1			1
Mtb KT-0085			1						1
Mtb KT-0109	3								0
Mtb KZN 1435			1						1
Mtb M1374						1			1
Mtb M1397					1		1		2
Mtb MAL010087									0
Mtb MAL020179						1			1
Mtb MAL020197			1						1
Mtb OFXR-6		1							1
Mtb SUMu011	3								0
Mtb TB_RSA08						1			1
Mtb TKK02_0069					2				2
Mtb TKK03_0037			1			1			2
Mtb TKK03_0089						1			1
Mtb TKK03_0094					1				1
Mtb TKK03_0099						1			1
Mtb TKK05MA					1				1
Mtb TKK05SA			1						1
Mtb TKK-01-0027					2				2
Mtb UT0014						1			1
Mtb UT0119					1				1
Mtb Wt 21231					1				1
Mtb XTB13-082					2				2
Total Δs		1	9	1	18	10	1	1	41