TESIS DE LA UNIVERSIDAD DE ZARAGOZA

2012

35

María Henar Alonso Ezcurra

Study of the insertion sequence IS6110 in Mycobacterium tuberculosis Beijing strains and its promoter activity

Departamento
Bioquímica y Biología Molecular y Celular

Director/es
Otal Gil, Isabel

ISSN 2254-7606

Prensas de la Universidad Universidad Zaragoza

Universidad Zaragoza
1542

Tesis Doctoral

STUDY OF THE INSERTION SEQUENCE IS6110 IN MYCOBACTERIUM TUBERCULOSIS BEIJING STRAINS AND ITS PROMOTER ACTIVITY

Autor

María Henar Alonso Ezcurra

Director/es

Otal Gil, Isabel

UNIVERSIDAD DE ZARAGOZA

Bioquímica y Biología Molecular y Celular

2012

FACULTAD DE MEDICINA

DEPARTAMENTO DE MICROBIOLOGÍA, MEDICINA PREVENTIVA Y SALUD PÚBLICA

"Study of the Insertion Sequence IS*6110* in *Mycobacterium tuberculosis* Beijing strains and its promoter activity"

Memoria presentada por María Henar Alonso Ezcurra

Licenciada en Bioquímica

Para optar al grado de Doctor por la Universidad de Zaragoza

Directora: Isabel Otal Gil

FACULTAD DE MEDICINA
Departamento de
Microbiología, Medicina
Preventiva y Salud Pública

D. Isabel Otal Gil, Profesora Titular del Departamento de Microbiología, Medicina Preventiva y Salud Pública, como directora de la Tesis Doctoral de **María Henar Alonso Ezcurra** titulada:

"Study of the Insertion Sequence IS*6110* in *Mycobacterium tuberculosis* Beijing strains and its promoter activity"

EXPONE:

Que esta Tesis Doctoral corresponde con el proyecto de tesis presentado y aprobado en su momento, no habiéndose producido ninguna variación.

Que esta Tesis Doctoral reúne los requisitos necesarios para optar al título de Doctor.
Por lo anterior, emito el presente **INFORME FAVORABLE**.

Zaragoza, Junio 2012

Fdo. Isabel Otal Gil

Esta tesis doctoral ha sido elaborada en el Departamento de Microbiología, Medicina Preventiva y Salud Pública, adscrita al programa de Doctorado del Departamento de Bioquímica y Biología Molecular y Celular, siendo María Henar Alonso Ezcurra beneficiaria de un contrato como investigador junior del Consorcio público Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERes), de una beca predoctoral del Proyecto de la Unión Europea del VII Programa Marco (HEALTH-F3-2008-200973) y de una beca predoctoral concedida por el Departamento de Ciencia, Tecnología y Universidad de la Diputación General de Aragón (referencia: B034/09).

Parte de este trabajo es fruto de una estancia de tres meses en el Instituto Pasteur de Corea bajo la supervisión de la Doctora Priscile Brodin, dentro de un Proyecto Europeo para el estudio de la virulencia de las cepas Beijing.

Este trabajo se ha realizado dentro de los siguientes proyectos:

- Proyecto de la Unión Europea del VII Programa Marco: *"Mycobacterium tuberculosis* W-Beijing genetic diversity and differential virulence and host immune responses" (Referencia: HEALTH-F3-2008-200973) desde 2008 al 2011.

- Proyecto del Ministerio de Ciencia y Tecnología "Estudios de los mecanismos de virulencia regulados por phoP en *Mycobacterium tuberculosis*. Aplicación para la construcción de una nueva generación de vacunas" (Referencia: BIO 2008-01561) desde 2009 al 2011.

Index of contents

Index of Figures

Index of Tables

Abreviations list

A	Adenine
A1 (Insertion)	IS*6110* between *dnaA-dnaN* of Beijing strains
ABB	AnnexinV binding buffer
Abs	Absorbance
ADC	albumin-dextrose-catalase
Ap	Ampicillin
AZYTRO	Azytromycin
BC	Before Christ
BCG vaccine or BCG	“Bacille de Calmette et Guérin” vaccine
bp	Base Pairs
BS	Beijing strains
BSL 3	Bio Safety Level 3
C	Cytosine
ºC	Degree celsius
CAS family	Central Middle Eastern Asian family
cDNA	Complementary DNA
CFU	Colony Forming Units
CM	Chloramphenicol
CYC	Cycloserin
CTAB	hexadecyltrimethylammonium bromide
DAT	diacyltrehaloses
DMEM	Dulbecco’s modified Eagle medium
DNA	Deoxyribonucleic acid
dNTPs	deoxyribonucleotides
DOTS	Directly Observed Therapy Short-Course
DR	Direct Repeats
DR Region	Direct Repeat Region
EAI family	East-African-Indian family
EDTA	Ethylenediaminetetraacetic acid
EMB	Ethambutol
FACS	Fluorescence activated cell sorting
FBS	fetal bovine serum
FMI	Fluorescence median intensity
G	Guanine
GFP	Green Fluorescent Protein

Gm	Gentamicin
GTC buffer	guanidinium thiocyanate buffer
h	hour
HCS	High copy strains, referred to high copy number of IS*6110*
H family	Haarlem family
Hyg	Hygromycin
INH	Isoniazid
IPT	Isoniazid Preventive Treatment
IR	Inverted Repeats
IS	Insertion Sequence
Km	Kanamycin
LAM family	Latin-American-Mediterranean family
LAM	lipoarabinomannan
LCS	Low copy strains, referred to low copy number of IS*6110*
LM	Lipopolysaccharides lipomannan
LMPCR	Ligation Mediated PCR
LSP	Large sequences of polymorphism
m	Mycolic acids
mAG	mycolyl arabinogalactan
MANU family	Indian TB family
MCS	Multiple Cloning Site
MDR	Multi Drug Resistant
MH-S	Murine immortalized alveolar macrophages
MIC	Minimum Inhibitory Concentration
min	minute
MOI	multiplicity of infection
mRNA	messenger RNA
MSH	Mycothiol
MTB	*Mycobacterium tuberculosis*
MTBC	*Mycobacterium tuberculosis* complex
ml	mililitres
µg	micrograms
µl	microlitres
µm	micrometre
NBS	Non-Beijing strains
NaAc	Sodium acetate
NaCl	Sodium chloride
ng	nanogram

nm	nanometres
NO	Nitric oxide
nsSNP	nonsynonymous Single Nucleotide Polymorphism
NTF region	genetic marker specific to the Beijing genotype
OADC	oleic acid-albumin-dextrose-catalase
OD	Optical Density
OP6110	Promoter Region of IS*6110*
ORF	Open Reading frame
OriC	Replication origin
PAT	polyacyltrehaloses
PAS	Para-aminosalicylic Acid
PAZ	Pyrazinamide
PBS	phosphate buffered saline
PCR	Polymerase Chain Reaction
PDIM	Phthiocerol dimycocerosates
PE	Proline-glutamic acid
PFA	Paraformaldehide
PPE	Proline-Pproline glutamic acid
PG	Peptidoglycan
PGG	Principal genetic groups
PGL	Phenolic Glycolipid
PIM	Phosphatidylinositol mannosides
plc region	Phospholipase C region
pmol	picomol
qRT-PCR	Quantitative-Reverse Transcription Polymerase Chain Reaction
RD	Region of Difference
RIF	Rifampicin
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic acid
rRNA	ribosomal RNA
ROI	Reactive Oxygen Intermediates
rmp	Revolutions per minute
RQ	Relative Quantification
RT-PCR	Reverse transcription Polymerase Chain Reaction
RvD	Regions absent in *M. tuberculosis* H37Rv
s	second
SCGs	SNP Cluster Groups
s.d.	Standard deviation

SM	Streptomycin
SNP	Single Nucleotide Polymorphism
sSNP	silent- Single Nucleotide Polymorphism
SL	Sulfolipids
ST	spoligotypes
T family	Modern TB strains family
tRNA	transfer RNA
TB	Tuberculosis
TDR	Totally Drug-Resistant
tsp	Transcriptional start point
U	Unit
UEPs	Unique-event poliporphisms
X family	IS-low-banding family
XDR	eXtensively Drug Resistant
WHO	World Health Organization
WT	Wild Type

Resumen

Mycobacterium tuberculosis es el agente causal de la tuberculosis (TB). Esta enfermedad ha acompañado a la raza humana a lo largo de su historia y hoy en día sigue siendo una de las enfermedades infecciosas que causa mayor número de muertes en el mundo. Aunque se han realizado grandes avances en la lucha contra esta enfermedad, sigue siendo un problema de salud mundial.

En la actualidad, una de las familias de *M. tuberculosis* que más atención atrae en todo el mundo es la familia Beijing, debido a que muestran importantes características patogénicas. Diferentes estudios han indicado que un tercio de los casos de tuberculosis a nivel mundial son causados por cepas de *M. tuberculosis* pertenecientes a la familia Beijing, siendo ésta una de las familias de *M. tuberculosis* con mayores tasas de morbilidad y mortalidad. Su capacidad de infección, así como su capacidad para propagarse rápidamente, probablemente se deban a ventajas genéticas y a mecanismos de virulencia aún no identificados. Entre los mecanismos propuestos causantes de las diferencias de virulencia entre los fenotipos de las cepas de *M. tuberculosis* se encuentran inserciones de la secuencia IS*6110*, deleciones duplicaciones y reorganizaciones genéticas. Se ha demostrado que esta secuencia puede aumentar la expresión de genes adyacentes actuando como un promotor móvil localizado en su extremo 3'. El hecho de que las cepas de la familia Beijing tengan un mayor número de copias de IS*6110* que otras familias, podría estar relacionado con las características especiales de virulencia y de transmisión de esta familia. La cepa de *M. tuberculosis* GC1237, perteneciente a la familia Beijing, ha demostrado una elevada capacidad de diseminación y transmisión dentro de una comunidad.

Con el fin de comprender mejor algunos de los mecanismos de virulencia de las cepas Beijing y relacionarlos con la secuencia de inserción IS*6110*, se caracterizó el aislado clínico de *M. tuberculosis* GC1237. Se localizaron todas las copias de este elemento en el genoma de la cepa GC1237 y se realizó un análisis más detallado de una copia de IS*6110* localizada a 31 nucleótidos del inicio del gen esencial Rv2179c y próxima al gen esencial *aroG.* Mediante el uso de varias técnicas se confirmó que esta copia de IS*6110* está actuando como un promotor móvil de Rv2179c, tanto en cultivo líquido como en el interior del macrófago. A partir de estos resultados podemos sugerir que la sobre-expresión de este gen podría ser ventajosa para la cepa Beijing GC1237, al menos en ciertos entornos, como en el interior de las células infectadas. Además, esta localización resultó ser única de esta cepa lo que permitió el diseño de un test rápido para su detección.

Por otro lado, en anteriores investigaciones se estudió la actividad promotora de IS*6110* en la copia de IS*6110* localizada en la región promotora del gen *phoP* en *M. bovis* MBZ. Debido a la condición MDR esta cepa, este estudio se llevó a cabo mediante el uso de dos plásmidos recombinantes en el interior *M. smegmatis* mc2155, en cual se observó un incremento en la transcripción del gen *phoP* en la construcción que contenía la IS*6110*. Con el objetivo de profundizar en el estudio del efecto de esta copia de IS*6110* sobre el gen *phoP*, se realizaron nuevas construcciones de plásmidos que contenían esta región y se introdujeron en *M. tuberculosis*. A continuación, se analizó la actividad promotora de IS*6110* por varios métodos y en condiciones diferentes, tanto en cultivo líquido como en el interior de macrófagos. Los resultados obtenidos confirmaron que esta copia de IS*6110* está actuando como secuencia promotora del gen *phoP*. Este hecho cobra gran importancia ya que el gen *phoP* codifica para un factor de transcripción que regula ~2% del genoma de *M. tuberculosis* y está implicado en virulencia.

Por último, dado que las cepas de *M. tuberculosis* Beijing contienen un mayor número de copias de IS*6110* lo que podría aportar alguna ventaja a este genotipo, se realizó una búsqueda de puntos de inserción de IS*6110* en 61 aislados clínicos, entre ellos 17 cepas Beijing y 44 cepas no-Beijing, los cuales presentaban un alto número de copias de este elemento. La ubicación de las copias de esta secuencia permitió la detección de los puntos de mayor frecuencia de inserción de IS*6110*, o “hot-spots”, en las cepas Beijing. Estas localizaciones podrían ser la clave de algunas ventajas del genotipo Beijing y su estudio debe de ser tenido en cuenta en una futura investigación.

Summary

Mycobacterium tuberculosis is the aetiological agent of tuberculosis (TB). This disease has walked together with human beings along our history, and nowadays is still one of the infectious diseases that causes more deaths in the world. Although in the last century great advances have been done in our fight against TB, this disease is still a worldwide health problem.

One of the families of *M. tuberculosis* currently attracting considerable worldwide attention is the Beijing family as they display important pathogenic features. Different studies have indicated that one-third of global TB cases are caused by this family assigning this lineage to one of the most successful mycobacterial families in terms of morbidity and mortality. The capacity of infection and the ability of *M. tuberculosis* Beijing strains to spread rapidly probably result from genetic advantages and unidentified mechanisms of virulence not yet thoroughly investigated. Among the mechanisms proposed to be responsible for differences in the virulence phenotypes of *M. tuberculosis* strains we find IS*6110* insertions, genetic reorganizations, genetic duplications and deletions, which have strong influence on fitness. It has been demonstrated that IS*6110* may increase the expression of neighboring genes through an outward-directed promoter in its 3' end. The fact that the Beijing lineage contains a larger number of IS*6110* copies than other lineages could be related with the special characteristics of this family in terms of virulence and capacity for rapid dissemination. The *M. tuberculosis* GC1237 strain, which belongs to Beijing family, has demonstrated an enhanced capacity to spread and transmit within a community.

In order to better understand some mechanisms of virulence of Beijing strains and related them with the insertion sequence IS*6110*, we have characterized the clinical isolate *M. tuberculosis* GC1237 strain locating all the copies of this element. We focused in the characterization of one copy located 31 bp upstream of the essential gene Rv2179c and close to the essential gene *aroG* and, by several methods, we confirmed that this copy of IS*6110* is acting as a mobile promoter both in broth and inside macrophages. From our results we can hypothesize that the over-expression of this gene could be advantageous for GC1237 at least in certain environments such as infecting macrophages. In addition, this location was unique of this strain allowing us the design of a rapid test for its detection.

On the other hand, in previous investigations, the promoter activity of IS*6110* was studied in the copy of IS*6110* located in the promoter region of *phoP* gene of *M. bovis* MBZ. Due to the MDR condition of this strain, this study was carried out by using two recombinant plasmids inside *M. smegmatis* mc^2155 obtaining an increment in the

transcription of *phoP* gene in the construction containing IS*6110*. To depth in the study of the promoter activity of this IS*6110* in *phoP* gene, new plasmid constructions containing this region were made and introduced in *M. tuberculosis.* After that, the promoter activity was analyzed by several methods and in different conditions both in broth and inside macrophages. Our results confirmed that this copy of IS*6110* is acting as a promoter of *phoP* gene. This fact could be important as this gene codifies for a transcription factor that regulates ~2% of *M. tuberculosis* genome and is involved in virulence.

Finally, as *M. tuberculosis* Beijing contain a larger number of IS*6110* copies and these could provide some advantages to this genotype, a scrutiny of points of insertion of IS*6110* was carried out in 61 clinical isolates with high number of copies of this element, 17 Beijing and 44 non-Beijing strains. The location of the copies of this sequence allowed us to detect new hot-spots for the insertion of IS*6110* specific of Beijing strains. These locations could be the key of some advantages of the Beijing genotype and its study should be taken into account in further research.

GENERAL INTRODUCTION

I.Tuberculosis, an old disease

Tuberculosis (TB) is a common and often deadly infectious disease caused by *Mycobacterium tuberculosis*. Tuberculosis usually attacks the lungs (as pulmonary TB) but can also affect almost any organ system (extrapulmonary TB). *M. tuberculosis* is an evolutionary recent pathogen which is extremely well adapted to the human host and its success could be due to its ability to persist for years. Currently, one third of the world's population is infected with *M. tuberculosis*, and new infections occur at a rate of one per second (134). However, most of these cases will not develop the full-blown disease; asymptomatic, latent infection is most common. About one in ten of these latent infections will eventually progress to active disease, which, if untreated, kills more than half of its victims.

The History of tuberculosis: from prehistory to nowadays

Tuberculosis is the greatest man killer in history; taking only the past two centuries into account, TB was responsible for the deaths of approximately one billion human beings (104). The first evidence was found in a cemetery near Heidelberg, in the Neolithic bone remains that show evidence of the type of angulation often seen with spinal tuberculosis. Some authors call TB the first disease known to mankind. Signs of the disease have also been found in Egyptian mummies dated between 3000 and 2400 Before Christ (BC) (32, 91, 106). The most convincing case was found in the mummy of priest Nesperehen, discovered by Grebart in 1881, which featured evidence of spinal tuberculosis with the characteristic psoas abscesses. Similar features were discovered on other mummies like that of the priest Philoc and throughout the cemeteries of Thebes. It appears that Akhenaten and his wife Nefertiti both died from tuberculosis, and evidence indicates that TB hospitals existed in Egypt as early as 1500 BC. The Ebers papyrus, an important Egyptian medical treatise from around 1550 BC, describes a pulmonary consumption associated with the cervical lymph nodes. It recommended that it has to be treated with the surgical lancing of the cyst and the application of a ground mixture of acacia seyal, peas, fruits, animal blood, insect blood, honey and salt.

Tuberculosis has also been called *consumption*, because it seemed to consume people from within, with a bloody cough, fever, pallor, and long relentless wasting. Another name was *phthisis* first appeared in Greek literature around 460 BC. Hippocrates identified phthisis as the most widespread disease of the times. It most commonly occurred between 18 and 35 years of age, and was almost always fatal. He wrote of patients with consumption, wasting away associated with chest pain and coughing, frequently with blood in the sputum (113). Although Aristotle believed that

the disease might be contagious, many of his contemporaries believed it to be hereditary. Galen, the most eminent Greek physician after Hippocrates, defined phthisis as the "ulceration of the lungs, thorax or throat, accompanied by a cough, fever, and consumption of the body by pus" (59).

It is believed that Europe became the epicenter of many TB epidemics from the 16th and 17th century due to the growing population, the industrialization and the development of large urban settings. Tuberculosis was supposed to be an incurable disease and it was called "*The Great White Plague*" or the "*White death*" because sufferers appear markedly pale. Death by tuberculosis was considered inevitable, being the principal cause of death in 1650. The high population density as well as the poor sanitary conditions that characterized most European and North American cities created a perfect environment for the propagation of the disease. The epidemic reached its peak in Europe in the first half of the 19th century, and it is estimated that as many as one quarter of Europeans died of TB at this time (113).

The hope: curable disease?

In 1854, Hermann Brehmer proposed the idea that "*tuberculosis was indeed a curable disease*". The introduction of the "sanatorium" cure provided the first big step toward treatment for tuberculosis (27, 59). Brehmer himself was a TB patient. His doctor advised him to move to a healthier climate, so he spent some time in the Himalayas and came home cured. This experience moved him to build the first sanatorium, a place where patients could get plenty of fresh air, exposure to sunlight and good nutrition. This setup became the blueprint for the subsequent development of sanatoriums. The measures available to doctors at the time were still modest. Improving social and sanitary conditions and ensuring adequate nutrition were all that could be done to strengthen the body's defenses against TB bacteria. Sanatoriums, founded throughout Europe and the United States, provided a dual function: to isolate the sick people from the general population, and to force the patients to rest, thus assisting the healing process.

In 1865, Jean Antoine Villemin demonstrated that consumption could be transmitted from humans or cattle to rabbits. In the light of this revolutionary evidence, he postulated that a specific microorganism caused the disease (59).

On March 24th 1882, Robert Koch presented his lecture "The etiology of tuberculosis". The experiments he described in the lecture defined tuberculosis as an infectious disease in accordance with his postulates (Figure 1). Once *M. tuberculosis* had been identified, the next step of Koch was to find a cure for the disease. Eight years later

Koch announced the discovery of a vaccine against tuberculosis consisting of glycerol extracts of pure cultures of tubercle bacilli. The vaccine candidate was assayed in a clinical trial with disastrous results but in 1908, Charles Mantoux found that it was an effective intradermic test for diagnosing tuberculosis. Regardless of this failure, Koch's observations settled the basis for microbiology known as the Koch's postulates. These comprise identification of the pathogen in affected tissue, growth of single clones of the pathogen in vitro, and establishment of a similar disease in the experimental animals by means of the pure culture (64).

With Edward Jenner's successful invention, showing that infection with cowpox would give immunity against smallpox in humans, many doctors placed their hopes on the use of *Mycobacterium bovis* (the agent that causes bovine TB) for the development of a vaccine against human TB. In 1908 Albert Calmette and Camille Guérin started the development of a vaccine against tuberculosis. They isolated from a dead cow *M. bovis* and for the next 13 years, these scientists grew every fortnight a new batch of bacteria in a solution of beef bile and potato. Eventually, the bacteria lost their ability to cause disease, but were still capable of stimulating the immune system to protect mice, pigs, guinea pigs and monkeys from tuberculosis. The vaccine was called "Bacille de Calmette et Guérin" or "BCG" and it was first administered to an infant in 1921 (Figure 1). A total of 8 million babies and nearly 14 million people were given the BCG vaccine in the International Tuberculosis Campaign, which ran through 1951. The project initially began in Europe in the aftermath of World War II. However, the program extended beyond Europe when UNICEF contributed two million dollars to expand the program to other continents. Since then more than a billion people have been vaccinated with BCG (9).

In 1943, Selman A. Waksman, who had been working for decades to find an effective antibiotic against *Mycobacterium tuberculosis*, was finally successful (Figure 1). Streptomycin purified from *Streptomyces griseus* was first administered to a human on November 20th, 1944. The results were extremely impressive. The disease immediately stopped its progression, the bacteria disappeared from the patient's sputum, and he recovered fully. A rapid succession of anti-TB drugs appeared in the following years: para-aminosalicylic acid (PAS, 1949), isoniazid (INH, 1952), pyrazinamide (PAZ, 1954), ethambutol (EMB, 1962) and rifampicin (RIF, 1963). This advance was very important because with the streptomycin (SM) treatment, resistant mutants appeared within a few months, endangering the success of antibiotic therapy. However, it was soon proven that using a combination of drugs would solve this problem (59).

The unavoidable return of TB

The hopes that the disease could be completely eliminated were dashed in the 1980s with its dramatic resurgence, the rise prevalence of multidrug-resistant strains (MDR) and the deadly form of MDR associated with HIV/AIDS (48).

In response to this resurgence, the World Health Organization (WHO) issued a declaration of a global health emergency in 1993 (135) (Figure 1). Every year, nearly half a million new cases of multidrug-resistant tuberculosis (MDR-TB) are estimated to occur worldwide. Even more alarming is the existence of at least 12 patients infected with tuberculosis that has become resistant to all the drugs used against the disease. It is called TDR, for Totally Drug-Resistant. In other words, they are untreatable as far as it is known. The first cases of these TDR were reported in Iran in 2009 and Italy in 2007 (72, 127).

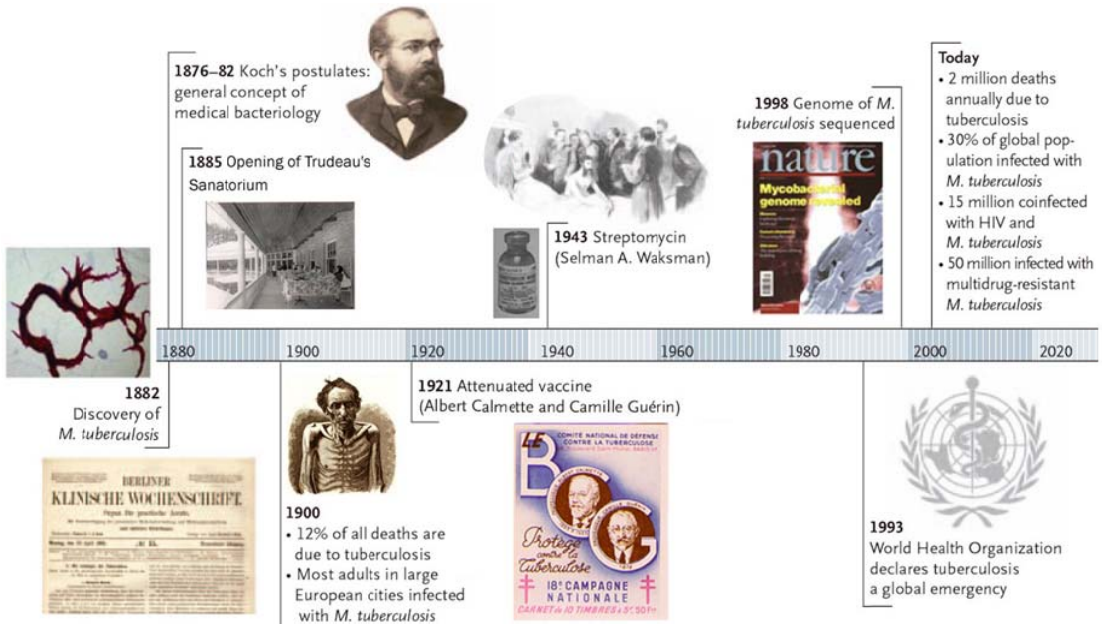


Figure 1: Timeline in tuberculosis research. It shows the most relevant highlights in the fight against this disease. Adapted from (64).

Tuberculosis is today the second major cause of death from any infectious disease, only surpassed by HIV/AIDS. In 2010, there were 8.8 million (range, 8.5–9.2 million) incident cases of TB, 1.1 million (range, 0.9–1.2 million) deaths from TB among HIV-negative people and an additional 0.35 million (range, 0.32–0.39 million) deaths from HIV-associated TB (134).

II.The disease: Pathogenesis and intracellular lifestyle of *M. tuberculosis*

Pathogenesis of *M. tuberculosis*

Tuberculosis is a contagious disease (113). Infection with *M. tuberculosis* occurs through inhalation of aerosols containing the bacteria which are spread by persons with active pulmonary TB. After inhalation, bacteria are settled in the alveoli and disseminated by the lymphatic circulation. Further dissemination to other parts of the lung and occasionally to other organs is achieved by haematogenous circulation. The most common form of the disease is pulmonary tuberculosis, although TB-meningitis, miliary (disseminated) tuberculosis, lymphadenitis, osteomyelitis and Pott's disease (affected bones) also occur (113). Primary infection leads to active disease in about 10% of infected individuals, in 80% of the cases in the period of two years (125). In the remaining 90%, the immune system controls the infection, and the individual is non-infectious and asymptomatic. In this clinical state the TB bacilli can lie dormant for years (latent TB) (102). However, when the immune system is weakened, the latent infection can reactivate (113, 125). In an HIV-infected person the risk of reactivation of latent TB is higher than 10% per year, compared to a lifetime risk of 10-20% for HIV-negative individuals (23, 111, 125).

Life inside the macrophages

Once the bacteria are phagocytosed by alveolar macrophages, mediated by specific macrophage receptors (36, 38, 62, 110), a localized proinflammatory response that leads to recruitment of mononuclear cells from neighboring blood vessels is induced. These cells will build the granuloma (necessary to contain infection) which consists of a group of infected macrophages, surrounded by foamy giant cells (multinuclear macrophages loaded with lipids) and macrophages with a layer of lymphocytes delineating its periphery (101-103) (Figure 2).

Activated macrophages promote phagosomal maturation by a process that finally involves fusion with lysosomes and formation of the phagolysosome. These vesicles provide a hostile environment for the bacilli including acid pH, reactive oxygen intermediates (ROI), lysosomal enzymes and toxic peptides. *M. tuberculosis* has evolved to survive within macrophages by arresting the normal phagosomal maturation at an early stage, thereby restricting its acidification and limiting fusion with lysosomes (90, 98). Modulation of phagosome maturation seems to be mainly mediated by both mycobacterial cell-wall lipids and other bacterial effectors (e.g., SapM or PknG) (46, 132, 133). Additionally, the screening of shotgun libraries has enabled the isolation of several *M. tuberculosis* mutants unable to induce phagosomal arrest (96, 118). Apart

from surviving within the macrophage, pathogenic mycobacteria have also been suggested to escape into the cytosol in an ESX-1-dependent manner (128). In addition, *M. tuberculosis* reduces phagosome acidification by the exclusion of the proton-ATPase (119) and decreases ROI and nitric oxide (NO) synthesis and cytokine production by modulating Ca^{2+} fluxes in infected macrophages (76). Moreover, *M. tuberculosis* has evolved multiple strategies to detoxify ROI and RNI (37). Recently, authophagy (29, 57) and apoptosis (68) have been postulated as effective antimycobacterial mechanisms.

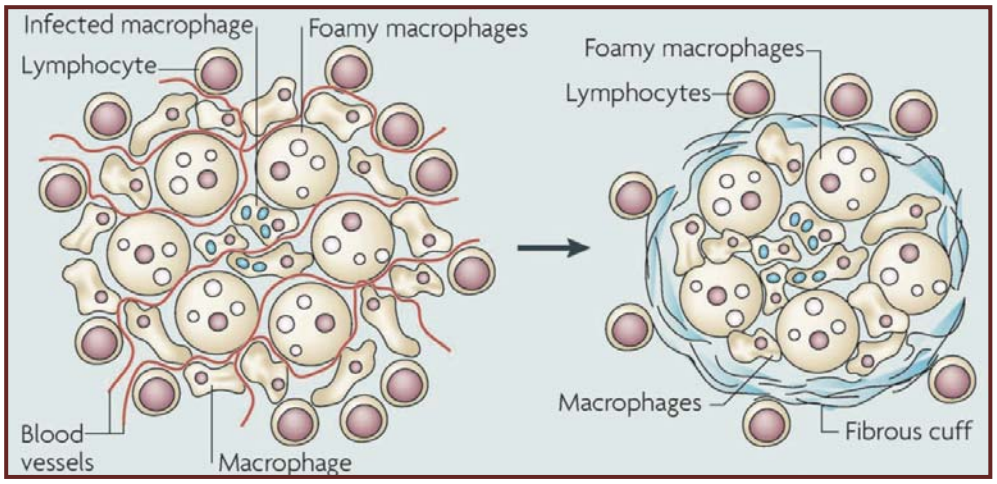


Figure 2: Progression of TB granuloma. In the early stage, the granuloma has a core of infected macrophages enclosed by foamy macrophages and other mononuclear phagocytes, surrounded by lymphocytes. As the granuloma matures, it develops a fibrous capsule that encases the macrophage core and excludes the majority of lymphocytes from the center of the structure. Adapted from (102).

Measures against tuberculosis:

- **Antibiotic therapy**

Tuberculosis treatment is difficult and requires long courses (around 6 to 12 months) of multiple antibiotics. It consist of an initial intensive phase of treatment (2 months) with RIF, INH, PAZ and either EMB or SM designed to kill actively growing and semi-dormant bacilli, followed by a continuation phase (4 months) with RIF and INH to eliminate residual bacilli and reduce the number of relapses. To complete the treatment it is critical to reduce the development of acquired drug resistance and therefore trained individual personnel supervises that the patient takes each dose of medication. This strategy is known as directly observed treatment short-course (DOTS) and was promoted as the official policy of the WHO in 1991 (136).

Isoniazid Preventive Treatment (IPT) for six months is recommended for high risk individuals, like HIV infected individuals, or children who are in close contact with an infectious person with the aim of decreasing the risk of primary infection with *M. tuberculosis* (63).

Another major challenge is the ever-increasing resistance to anti-TB drugs. Multidrug-resistant TB refers to TB caused by *M. tuberculosis* isolates that are resistant to the most effective drugs, INH and RIF. MDR strains take longer to treat (up to two years) with second-line drugs, which are more than 100 times more expensive, less effective and also much more toxic to the patients (49). The virtually untreatable extensively drug-resistant TB (XDR-TB) is defined as MDR-TB which is also resistant any of the fluoroquinolones and second-line anti-TB injectable drugs (Amikacin, Kanamycin (Km) or Capreomycin).

- **Vaccination**

BCG is one of the most widely used vaccine today (> 80 % of neonates and infants in some countries). Although it does not prevent primary infection or reactivation of latent pulmonary TB, the principal source of bacillary spread in the community, BCG protects against disseminated forms of TB, especially against meningitis in children. For that reason, BCG vaccination is still recommended by the WHO and applied widely (115, 123).

After its first use in humans in 1921, propagation of BCG in non-standardized conditions in different laboratories all around the world gave rise to different variant strains that differ in genotype and phenotypic characteristics (9). Strain differences could be one cause of the variable protective efficacy afforded by BCG (44), but other factors such as population genetics or exposure to environmental mycobacteria could be implicated (28).

Due to this inconsistency in protective efficacy, and to the inability of BCG to prevent pulmonary forms of tuberculosis and consequently, to prevent transmission of the disease, great efforts are being made by governments, research institutions and private foundations for constructing and testing new promising vaccine candidates.

III.Biology of the bacillus and mycobacterial genetics

General characteristics of mycobacteria

Mycobacterium genus comprises a number of Gram-positive, acid-fast, rod-shaped aerobic bacteria and it is the only member of the family *Mycobacteriaceae* within the order *Actinomycetales* (117). Like other closely related *Actinomycetales*, such as *Nocardia* and *Corynebacterium*, mycobacteria have unusually high genomic DNA GC content and are capable of producing mycolic acids (m) as major components of their cell wall. Phenotypically *M. tuberculosis* is catalase and nitrate reductase positive, non-motile and non-sporulating of 2-5μm in length and 0.2-0.5μm in width.

To date, more than 100 species of mycobacteria have been described (122). Most of these species are environmental saprophytes. Phylogenetic analyses and phenotypic characteristics, such as growth rate and pigmentation, can be used to classify these species. The classical distinction between rapid and slow growth is based on the ability of strains to develop clearly visible colonies in less or more than 7 days, respectively. Rapid growers (< 7 days) are free, environmental, saprophytic species. Interestingly, the slow growing mycobacteria comprise most of the pathogenic species, including *M. tuberculosis* which has a generation time of ~24 hours. This contributes to the chronic nature of the infection, requires long-term treatment to fully eradicate the pathogen and imposes a barrier for microbiologists since it takes up to four weeks for a single *M. tuberculosis* cell to become a colony on solid media.

The various etiologic agents of TB are clustered in the *M. tuberculosis* complex (MTBC), which currently includes eight species: The human-adapted strains *M. tuberculosis*, *M. africanum* and *M. canettii* (being *M. canettii* the most divergent within the MTBC) and MTBC also includes the animal-adapted strains *M. bovis* (primarily a pathogen of cattle), *M. caprae* (a pathogen of goats), *M. microti* (a pathogen of rodents) and *M. pinnipedii* (a pathogen of seals and sea lions) and, recently *M. mungi* was isolated from mongoose (2, 89). Mycobacteria grouped in MTBC are characterized by 99.9% similarity at the nucleotide level and identical 16S rRNA but that differ in terms of phenotypes and host preference (13).

The envelope of mycobacteria

The unique cell wall of *M. tuberculosis* forms a waxy envelope and has been established as an important factor leading to bacterial virulence and survival. This envelope differs considerably from cell wall of both Gram-positive and Gram-negative bacteria. Even though they are structurally more closely related to Gram-positive bacteria, mycobacteria do not fit into the Gram-positive category as the molecules

attached to the cell wall are distinctively lipids rather than proteins or polysaccharides. Frequently, they do not retain the crystal violet and appear as “ghosts” after Gram staining (7). Therefore, acid-fast stains, such as the Ziehl-Neelsen stain or the fluorescent auramine-rhodamine stain are recommended for the detection of mycobacteria.

A characteristic feature of their cell envelope is the high proportion of lipid-rich molecules. Schematically, the cell envelope from the inside to the outside, comprises a typical bacterial plasma membrane, a cell wall and an external capsule, which contains proteins, polysaccharides and lipids (Figure 3) (26). The outer layer presents various non-covalently attached lipids and glycolipids esterified with multimethyl-branched long-chain fatty acids (61, 86). These complex lipids are restricted, with few exceptions, to pathogenic mycobacteria and are suggested to play important roles in pathogenicity. In *M. tuberculosis*, these lipids include the phthiocerol dimycocerosates (PDIM) (5, 18, 24, 92, 100) and the closely related phenolic glycolipids (PGL) (93, 97, 112), the trehalose ester families that include sulfolipids (SL) (22, 31, 54), diacyltrehaloses (DAT) and polyacyltrehaloses (PAT) (69, 99), and the family of mannosyl-β-1-phosphomycoketides (80). This highly hydrophobic envelope decreases the permeability and susceptibility to degradation and makes the bacterium naturally resistant to most of the antibiotics and chemical agents and confers unique staining properties.

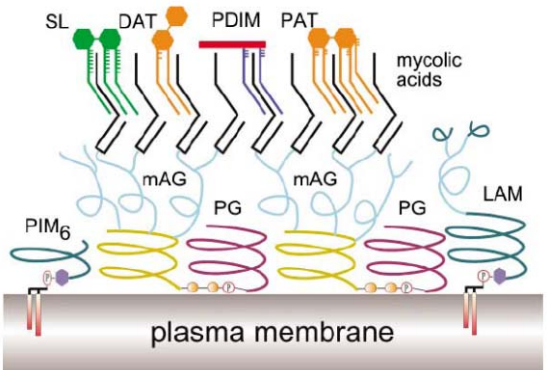


Figure 3: Arrangement of structural components in the cell envelope structure of *M.tuberculosis.* The cell wall consists of a peptidoglycan (PG) covalently attached to the mycolyl arabinogalactan (mAG), which is in turn esterified by mycolic acids (m). The lipopolysacharides lipomannan (LM) and lipoarabinomannan (LAM) are anchored to the plasmatic membrane through phosphatidylinositol mannosides (PIM) and they extend towards the exterior of the cell wall (86).

Mycobacterial genomics

The complete genome sequence of the most studied and the best understood strain of *M. tuberculosis*, H37Rv, was obtained in 1998 (113) and revised and re-annotated in 2002 and it is in continuous revision (as of August 2010; NCBI Entrez Genome Project). Complete genome sequences are now available for five *M. tuberculosis* strains, CDC1551, F11, H37Ra, H37Rv and KZN1435, and sequencing of 71 other *M. tuberculosis* genomes is in progress providing a valuable tool for research using genetic techniques (70). This information is available in sequencing websites, as well as in several online databases (http://tuberculist.epfl.ch/, http://www.ncbi.nlm.nih.gov/ or http://www.microbesonline.org/).

The genome comprises 4,411,532 bp, with a G+C content of 65.9% (21). Only a few regions show a skew in this G+C content. A conspicuous group of genes with a very high G+C content (> 80%) appear to be unique in mycobacteria and belong to the family of PE or PPE proteins. In turn, the few genes with particularly low (< 50%) G+C content are those coding for transmembrane proteins or polyketide synthases. This deviation to low G+C content is believed to be a consequence of the required hydrophobic amino acids, essential in any transmembrane domain, that are coded by low G+C content codons. The last re-annotation identified 4090 genes thought to encode 3,989 proteins and 71 stable RNAs (Figure 4) (70).

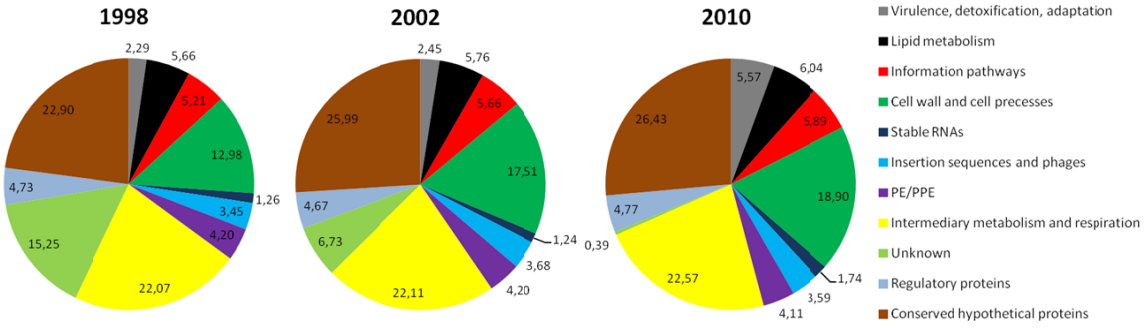


Figure 4: Functional classification of *M. tuberculosis* H37Rv genes. Timeline of the genomic content of *M. tuberculosis* H37Rv. Adapted from (70).

Some outstanding features include the great abundance of genes involved in lipid metabolism which is indicative of the highly active metabolism mainly in relation with the cell envelope. There are genes encoding for 250 distinct enzymes involved in fatty acid metabolism, compared to only 50 in the genome of *Escherichia coli* (20). The PE-PPE family of proteins, which represents approximately 4% of the genome, includes 168 genes. The names come from Pro-Glu (PE) and Pro-Pro-Glu (PPE) sequences

found in the two conserved N-terminal regions in each of these protein families that are approximately 110 and 180 amino acids long, respectively. Although the function of the members of the PE and PPE protein families has not been established, they are suggested to be involved in antigenic variation and disease pathogenesis (29). Tubercle bacilli also contain complex regulatory machinery comprising 11 two component systems, 11 serine/threonine protein kinases, 13 σ factors, and over 100 repressors and activators, suggesting a tight regulation of transcription in response to different stimuli. Interestingly, the analysis of the DNA metabolic system of *M. tuberculosis* indicates a very efficient DNA repair system, in other words, replication machinery of exceptionally high fidelity. The genome of *M. tuberculosis* lacks the MutS-based mismatch repair system. However, this absence is overcome by the presence of nearly 45 genes related to DNA repair mechanisms (87), including three copies of the *mutT* gene. This gene encodes the enzyme in charge of removing oxidized guanines whose incorporation during replication causes base-pair mismatching (20, 87). Repetitive DNA is very important in mycobacterial genomes. Elements like Tandem Repeats, Interspersed Repeats, and mobile genetic elements are common. In H37Rv, 56 loci with similarity to IS elements have been found that can be classified into major IS families (IS3, IS5, IS605 or IS21). Mobile genetic elements will be widely analyzed in the following section as they could have an important role in genome plasticity and the virulence of the bacilli.

Mobile genetic elements in mycobacteria

Insertion sequence (IS) is a short DNA mobile genetic element coding for proteins involved in the transposition activity, which allows it to spread within the genome. ISs are widely distributed in prokaryotes and can be grouped into different families based on structure characteristics and transposase similarities (75). More than 46 ISs from different species have been located and identified in *Mycobacterium*, mostly on the basis of sequence similarities (15).

In the genome of the members of MTBC it has been possible to find dispersed IS elements that could be included in several of the following families attending to their characteristics: IS*3*, IS*5*, IS*21*, IS*30*, IS*110*, IS*256*; IS*1535*, IS*L3* and other IS-*like* elements (53).

The IS*3* family consist of an extensive set of insertion elements in bacteria. The features that characterize this family are their length, between 1200 and 1600 bp, and their inverted repeats (IRs), between 20 and 40 bp long, as well as the presence of two overlapping open reading frames (ORFs: *orf*A and *orf*B) (75, 81). After the insertion, a duplication of 3 or 4 bp occurs at the insertion point (85). Recombination is also

another mechanism participating in the changes of the location of ISs along the genomes. All those mechanisms lead to IS-mediated gene rearrangements, inversions, deletions and duplications in the bacterial genomes. Examples of sequences of this family in MTBC are IS*1540*, IS*1604*, IS*1556*/*990* and IS*6110*. The IS*1540*, IS*1604* and IS*1556*/*990*, have missed the IRs or contain mutations in *orf*B making them supposedly inactive and non-functional (34, 82). However, the most representative member of this family, IS*6110*, has the two overlapping ORFs intact and a functional transposase.

The insertion sequence IS*6110*

IS*6110* was initially named IS*986*. Subsequently, a related element from *M. bovis* BCG was sequenced by Hermans *et al*. (58) and is referred to as IS*987*. The three sequences IS*6110*, IS*986* and IS*987* were practically identical and to avoid confusion, it was decided to use the same name IS*6110* (81, 120). IS*6110* is a genomic insertion element 1361 bp long, it shows 28 bp imperfect IRs, and due to the mechanism of transposition it generates duplications of 3 or 4 bp (known as Direct Repeat, DR) next to the insertion site (Figure 5). It may be present up to 25 copies per genome in *M. tuberculosis* (15) and only a few number of strains have no copies of this element (73). Recently, it has been suggested variations into the sequence of IS*6110* copies from different strains of *M. tuberculosis*, which could have implications in its usefulness as target of PCR detection (109).

Figure 5: Complete sequence of IS*6110*. The direct repeats (DR) (red) and the invert repeats (IR) (green) flanking the transposase gene are underlined.

- **Its success as a diagnostic and epidemiological tool**

Soon after the discovery of IS*6110* as a specific element in MTBC, its usefulness as diagnostic tool was explored (121). Subsequently, at the beginning of the nineties it was demonstrated that two strains isolated from different episodes of a patient had the same IS*6110*-RFLP in turn, a high degree of polymorphism was observed between strains isolated from different patients (95). The fact that IS*6110* varies in copy number and it is randomly distributed throughout the genome, along with its stability over time showed usefulness in genotyping of the MTBC. DNA fingerprinting of *M. tuberculosis*, standardised in 1993 (129), has shown to be of great value in comparison of the results obtained by different laboratories, distinguishing between recent transmission and reactivation, reinfection, mixed infections, studies of outbreaks, and confirmation or ruling out laboratory errors. It has also been useful to identify some strains that may differ in transmission, suggesting that more virulent strains could show different pathogenesis and epidemiologic characteristics. Additionally, in some cases, the location of one copy of IS*6110* specific to one strain can be used as a useful tool in its rapid diagnosis allowing the identification and differentiation of this particular strain (85).

- **Where is IS*6110* inserted?**

IS*6110* does not have a known target or consensus sequence, it has been found within ORFs and intergenic regions (39). The interruption of coding regions can be seen as a sort of natural knock-out mutation of the target gene. The IS*6110* was found to be inserted more often in some genome regions (hot-spots). The first identified hot-spot was the Direct Repeat region (the DR region) (58) and with minor exceptions, all members of the MTBC carry a copy of the IS*6110* integrated in that locus. Another study shows the possible conserved sites of insertion in strains of *M. tuberculosis* with less than seven copies (45); other hot-spots have also been shown such as the IS*1547* (*iplA and iplB* loci) (40, 41), the phospholipase C regions (locus *plcABC* and *plcD* gene) (60, 131), members of the PPE gene family (84) and the origin of replication (*oriC*) (67, 126).

The locations of the IS*6110* along the genome are not equally distributed. This sequence was found to be inserted more often in some genome regions, whereas others lacked it (21). Nearly up to the first 800 nucleotides from the *oriC* fail in carrying copies of IS*6110* in the strain H37Rv. Besides, IS*6110* was otherwise located more or less randomly along the rest of the genome (Figure 6). The conclusion was that this part of the genome could be more abundant in essential genes. This result also was seen when studies of other strains were accomplished (3, 8, 108).

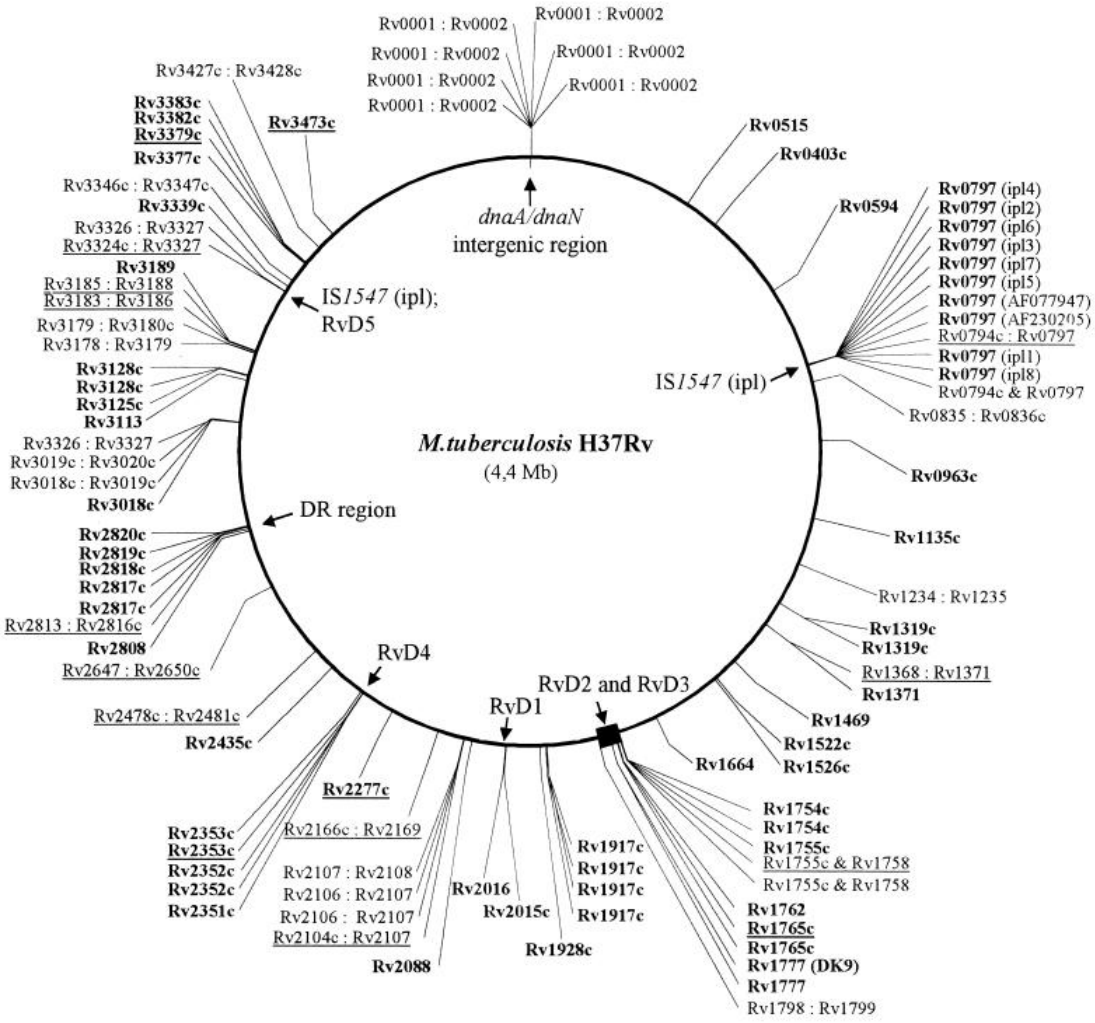


Figure 6: Example of locations of IS*6110* relative to ORFs of the *M. tuberculosis* H37Rv. The ORF numbers of the disrupted genes are indicated in bold, while the intergenic insertions are represented by two ORF numbers separated by a colon. The *M. tuberculosis* H37Rv insertion sites are underlined. Adapted from (107).

- **IS*6110* mediates genome plasticity of members of the MTBC**

In general, many mechanisms can be related to changes in the bacterial genomes, being those mediated by ISs one of the most relevant and better studied. It was considered that among 5 to 15% of spontaneous mutations in the bacterial genomes were due to changes in the IS locations. The more common mechanism used by an insertion sequence to move along genome is transposition. Another observed mechanism is recombination between two elements (83). For example, the deletion of the region RvD2 from *M. tuberculosis* H37Rv genome relative to *M. bovis* is due to the homologous recombination between two closed copies of IS*6110* (14, 39, 52). All those changes could be a risk to the bacteria's genomes integrity, being the carriage of mobile IS either a potential enemy with deadly influence on the bacterial fitness or a helpful ally contributing to the improvement of that fitness.

In summary, many events have been associated with IS*6110*, such as participation in adaptation to a particular host (94), activation of genes during infection (105), participation in the evolution as an epidemiological marker as we mention above (129) and activation of downstream genes with an activity promoter orientation-dependent (3, 105, 114).

The promoter activity of IS6110 will be widely studied in chapter 2.

IV.The phylogeny and spoligotype of *M. tuberculosis*

TB researchers are relying on genomics and DNA sequence-based approaches to study the evolution and population genetics of *M. tuberculosis*. The analysis of synonymous single-nucleotide polymorphisms (SNPs) and the study of genomic deletions have led to the development of a phylogenetic structure that allows understanding the relationship and evolution of members of the MTBC.

The Phylogeny of *M. tuberculosis*

Silent mutations in housekeeping genes were the first candidates to be selected as evolutionary markers. Based on the combination of two polymorphisms that occur at high frequency in the genes encoding catalase-peroxidase (*katG*) and subunit A of gyrase (*gyrA*), Sreevatsan *et al.* proposed the first classification of *M. tuberculosis* in 1997. Three Principal Genetic Groups (PGG 1-3) were identified (Figure 7). Group 1 or PGG1 is allied with *M. bovis, M. microtti and M. africanum* which is evolutionary closer to the common ancestor (116).

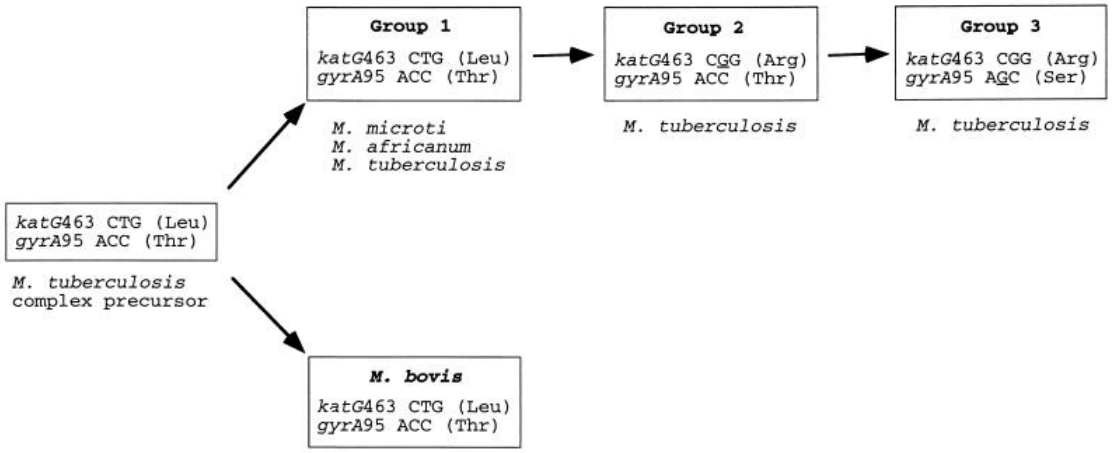


Figure 7: Broad evolutionary scenario for *M. tuberculosis* complex. Adapted from (116)

To group *M. tuberculosis*, Baker *et al*. divided *M. tuberculosis* into four main lineages (numbered I to IV) with the analysis of silent nucleotide sequence (sSNP) variation for seven genes associated with antimicrobial drug resistance. The *M. bovis* isolates formed an additional lineage (6) (Figure 8A). This phylogenetic structure is in concordance with PGG 1-3 of Sreevatsan *et al*. (116).

Gagneux *et al*. screened phylogenetically informative large sequences of polymorphisms (LSPs) in a global sample of 875 *M. tuberculosis* clinical isolates from 80 countries (47). All isolates belonged to one of six major lineages (Figure 8B), including two lineages that are traditionally referred to as *M. africanum*. The other four lineages were congruent with the four lineages defined by Baker *et al*.(6). These observations show that mycobacterial lineages are adapted to particular human populations.

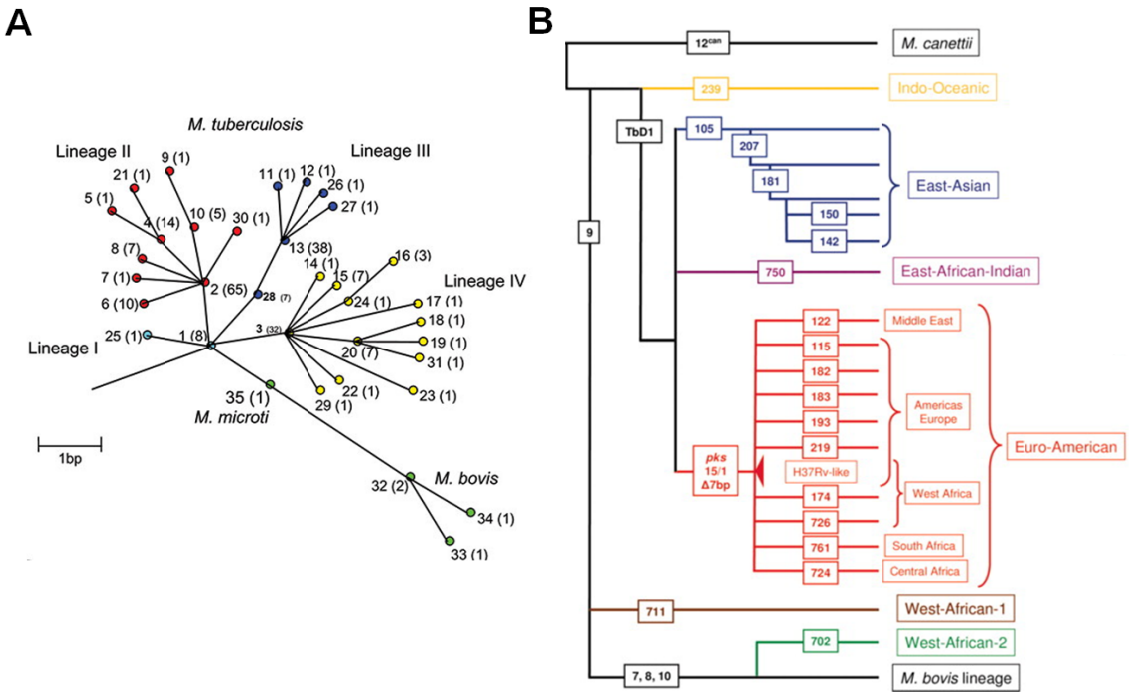


Figure 8: Two global phylogenic trees of *M. tuberculosis* complex. (A) Phylogenetic tree of *M. tuberculosis* and *M. bovis* based on 37 sSNP in 225 isolates described by Baker *et al.* Adapted from (6). **(B)** Gagneaux's global phylogeny for *M. tuberculosis* based on LSPs; the names of the lineage-defining LSPs or regions of difference are shown in rectangles. Adapted from (47).

Gutaker *et al.* using over 230 sSNP examined the phylogenetic relationship of *M. tuberculosis* isolates recovered from geographically distinct populations and found 8 distinct *M. tuberculosis* clusters (56). More recently, analysis of 5069 *M. tuberculosis* clinical isolates retrieved from four population-based studies, using the 36 most informative sSNP (from the 230 sSNP previously reported), as well as selected nsSNP and intergenic SNP, found a similar phylogenetic framework of 8 clusters (I to VIII), plus an additional cluster (II.A) (totaling IX clusters) (55) (Figure 9A).

These results showed a nonrandom relationship between genetic cluster and the 3 principal genetic groups of species. PGG1 strains were assigned to clusters I and II, PGG2 organism were assigned to clusters III-VI and PGG3 isolates were assigned to clusters VII and VIII. Cluster II.A included PGG1 (44%) and PGG2 (56%). Moreover, there is a correlation between sSNP lineages (6) and sSNP clusters (55).

Finally, Filliol *et al.* used a very similar approach to the study by Gutaker *et al*. to identify 159 sSNP and screened 212 isolates of *M. tuberculosis* and *M. bovis* and identified 7 phylogenetically distinct SNP cluster groups (SCGs), one SCG being specific to *M. bovis* (totaling 10 lineages) (43) (Figure 9B). The SCGs were strongly associated with the geographical origin of the *M. tuberculosis* samples and the birthplace of the human hosts. Again, six of the nine *M. tuberculosis* lineages belonged to the closely related PGG2 and 3, whereas the other three corresponded to the lineages defined by the other studies explained previously (6, 47).

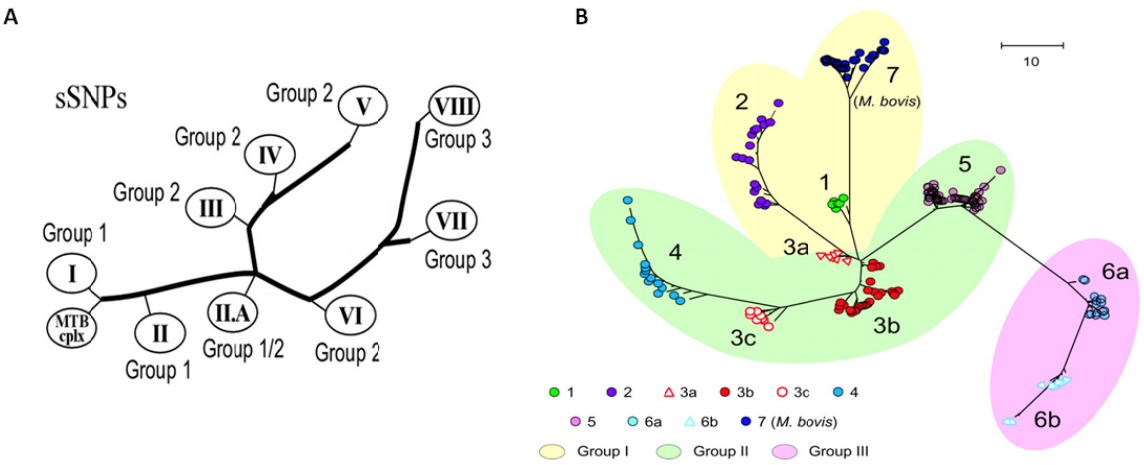


Figure 9: Comparison of two global phylogenetic trees of *M. tuberculosis* based on sSNP. (A) sSNP cluster groups based on 36 sSNP. Adapted from (55) **(B)** sSNP cluster groups (SCGs) based on 159 sSNP. Adapted from (43). The Principal Genetic Groups described by Sreevatsan *et al*. (116) are included.

In summary, applying LSPs or SNPs to the analysis of globally sampled collection of *M. tuberculosis* isolates has led to the same conclusion: the population structure of *M. tuberculosis* consists of four main strain lineages, plus two lineages comprising strains traditionally referred to as *M. africanum*.

Spoligotype of *M. tuberculosis*

Based on the genotype technique spoligotyping, we can group the global and local geographical structures of MTBC populations in families (16). Generally, a strain family can be described as a group of isolates that share specific biomarkers or properties indicative of a recent ancestor. MTBC strains contain a distinct chromosomal region consisting of multiple 36-bp direct repeats (DRs) interspersed by unique spacer DNA sequences (35 to 41 bp). Spoligotyping is based in the detection of 43 interspersed spacer sequences. Membranes spoted with 43 synthetic oligonucleotides are hybridized with labeled PCR-amplified DR region resulting in a pattern that can be detected by chemiluminiscence. The results are highly reproducible, and the binary (present/absent) data generated can be easily interpreted and computcrized and are amenable to intralaboratory comparisons (79). SpolDB4, the international spoligotyping database, contains 1939 different spoligotypes (ST), representing a total of 39,295 *M. tuberculosis* strains from 122 countries, and these are organized into 22 lineages/sub-lineages (16). Among these families we can enhance the Beijing family, the Central-Asian (CAS) family with two principal sublineages (CAS1 and CAS2), Zero, whose name comes from the absence of IS*6110* insertion element (“Zero copy”), MANU (sub-divided into MANU1 to MANU3), East-African-Indian (EAI) family, Haarlem family (H) with 4 sub lineages (H1-H4), Latin-American-Mediterranean (LAM), T family (modern TB strains) which is stratified into 5 sub-clades (T1-T5) and X (X1-X3 sub-lineages) which is well characterized as IS*6110* low-banding family.

- **Variation in spoligotype among the IX genetic cluster groups and the SCGs.**

Analysis of the relationship between the SNP-based phylogenetic trees (17, 55, 79) showed that some spoligotype clades were phylogenetically accurate with the SNP-derived phylogenetic framework trees while others were less so. For example, the spoligotype-defined Beijing clade was exclusively present in SCG-2 and exclusive to cluster II of the IX cluster groups described by Gutaker (Figure 10).

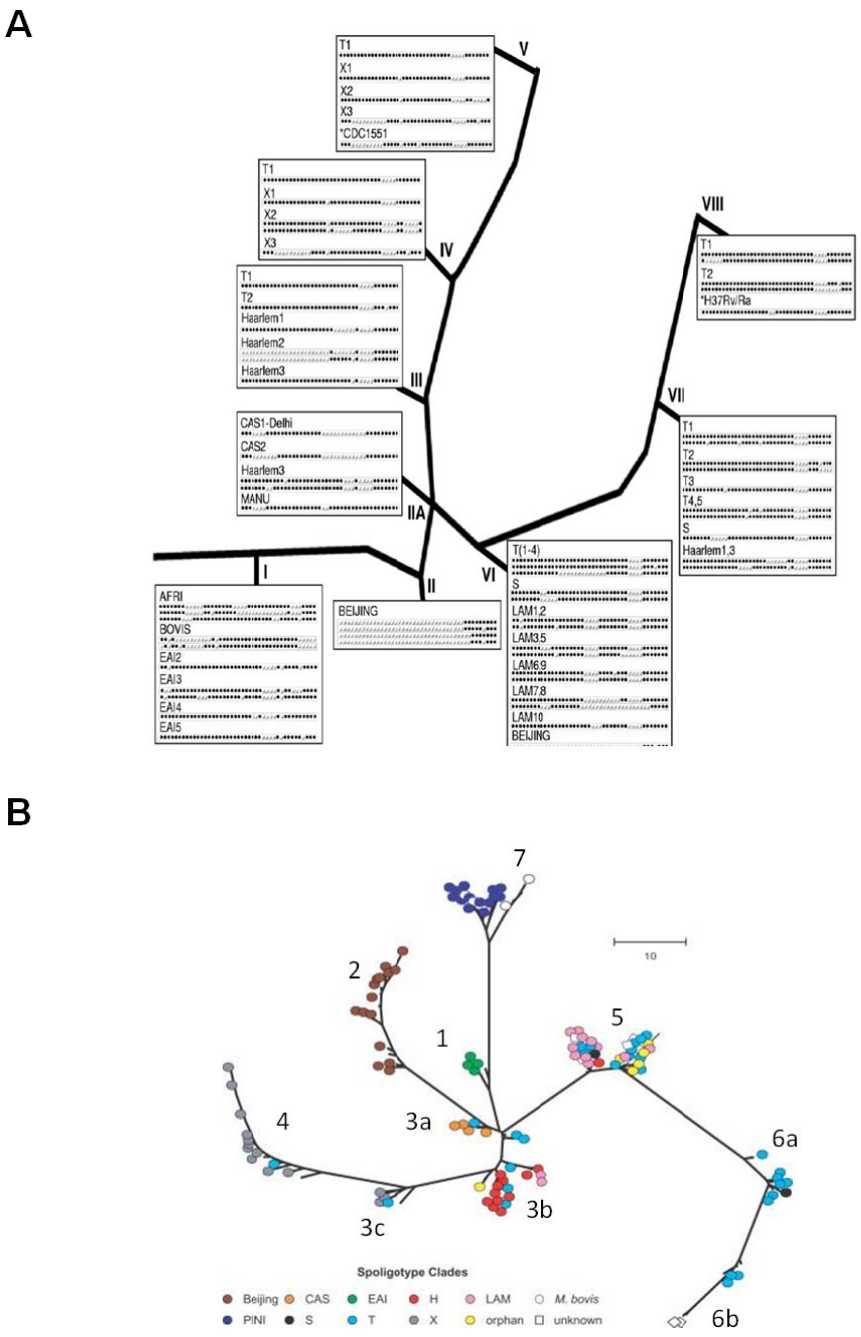


Figure 10: Distribution of the spoligotype clades on the SNP-based phylogeny. **(A)** Spoligotype patterns superimposed on the SNP-derived phylogentetic tree based on the IX clusters. Adapted from (79). **(B)** Distribution of the spoligotype patterns among the ten cluster groups (SCGs). Isolates are indicated by a dot, which is color coded according to the spoligotype clade assignment. Adapted from (43).

Although there are few studies showing a clear association between lineage and transmission capacity, it is now clear that the M. tuberculosis Beijing family globally has shown high transmissibility and virulence capacity more than any other lineage.

The *M. tuberculosis* Beijing family

The *M. tuberculosis* Beijing genotype family, also named East Asian Lineage (47), constitutes a homogeneous group of strains sharing a closely related IS*6110* RFLP patterns containing a high number of bands, identical spoligotyiping (deletion of spacers 1 to 34 in the Direct Repeat region) (130), an insertion of IS*6110* between *dnaA-dnaN* genes (called insertion A1) (67) and one or two IS*6110* copies in a DNA region called NTF locus (66, 67). The insertion of IS*6110* in these regions subdivide Beijing genotype in *ancestral or atypical* and *modern or typical* being the modern sublineage which contains at least one copy of IS*6110* upstream of the NTF region (88).

Moreover, the deletion of the regions of difference (RD) 105 and RD207, which are LSP markers, are present in all Beijing strains (124) (detailed in chapter 1). A few outliers that do not meet all of the above-suggested criteria for identification have also been described, as strains with “Beijing-like” spoligo patterns (25) (lacking one or more of the last nine spacers) and strains that harbor the Beijing genotype and the A1 insertion but which differ significantly in their IS*6110* banding patterns from the regular Beijing strains (67).

Strains of Beijing genotype were first described in China and neighbouring countries in 1995. The name comes from the finding that more than 80% of strains from the Beijing area were of this type (130). However, currently they are widespread in many regions of the world and frequently cause epidemic outbreaks (10, 19, 42, 50). Different studies have indicated that one-third of global TB cases is caused by Beijing family strains assigning this lineage to one of the most successful mycobacterial families in terms of morbidity and mortality (10). In some areas, such as Vietnam, Cuba, and Estonia, Beijing strains were found to be strongly associated with drug resistance (4, 12, 30, 65). This is consistent with the prolific spread of the (multidrug)-resistant strain W and its variants in the 1990s in Northern America (11). This strain and its variants exhibit highly similar IS*6110* RFLP patterns and an identical spoligo pattern. It is now known that the W family and the Beijing family represent the same genotype which was concurrently identified in North America (12) and Asia (130). Both groups of strains are referred as Beijing family (Beijing/W or W-Beijing). A second group of strains with distantly related IS*6110* RFLP profiles have been determined to belong to the Beijing lineage by other molecular techniques and are referred to as *ancestral* strains (88). They share a common predecessor with Beijing family strains, as illustrated by the extensive shared molecular characteristics, including chromosomal insertions and deletions, and single nucleotide polymorphisms (35).

The widely distributed association of drug resistance and the Beijing genotype suggest that these strains may have a particular propensity for acquiring drug resistance (51). It was reported that Beijing strains carry mutations in putative mutators genes, and this may explain a higher adaptability of these bacteria to stress conditions such as exposure to antituberculosis drugs and the hostile intracellular environment (35).

A number of selective advantages have been associated with its success including a lower efficacy of BCG for Beijing strains (74, 130), higher virulence (74, 78), ability to induce a differential immune response (33, 77), higher transmissibility (10) and enhanced capacity to grow in human macrophages and monocytes (71, 137) and an enhanced capacity to acquire drug resistance (1, 11, 12). However, unidentified virulence factors and the modulation of specific host responses are not yet thoroughly investigated.

References

1. **Agerton, T. B., S. E. Valway, R. J. Blinkhorn, K. L. Shilkret, R. Reves, W. W. Schluter, B. Gore, C. J. Pozsik, B. B. Plikaytis, C. Woodley, and I. M. Onorato.** 1999. Spread of strain W, a highly drug-resistant strain of Mycobacterium tuberculosis, across the United States. Clin Infect Dis **29:**85-92; discussion 93-5.
2. **Alexander, K. A., P. N. Laver, A. L. Michel, M. Williams, P. D. van Helden, R. M. Warren, and N. C. Gey van Pittius.** Novel Mycobacterium tuberculosis complex pathogen, M. mungi. Emerg Infect Dis **16:**1296-9.
3. **Alonso, H., J. I. Aguilo, S. Samper, J. A. Caminero, M. I. Campos-Herrero, B. Gicquel, R. Brosch, C. Martin, and I. Otal.** Deciphering the role of IS6110 in a highly transmissible Mycobacterium tuberculosis Beijing strain, GC1237. Tuberculosis (Edinb).
4. **Anh, D. D., M. W. Borgdorff, L. N. Van, N. T. Lan, T. van Gorkom, K. Kremer, and D. van Soolingen.** 2000. Mycobacterium tuberculosis Beijing genotype emerging in Vietnam. Emerg Infect Dis **6:**302-5.
5. **Astarie-Dequeker, C., L. Le Guyader, W. Malaga, F. K. Seaphanh, C. Chalut, A. Lopez, and C. Guilhot.** 2009. Phthiocerol dimycocerosates of M. tuberculosis participate in macrophage invasion by inducing changes in the organization of plasma membrane lipids. PLoS Pathog **5:**e1000289.
6. **Baker, L., T. Brown, M. C. Maiden, and F. Drobniewski.** 2004. Silent nucleotide polymorphisms and a phylogeny for Mycobacterium tuberculosis. Emerg Infect Dis **10:**1568-77.
7. **Barrera, L.** 2007. The basics of clinical bacteriology. *In* J. C. Palomino, S. Cardoso-Leao, and V. Ritacco (ed.), Tuberculosis. From basic science to patient care.
8. **Beggs, M. L., K. D. Eisenach, and M. D. Cave.** 2000. Mapping of IS6110 insertion sites in two epidemic strains of Mycobacterium tuberculosis. J Clin Microbiol **38:**2923-8.
9. **Behr, M. A.** 2002. BCG--different strains, different vaccines? Lancet Infect Dis **2:**86-92.
10. **Bifani, P. J., B. Mathema, N. E. Kurepina, and B. N. Kreiswirth.** 2002. Global dissemination of the Mycobacterium tuberculosis W-Beijing family strains. Trends Microbiol **10:**45-52.
11. **Bifani, P. J., B. Mathema, Z. Liu, S. L. Moghazeh, B. Shopsin, B. Tempalski, J. Driscol, R. Frothingham, J. M. Musser, P. Alcabes, and B. N. Kreiswirth.** 1999. Identification of a W variant outbreak of Mycobacterium tuberculosis via population-based molecular epidemiology. JAMA **282:**2321-7.
12. **Bifani, P. J., B. B. Plikaytis, V. Kapur, K. Stockbauer, X. Pan, M. L. Lutfey, S. L. Moghazeh, W. Eisner, T. M. Daniel, M. H. Kaplan, J. T. Crawford, J. M. Musser, and B. N. Kreiswirth.** 1996. Origin and interstate spread of a New York City multidrug-resistant Mycobacterium tuberculosis clone family. Jama **275:**452-7.
13. **Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen, and S. T. Cole.** 2002. A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci U S A **99:**3684-9.
14. **Brosch, R., W. J. Philipp, E. Stavropoulos, M. J. Colston, S. T. Cole, and S. V. Gordon.** 1999. Genomic analysis reveals variation between Mycobacterium tuberculosis H37Rv and the attenuated M. tuberculosis H37Ra strain. Infect Immun **67:**5768-74.
15. **Brosch, R. G., S.V.; Eiglmeier, K.; Garnier, T.; Tekaia, F.; Yeramian, E. & Cole, S.T. .** 2000. Genomics, biology, and evolution of the *Mycobacterium tuberculosis* complex. *Molecular Genetics of Mycobacteria*, **ISBN 1-55581-191-4:**19-36.
16. **Brudey, K., J. R. Driscoll, L. Rigouts, W. M. Prodinger, A. Gori, S. A. Al-Hajoj, C. Allix, L. Aristimuno, J. Arora, V. Baumanis, L. Binder, P. Cafrune, A. Cataldi, S. Cheong, R. Diel, C. Ellermeier, J. T. Evans, M. Fauville-Dufaux, S. Ferdinand, D. Garcia de Viedma, C. Garzelli, L. Gazzola, H. M. Gomes, M. C. Guttierez, P. M. Hawkey, P. D. van Helden, G. V. Kadival, B. N. Kreiswirth, K. Kremer, M. Kubin, S.**

P. Kulkarni, B. Liens, T. Lillebaek, M. L. Ho, C. Martin, I. Mokrousov, O. Narvskaia, Y. F. Ngeow, L. Naumann, S. Niemann, I. Parwati, Z. Rahim, V. Rasolofo-Razanamparany, T. Rasolonavalona, M. L. Rossetti, S. Rusch-Gerdes, A. Sajduda, S. Samper, I. G. Shemyakin, U. B. Singh, A. Somoskovi, R. A. Skuce, D. van Soolingen, E. M. Streicher, P. N. Suffys, E. Tortoli, T. Tracevska, V. Vincent, T. C. Victor, R. M. Warren, S. F. Yap, K. Zaman, F. Portaels, N. Rastogi, and C. Sola. 2006. Mycobacterium tuberculosis complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol **6:**23.

17. **Brudey, K., I. Filliol, S. Ferdinand, V. Guernier, P. Duval, B. Maubert, C. Sola, and N. Rastogi.** 2006. Long-term population-based genotyping study of Mycobacterium tuberculosis complex isolates in the French departments of the Americas. J Clin Microbiol **44:**183-91.
18. **Camacho, L. R., D. Ensergueix, E. Perez, B. Gicquel, and C. Guilhot.** 1999. Identification of a virulence gene cluster of Mycobacterium tuberculosis by signature-tagged transposon mutagenesis. Mol Microbiol **34:**257-67.
19. **Caminero, J. A., M. J. Pena, M. I. Campos-Herrero, J. C. Rodriguez, I. Garcia, P. Cabrera, C. Lafoz, S. Samper, H. Takiff, O. Afonso, J. M. Pavon, M. J. Torres, D. van Soolingen, D. A. Enarson, and C. Martin.** 2001. Epidemiological evidence of the spread of a Mycobacterium tuberculosis strain of the Beijing genotype on Gran Canaria Island. Am J Respir Crit Care Med **164:**1165-70.
20. **Cole, S. T.** 1999. Learning from the genome sequence of Mycobacterium tuberculosis H37Rv. FEBS Lett **452:**7-10.
21. **Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, 3rd, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell.** 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature **393:**537-44.
22. **Converse, S. E., J. D. Mougous, M. D. Leavell, J. A. Leary, C. R. Bertozzi, and J. S. Cox.** 2003. MmpL8 is required for sulfolipid-1 biosynthesis and Mycobacterium tuberculosis virulence. Proc Natl Acad Sci U S A **100:**6121-6.
23. **Corbett, E. L., C. J. Watt, N. Walker, D. Maher, B. G. Williams, M. C. Raviglione, and C. Dye.** 2003. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. Arch Intern Med **163:**1009-21.
24. **Cox, J. S., B. Chen, M. McNeil, and W. R. Jacobs, Jr.** 1999. Complex lipid determines tissue-specific replication of Mycobacterium tuberculosis in mice. Nature **402:**79-83.
25. **Chan, M. Y., M. Borgdorff, C. W. Yip, P. E. de Haas, W. S. Wong, K. M. Kam, and D. Van Soolingen.** 2001. Seventy percent of the Mycobacterium tuberculosis isolates in Hong Kong represent the Beijing genotype. Epidemiol Infect **127:**169-71.
26. **Daffe, M., and P. Draper.** 1998. The envelope layers of mycobacteria with reference to their pathogenicity. Adv Microb Physiol **39:**131-203.
27. **Daniel, T. M.** 2006. The history of tuberculosis. Respir Med **100:**1862-70.
28. **Demangel, C., T. Garnier, I. Rosenkrands, and S. T. Cole.** 2005. Differential effects of prior exposure to environmental mycobacteria on vaccination with Mycobacterium bovis BCG or a recombinant BCG strain expressing RD1 antigens. Infect Immun **73:**2190-6.
29. **Deretic, V., S. Singh, S. Master, J. Harris, E. Roberts, G. Kyei, A. Davis, S. de Haro, J. Naylor, H. H. Lee, and I. Vergne.** 2006. Mycobacterium tuberculosis inhibition of phagolysosome biogenesis and autophagy as a host defence mechanism. Cell Microbiol **8:**719-27.
30. **Diaz, R., K. Kremer, P. E. de Haas, R. I. Gomez, A. Marrero, J. A. Valdivia, J. D. van Embden, and D. van Soolingen.** 1998. Molecular epidemiology of tuberculosis in

Cuba outside of Havana, July 1994-June 1995: utility of spoligotyping versus IS6110 restriction fragment length polymorphism. Int J Tuberc Lung Dis **2:**743-50.

31. **Domenech, P., M. B. Reed, C. S. Dowd, C. Manca, G. Kaplan, and C. E. Barry, 3rd.** 2004. The role of MmpL8 in sulfatide biogenesis and virulence of Mycobacterium tuberculosis. J Biol Chem **279:**21257-65.
32. **Donoghue, H. D., M. Spigelman, C. L. Greenblatt, G. Lev-Maor, G. K. Bar-Gal, C. Matheson, K. Vernon, A. G. Nerlich, and A. R. Zink.** 2004. Tuberculosis: from prehistory to Robert Koch, as revealed by ancient DNA. Lancet Infect Dis **4:**584-92.
33. **Dormans, J., M. Burger, D. Aguilar, R. Hernandez-Pando, K. Kremer, P. Roholl, S. M. Arend, and D. van Soolingen.** 2004. Correlation of virulence, lung pathology, bacterial load and delayed type hypersensitivity responses after infection with different Mycobacterium tuberculosis genotypes in a BALB/c mouse model. Clin Exp Immunol **137:**460-8.
34. **Dziadek, J., A. Sajduda, J. W. Dale, and J. J. McFadden.** 1998. IS990, a new species-specific insertion-sequence-related element of the Mycobacterium tuberculosis complex. Microbiology **144 (Pt 12):**3407-12.
35. **Ebrahimi-Rad, M., P. Bifani, C. Martin, K. Kremer, S. Samper, J. Rauzier, B. Kreiswirth, J. Blazquez, M. Jouan, D. van Soolingen, and B. Gicquel.** 2003. Mutations in putative mutator genes of Mycobacterium tuberculosis strains of the W-Beijing family. Emerg Infect Dis **9:**838-45.
36. **Ehlers, M. R., and M. Daffe.** 1998. Interactions between Mycobacterium tuberculosis and host cells: are mycobacterial sugars the key? Trends Microbiol **6:**328-35.
37. **Ehrt, S., and D. Schnappinger.** 2009. Mycobacterial survival strategies in the phagosome: defence against host stresses. Cell Microbiol **11:**1170-8.
38. **Ernst, J. D.** 1998. Macrophage receptors for Mycobacterium tuberculosis. Infect Immun **66:**1277-81.
39. **Fang, Z., C. Doig, D. T. Kenna, N. Smittipat, P. Palittapongarnpim, B. Watt, and K. J. Forbes.** 1999. IS6110-mediated deletions of wild-type chromosomes of Mycobacterium tuberculosis. J Bacteriol **181:**1014-20.
40. **Fang, Z., and K. J. Forbes.** 1997. A Mycobacterium tuberculosis IS6110 preferential locus (ipl) for insertion into the genome. J Clin Microbiol **35:**479-81.
41. **Fang, Z., D. T. Kenna, C. Doig, D. N. Smittipat, P. Palittapongarnpim, B. Watt, and K. J. Forbes.** 2001. Molecular evidence for independent occurrence of IS6110 insertions at the same sites of the genome of Mycobacterium tuberculosis in different clinical isolates. J Bacteriol **183:**5279-84.
42. **Filliol, I., J. R. Driscoll, D. van Soolingen, B. N. Kreiswirth, K. Kremer, G. Valetudie, D. A. Dang, R. Barlow, D. Banerjee, P. J. Bifani, K. Brudey, A. Cataldi, R. C. Cooksey, D. V. Cousins, J. W. Dale, O. A. Dellagostin, F. Drobniewski, G. Engelmann, S. Ferdinand, D. Gascoyne-Binzi, M. Gordon, M. C. Gutierrez, W. H. Haas, H. Heersma, E. Kassa-Kelembho, M. L. Ho, A. Makristathis, C. Mammina, G. Martin, P. Mostrom, I. Mokrousov, V. Narbonne, O. Narvskaya, A. Nastasi, S. N. Niobe-Eyangoh, J. W. Pape, V. Rasolofo-Razanamparany, M. Ridell, M. L. Rossetti, F. Stauffer, P. N. Suffys, H. Takiff, J. Texier-Maugein, V. Vincent, J. H. de Waard, C. Sola, and N. Rastogi.** 2003. Snapshot of moving and expanding clones of Mycobacterium tuberculosis and their global distribution assessed by spoligotyping in an international study. J Clin Microbiol **41:**1963-70.
43. **Filliol, I., A. S. Motiwala, M. Cavatore, W. Qi, M. H. Hazbon, M. Bobadilla del Valle, J. Fyfe, L. Garcia-Garcia, N. Rastogi, C. Sola, T. Zozio, M. I. Guerrero, C. I. Leon, J. Crabtree, S. Angiuoli, K. D. Eisenach, R. Durmaz, M. L. Joloba, A. Rendon, J. Sifuentes-Osornio, A. Ponce de Leon, M. D. Cave, R. Fleischmann, T. S. Whittam, and D. Alland.** 2006. Global phylogeny of Mycobacterium tuberculosis based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. J Bacteriol **188:**759-72.

44. **Fine, P. E.** 1995. Variation in protection by BCG: implications of and for heterologous immunity. Lancet **346:**1339-45.
45. **Fomukong, N., M. Beggs, H. el Hajj, G. Templeton, K. Eisenach, and M. D. Cave.** 1997. Differences in the prevalence of IS6110 insertion sites in Mycobacterium tuberculosis strains: low and high copy number of IS6110. Tuber Lung Dis **78:**109-16.
46. **Fratti, R. A., J. Chua, I. Vergne, and V. Deretic.** 2003. Mycobacterium tuberculosis glycosylated phosphatidylinositol causes phagosome maturation arrest. Proc Natl Acad Sci U S A **100:**5437-42.
47. **Gagneux, S., K. DeRiemer, T. Van, M. Kato-Maeda, B. C. de Jong, S. Narayanan, M. Nicol, S. Niemann, K. Kremer, M. C. Gutierrez, M. Hilty, P. C. Hopewell, and P. M. Small.** 2006. Variable host-pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A **103:**2869-73.
48. **Gandhi, N. R., A. Moll, A. W. Sturm, R. Pawinski, T. Govender, U. Lalloo, K. Zeller, J. Andrews, and G. Friedland.** 2006. Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. Lancet **368:**1575-80.
49. **Geerligs, W. A., R. Van Altena, W. C. M. De Lange, D. Van Soolingen, and T. S. Van Der Werf.** 2000. Multidrug-resistant tuberculosis: long-term treatment outcome in the Netherlands. Int J Tuberc Lung Dis **4:**758-64.
50. **Glynn, J. R., A. C. Crampin, H. Traore, M. D. Yates, F. D. Mwaungulu, B. M. Ngwira, S. D. Chaguluka, D. T. Mwafulirwa, S. Floyd, C. Murphy, F. A. Drobniewski, and P. E. Fine.** 2005. Mycobacterium tuberculosis Beijing genotype, northern Malawi. Emerg Infect Dis **11:**150-3.
51. **Glynn, J. R., J. Whiteley, P. J. Bifani, K. Kremer, and D. van Soolingen.** 2002. Worldwide occurrence of Beijing/W strains of Mycobacterium tuberculosis: a systematic review. Emerg Infect Dis **8:**843-9.
52. **Gordon, S. V., R. Brosch, A. Billault, T. Garnier, K. Eiglmeier, and S. T. Cole.** 1999. Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. Mol Microbiol **32:**643-55.
53. **Gordon, S. V., B. Heym, J. Parkhill, B. Barrell, and S. T. Cole.** 1999. New insertion sequences and a novel repeated sequence in the genome of Mycobacterium tuberculosis H37Rv. Microbiology **145 (Pt 4):**881-92.
54. **Guiard, J., A. Collmann, L. F. Garcia-Alles, L. Mourey, T. Brando, L. Mori, M. Gilleron, J. Prandi, G. De Libero, and G. Puzo.** 2009. Fatty acyl structures of mycobacterium tuberculosis sulfoglycolipid govern T cell response. J Immunol **182:**7030-7.
55. **Gutacker, M. M., B. Mathema, H. Soini, E. Shashkina, B. N. Kreiswirth, E. A. Graviss, and J. M. Musser.** 2006. Single-nucleotide polymorphism-based population genetic analysis of Mycobacterium tuberculosis strains from 4 geographic sites. J Infect Dis **193:**121-8.
56. **Gutacker, M. M., J. C. Smoot, C. A. Migliaccio, S. M. Ricklefs, S. Hua, D. V. Cousins, E. A. Graviss, E. Shashkina, B. N. Kreiswirth, and J. M. Musser.** 2002. Genome-wide analysis of synonymous single nucleotide polymorphisms in Mycobacterium tuberculosis complex organisms: resolution of genetic relationships among closely related microbial strains. Genetics **162:**1533-43.
57. **Gutierrez, M. G., S. S. Master, S. B. Singh, G. A. Taylor, M. I. Colombo, and V. Deretic.** 2004. Autophagy is a defense mechanism inhibiting BCG and Mycobacterium tuberculosis survival in infected macrophages. Cell **119:**753-66.
58. **Hermans, P. W., D. van Soolingen, E. M. Bik, P. E. de Haas, J. W. Dale, and J. D. van Embden.** 1991. Insertion element IS987 from Mycobacterium bovis BCG is located in a hot-spot integration region for insertion elements in Mycobacterium tuberculosis complex strains. Infect Immun **59:**2695-705.
59. **Herzog, H.** 1998. History of tuberculosis. Respiration **65:**5-15.
60. **Ioerger, T. R., Y. Feng, K. Ganesula, X. Chen, K. M. Dobos, S. Fortune, W. R. Jacobs, Jr., V. Mizrahi, T. Parish, E. Rubin, C. Sassetti, and J. C. Sacchettini.**

Variation among genome sequences of H37Rv strains of Mycobacterium tuberculosis from multiple laboratories. J Bacteriol **192:**3645-53.

61. **Jackson, M., G. Stadthagen, and B. Gicquel.** 2007. Long-chain multiple methyl-branched fatty acid-containing lipids of Mycobacterium tuberculosis: biosynthesis, transport, regulation and biological activities. Tuberculosis (Edinb) **87:**78-86.
62. **Kang, B. K., and L. S. Schlesinger.** 1998. Characterization of mannose receptor-dependent phagocytosis mediated by Mycobacterium tuberculosis lipoarabinomannan. Infect Immun **66:**2769-77.
63. **Kaufmann, S., and P. v. Helden (ed.).** 2008. Handbook of Tuberculosis.
64. **Kaufmann, S. H.** 2005. Robert Koch, the Nobel Prize, and the ongoing threat of tuberculosis. N Engl J Med **353:**2423-6.
65. **Kruuner, A., S. E. Hoffner, H. Sillastu, M. Danilovits, K. Levina, S. B. Svenson, S. Ghebremichael, T. Koivula, and G. Kallenius.** 2001. Spread of drug-resistant pulmonary tuberculosis in Estonia. J Clin Microbiol **39:**3339-45.
66. **Kurepina, N., E. Likhoshvay, E. Shashkina, B. Mathema, K. Kremer, D. van Soolingen, P. Bifani, and B. N. Kreiswirth.** 2005. Targeted hybridization of IS6110 fingerprints identifies the W-Beijing Mycobacterium tuberculosis strains among clinical isolates. J Clin Microbiol **43:**2148-54.
67. **Kurepina, N. E., S. Sreevatsan, B. B. Plikaytis, P. J. Bifani, N. D. Connell, R. J. Donnelly, D. van Sooligen, J. M. Musser, and B. N. Kreiswirth.** 1998. Characterization of the phylogenetic distribution and chromosomal insertion sites of five IS6110 elements in Mycobacterium tuberculosis: non-random integration in the dnaA-dnaN region. Tuber Lung Dis **79:**31-42.
68. **Lee, J., M. Hartman, and H. Kornfeld.** 2009. Macrophage apoptosis in tuberculosis. Yonsei Med J **50:**1-11.
69. **Lee, K. S., V. S. Dubey, P. E. Kolattukudy, C. H. Song, A. R. Shin, S. B. Jung, C. S. Yang, S. Y. Kim, E. K. Jo, J. K. Park, and H. J. Kim.** 2007. Diacyltrehalose of Mycobacterium tuberculosis inhibits lipopolysaccharide- and mycobacteria-induced proinflammatory cytokine production in human monocytic cells. FEMS Microbiol Lett **267:**121-8.
70. **Lew, J. M., A. Kapopoulou, L. M. Jones, and S. T. Cole.** TubercuList--10 years after. Tuberculosis (Edinb) **91:**1-7.
71. **Li, Q., C. C. Whalen, J. M. Albert, R. Larkin, L. Zukowski, M. D. Cave, and R. F. Silver.** 2002. Differences in rate and variability of intracellular growth of a panel of Mycobacterium tuberculosis clinical isolates within a human monocyte model. Infect Immun **70:**6489-93.
72. **Loewenberg, S.** India reports cases of totally drug-resistant tuberculosis. Lancet **379:**205.
73. **Lok, K. H., W. H. Benjamin, Jr., M. E. Kimerling, V. Pruitt, M. Lathan, J. Razeq, N. Hooper, W. Cronin, and N. E. Dunlap.** 2002. Molecular differentiation of Mycobacterium tuberculosis strains without IS6110 insertions. Emerg Infect Dis **8:**1310-3.
74. **Lopez, B., D. Aguilar, H. Orozco, M. Burger, C. Espitia, V. Ritacco, L. Barrera, K. Kremer, R. Hernandez-Pando, K. Huygen, and D. van Soolingen.** 2003. A marked difference in pathogenesis and immune response induced by different Mycobacterium tuberculosis genotypes. Clin Exp Immunol **133:**30-7.
75. **Mahillon, J., and M. Chandler.** 1998. Insertion sequences. Microbiol Mol Biol Rev **62:**725-74.
76. **Malik, Z. A., C. R. Thompson, S. Hashimi, B. Porter, S. S. Iyer, and D. J. Kusner.** 2003. Cutting edge: Mycobacterium tuberculosis blocks Ca2+ signaling and phagosome maturation in human macrophages via specific inhibition of sphingosine kinase. J Immunol **170:**2811-5.
77. **Manca, C., M. B. Reed, S. Freeman, B. Mathema, B. Kreiswirth, C. E. Barry, 3rd, and G. Kaplan.** 2004. Differential monocyte activation underlies strain-specific Mycobacterium tuberculosis pathogenesis. Infect Immun **72:**5511-4.

78. **Manca, C., L. Tsenova, A. Bergtold, S. Freeman, M. Tovey, J. M. Musser, C. E. Barry, 3rd, V. H. Freedman, and G. Kaplan.** 2001. Virulence of a Mycobacterium tuberculosis clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN-alpha /beta. Proc Natl Acad Sci U S A **98:**5752-7.
79. **Mathema, B., N. E. Kurepina, P. J. Bifani, and B. N. Kreiswirth.** 2006. Molecular epidemiology of tuberculosis: current insights. Clin Microbiol Rev **19:**658-85.
80. **Matsunaga, I., A. Bhatt, D. C. Young, T. Y. Cheng, S. J. Eyles, G. S. Besra, V. Briken, S. A. Porcelli, C. E. Costello, W. R. Jacobs, Jr., and D. B. Moody.** 2004. Mycobacterium tuberculosis pks12 produces a novel polyketide presented by CD1c to T cells. J Exp Med **200:**1559-69.
81. **McAdam, R. A., P. W. Hermans, D. van Soolingen, Z. F. Zainuddin, D. Catty, J. D. van Embden, and J. W. Dale.** 1990. Characterization of a Mycobacterium tuberculosis insertion sequence belonging to the IS3 family. Mol Microbiol **4:**1607-13.
82. **McAdam, R. A. Q., S. & Guilhot, C.** 2000. Mycobacterial Tansposons and their applications. *Molecular Genetics of Mycobacteria* **American Society for Microbiology Press, ISBN 1-55581-191-4, Washington DC, USA.:**69−84.
83. **McEvoy, C. R., A. A. Falmer, N. C. Gey van Pittius, T. C. Victor, P. D. van Helden, and R. M. Warren.** 2007. The role of IS6110 in the evolution of Mycobacterium tuberculosis. Tuberculosis (Edinb) **87:**393-404.
84. **McEvoy, C. R., R. M. Warren, P. D. van Helden, and N. C. Gey van Pittius.** 2009. Multiple, independent, identical IS6110 insertions in Mycobacterium tuberculosis PPE genes. Tuberculosis (Edinb) **89:**439-42.
85. **Mendiola, M. V., C. Martin, I. Otal, and B. Gicquel.** 1992. Analysis of the regions responsible for IS6110 RFLP in a single Mycobacterium tuberculosis strain. Res Microbiol **143:**767-72.
86. **Minnikin, D. E., L. Kremer, L. G. Dover, and G. S. Besra.** 2002. The methyl-branched fortifications of Mycobacterium tuberculosis. Chem Biol **9:**545-53.
87. **Mizrahi, V., and S. J. Andersen.** 1998. DNA repair in Mycobacterium tuberculosis. What have we learnt from the genome sequence? Mol Microbiol **29:**1331-9.
88. **Mokrousov, I., O. Narvskaya, T. Otten, A. Vyazovaya, E. Limeschenko, L. Steklova, and B. Vyshnevskyi.** 2002. Phylogenetic reconstruction within Mycobacterium tuberculosis Beijing genotype in northwestern Russia. Res Microbiol **153:**629-37.
89. **Mostowy, S., J. Inwald, S. Gordon, C. Martin, R. Warren, K. Kremer, D. Cousins, and M. A. Behr.** 2005. Revisiting the evolution of Mycobacterium bovis. J Bacteriol **187:**6386-95.
90. **Mueller, P., and J. Pieters.** 2006. Modulation of macrophage antimicrobial mechanisms by pathogenic mycobacteria. Immunobiology **211:**549-56.
91. **Nerlich, A. G., C. J. Haas, A. Zink, U. Szeimies, and H. G. Hagedorn.** 1997. Molecular evidence for tuberculosis in an ancient Egyptian mummy. Lancet **350:**1404.
92. **Onwueme, K. C., C. J. Vos, J. Zurita, J. A. Ferreras, and L. E. Quadri.** 2005. The dimycocerosate ester polyketide virulence factors of mycobacteria. Prog Lipid Res **44:**259-302.
93. **Ordway, D., M. Henao-Tamayo, M. Harton, G. Palanisamy, J. Troudt, C. Shanley, R. J. Basaraba, and I. M. Orme.** 2007. The hypervirulent Mycobacterium tuberculosis strain HN878 induces a potent TH1 response followed by rapid down-regulation. J Immunol **179:**522-31.
94. **Otal, I., A. B. Gomez, K. Kremer, P. de Haas, M. J. Garcia, C. Martin, and D. van Soolingen.** 2008. Mapping of IS6110 insertion sites in Mycobacterium bovis isolates in relation to adaptation from the animal to human host. Vet Microbiol **129:**333-41.
95. **Otal, I., C. Martin, V. Vincent-Levy-Frebault, D. Thierry, and B. Gicquel.** 1991. Restriction fragment length polymorphism analysis using IS6110 as an epidemiological marker in tuberculosis. J Clin Microbiol **29:**1252-4.

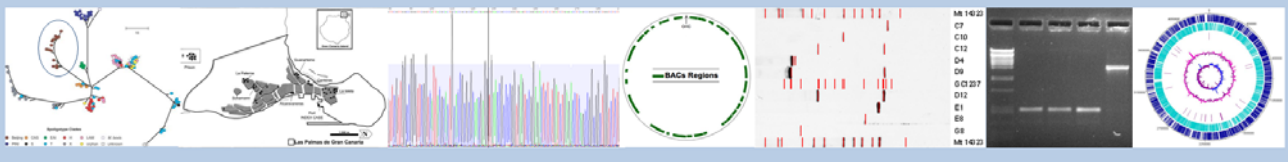
96. **Pethe, K., D. L. Swenson, S. Alonso, J. Anderson, C. Wang, and D. G. Russell.** 2004. Isolation of Mycobacterium tuberculosis mutants defective in the arrest of phagosome maturation. Proc Natl Acad Sci U S A **101:**13642-7.
97. **Reed, M. B., P. Domenech, C. Manca, H. Su, A. K. Barczak, B. N. Kreiswirth, G. Kaplan, and C. E. Barry, 3rd.** 2004. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. Nature **431:**84-7.
98. **Rohde, K., R. M. Yates, G. E. Purdy, and D. G. Russell.** 2007. Mycobacterium tuberculosis and the environment within the phagosome. Immunol Rev **219:**37-54.
99. **Rousseau, C., O. Neyrolles, Y. Bordat, S. Giroux, T. D. Sirakova, M. C. Prevost, P. E. Kolattukudy, B. Gicquel, and M. Jackson.** 2003. Deficiency in mycolipenate- and mycosanoate-derived acyltrehaloses enhances early interactions of Mycobacterium tuberculosis with host cells. Cell Microbiol **5:**405-15.
100. **Rousseau, C., N. Winter, E. Pivert, Y. Bordat, O. Neyrolles, P. Ave, M. Huerre, B. Gicquel, and M. Jackson.** 2004. Production of phthiocerol dimycocerosates protects Mycobacterium tuberculosis from the cidal activity of reactive nitrogen intermediates produced by macrophages and modulates the early immune response to infection. Cell Microbiol **6:**277-87.
101. **Russell, D. G.** 2001. Mycobacterium tuberculosis: here today, and here tomorrow. Nat Rev Mol Cell Biol **2:**569-77.
102. **Russell, D. G.** 2007. Who puts the tubercle in tuberculosis? Nat Rev Microbiol **5:**39-47.
103. **Russell, D. G., P. J. Cardona, M. J. Kim, S. Allain, and F. Altare.** 2009. Foamy macrophages and the progression of the human tuberculosis granuloma. Nat Immunol **10:**943-8.
104. **Ryan, F.** 1992. Tuberculosis: The greatest story never told. Swift Publishers, Worcestershire, Great Britain.
105. **Safi, H., P. F. Barnes, D. L. Lakey, H. Shams, B. Samten, R. Vankayalapati, and S. T. Howard.** 2004. IS6110 functions as a mobile, monocyte-activated promoter in Mycobacterium tuberculosis. Mol Microbiol **52:**999-1012.
106. **Salo, W. L., A. C. Aufderheide, J. Buikstra, and T. A. Holcomb.** 1994. Identification of Mycobacterium tuberculosis DNA in a pre-Columbian Peruvian mummy. Proc Natl Acad Sci U S A **91:**2091-4.
107. **Sampson, S., R. Warren, M. Richardson, G. van der Spuy, and P. van Helden.** 2001. IS6110 insertions in Mycobacterium tuberculosis: predominantly into coding regions. J Clin Microbiol **39:**3423-4.
108. **Sampson, S. L., R. M. Warren, M. Richardson, G. D. van der Spuy, and P. D. van Helden.** 1999. Disruption of coding regions by IS6110 insertion in Mycobacterium tuberculosis. Tuber Lung Dis **79:**349-59.
109. **Sankar, S., S. Kuppanan, B. Balakrishnan, and B. Nandagopal.** Analysis of sequence diversity among IS6110 sequence of Mycobacterium tuberculosis: possible implications for PCR based detection. Bioinformation **6:**283-5.
110. **Schlesinger, L. S.** 1993. Macrophage phagocytosis of virulent but not attenuated strains of Mycobacterium tuberculosis is mediated by mannose receptors in addition to complement receptors. J Immunol **150:**2920-30.
111. **Selwyn, P. A., D. Hartel, V. A. Lewis, E. E. Schoenbaum, S. H. Vermund, R. S. Klein, A. T. Walker, and G. H. Friedland.** 1989. A prospective study of the risk of tuberculosis among intravenous drug users with human immunodeficiency virus infection. N Engl J Med **320:**545-50.
112. **Sinsimer, D., G. Huet, C. Manca, L. Tsenova, M. S. Koo, N. Kurepina, B. Kana, B. Mathema, S. A. Marras, B. N. Kreiswirth, C. Guilhot, and G. Kaplan.** 2008. The phenolic glycolipid of Mycobacterium tuberculosis differentially modulates the early host cytokine response but does not in itself confer hypervirulence. Infect Immun **76:**3027-36.
113. **Smith, I.** 2003. Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence. Clin Microbiol Rev **16:**463-96.

114. **Soto, C. Y., M. C. Menendez, E. Perez, S. Samper, A. B. Gomez, M. J. Garcia, and C. Martin.** 2004. IS6110 mediates increased transcription of the phoP virulence gene in a multidrug-resistant clinical isolate responsible for tuberculosis outbreaks. J Clin Microbiol **42:**212-9.
115. **Soysal, A., K. A. Millington, M. Bakir, D. Dosanjh, Y. Aslan, J. J. Deeks, S. Efe, I. Staveley, K. Ewer, and A. Lalvani.** 2005. Effect of BCG vaccination on risk of Mycobacterium tuberculosis infection in children with household tuberculosis contact: a prospective community-based study. Lancet **366:**1443-51.
116. **Sreevatsan, S., X. Pan, K. E. Stockbauer, N. D. Connell, B. N. Kreiswirth, T. S. Whittam, and J. M. Musser.** 1997. Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci U S A **94:**9869-74.
117. **Stackebrandt, E., F. A. Rainey, and N. L. Ward-Rainey.** 1997. Proposal for a new hierarchic classification system Actinobacteria classis nov Int J Syst Evol Microbiol **47:**479-91
118. **Stewart, G. R., J. Patel, B. D. Robertson, A. Rae, and D. B. Young.** 2005. Mycobacterial mutants with defective control of phagosomal acidification. PLoS Pathog **1:**269-78.
119. **Sturgill-Koszycki, S., P. H. Schlesinger, P. Chakraborty, P. L. Haddix, H. L. Collins, A. K. Fok, R. D. Allen, S. L. Gluck, J. Heuser, and D. G. Russell.** 1994. Lack of acidification in Mycobacterium phagosomes produced by exclusion of the vesicular proton-ATPase. Science **263:**678-81.
120. **Thierry, D., A. Brisson-Noel, V. Vincent-Levy-Frebault, S. Nguyen, J. L. Guesdon, and B. Gicquel.** 1990. Characterization of a Mycobacterium tuberculosis insertion sequence, IS6110, and its application in diagnosis. J Clin Microbiol **28:**2668-73.
121. **Thierry, D., M. D. Cave, K. D. Eisenach, J. T. Crawford, J. H. Bates, B. Gicquel, and J. L. Guesdon.** 1990. IS6110, an IS-like element of Mycobacterium tuberculosis complex. Nucleic Acids Res **18:**188.
122. **Tortoli, E.** 2006. The new mycobacteria: an update. FEMS Immunol Med Microbiol **48:**159-78.
123. **Trunz, B. B., P. Fine, and C. Dye.** 2006. Effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness. Lancet **367:**1173-80.
124. **Tsolaki, A. G., S. Gagneux, A. S. Pym, Y. O. Goguet de la Salmoniere, B. N. Kreiswirth, D. Van Soolingen, and P. M. Small.** 2005. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of Mycobacterium tuberculosis. J Clin Microbiol **43:**3185-91.
125. **Tufariello, J. M., J. Chan, and J. L. Flynn.** 2003. Latent tuberculosis: mechanisms of host and bacillus that contribute to persistent infection. Lancet Infect Dis **3:**578-90.
126. **Turcios, L., Y. Casart, I. Florez, J. de Waard, and L. Salazar.** 2009. Characterization of IS6110 insertions in the dnaA-dnaN intergenic region of Mycobacterium tuberculosis clinical isolates. Clin Microbiol Infect **15:**200-3.
127. **Udwadia, Z. F., R. A. Amale, K. K. Ajbani, and C. Rodrigues.** Totally drug-resistant tuberculosis in India. Clin Infect Dis **54:**579-81.
128. **van der Wel, N., D. Hava, D. Houben, D. Fluitsma, M. van Zon, J. Pierson, M. Brenner, and P. J. Peters.** 2007. M. tuberculosis and M. leprae translocate from the phagolysosome to the cytosol in myeloid cells. Cell **129:**1287-98.
129. **van Embden, J. D., M. D. Cave, J. T. Crawford, J. W. Dale, K. D. Eisenach, B. Gicquel, P. Hermans, C. Martin, R. McAdam, T. M. Shinnick, and et al.** 1993. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol **31:**406-9.
130. **van Soolingen, D., L. Qian, P. E. de Haas, J. T. Douglas, H. Traore, F. Portaels, H. Z. Qing, D. Enkhsaikan, P. Nymadawa, and J. D. van Embden.** 1995. Predominance of a single genotype of Mycobacterium tuberculosis in countries of east Asia. J Clin Microbiol **33:**3234-8.

131. **Vera-Cabrera, L., M. A. Hernandez-Vera, O. Welsh, W. M. Johnson, and J. Castro-Garza.** 2001. Phospholipase region of Mycobacterium tuberculosis is a preferential locus for IS6110 transposition. J Clin Microbiol **39:**3499-504.
132. **Vergne, I., R. A. Fratti, P. J. Hill, J. Chua, J. Belisle, and V. Deretic.** 2004. Mycobacterium tuberculosis phagosome maturation arrest: mycobacterial phosphatidylinositol analog phosphatidylinositol mannoside stimulates early endosomal fusion. Mol Biol Cell **15:**751-60.
133. **Walburger, A., A. Koul, G. Ferrari, L. Nguyen, C. Prescianotto-Baschong, K. Huygen, B. Klebl, C. Thompson, G. Bacher, and J. Pieters.** 2004. Protein kinase G from pathogenic mycobacteria promotes survival within macrophages. Science **304:**1800-4.
134. **WHO.** 2011. Global tuberculosis control 2011. World Health Organization/HTM/TB/2011.16

135. **WHO.** 1994. TB: a global emergency. World Health Organization. .
136. **WHO.** 1999. What is DOTS? A guide to understanding the WHO-recommended TB control strategy known as DOTS. World Health Organization. .
137. **Zhang, M., J. Gong, Z. Yang, B. Samten, M. D. Cave, and P. F. Barnes.** 1999. Enhanced capacity of a widespread strain of Mycobacterium tuberculosis to grow in human macrophages. J Infect Dis **179:**1213-7.

CHAPTER 1: Characterization of *M. tuberculosis* Beijing strain GC1237: Study of the copies of IS*6110* and other genetic markers



Introduction

Since the first description in 1995, *M. tuberculosis* Beijing strains became a main health problem worldwide. Strains of this genotype are widespread in many regions of the world and frequently cause epidemic outbreaks (5, 10, 20). The distribution of Beijing strains differs, being low in Western Europe, although a slight increase in the number of Beijing strains has been detected over time (5). Recent studies have illustrated the high frequency of this genotype in different regions of China being around 80%. Moreover, according to spoligotyping database (SpolDB4), strains from this genotype were found to be present in the largest number of countries globally (13% of global isolates) (22). This family is generally considered to be associated with drug-resistance (2, 6, 7), although this association has not been found in all geographic settings (1, 20).

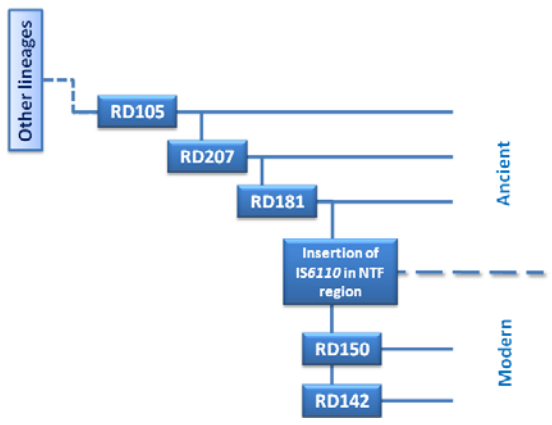
The predominance of this lineage probably results from genetic advantages, including unidentified virulence factors and the modulation of specific host responses not yet thoroughly investigated. For example, a study using human macrophages found that Beijing strains grew significantly faster than non-Beijing strains (49). However, when it was analyzed whether this characteristic was common to Beijing strains, it was found that not all Beijing strains grew rapidly (46). In addition to this, there are some studies that relate hypervirulence of Beijing strains with production of PGL (12, 39), which is a putative virulence factor that attenuates the host's innate immune response and ability to control infection (39). The *pks15/1* locus, described to be polymorphic among members of the MTBC (14), is involved in the biosynthesis of the PGL and it is characteristic of Beijing strains to have an intact *pks15/1* region for which they could produce a functional Pks15/1 synthase (14, 43, 45). However, when monocytes were infected with a non-Beijing strain transformed with the functional *pks15/1* gene, growth was similar to the wild-type non-Beijing strain, suggesting that PGL did not influence growth of *M. tuberculosis* (22). Another example of virulence factor could be that Beijing strains have a basal level of transcription for the dormancy regulon genes (*dosR, Rv3130c, hspX, fdxA* and *narX*) that is up to 50-fold higher than that of non-Beijing genotype strains (40). More recently, it was found that Beijing strains possess two copies of *dosR* which was suggested to be partly responsible for the constitutive *dosR* over-expression phenotype (15).

Beijing strains are currently attracting considerable worldwide attention because they display important pathogenic features (30, 49).

Subfamilies Beijing

Members of this family were subdivided into modern/typical and ancient/atypical sublineages based on the analysis of the NTF locus (34). In addition to the characteristic deletion of the DR locus, Tsolaki *et al.* described LSPs (RD105, RD142, RD150, and RD181) which further divided this family into four monophyletic subgroups (47). They sought LSPs that were unique-event polymorphism (UEPs) within Beijing genotype. A total of 21 distinct LSPs were identified that were present in H37Rv but absent from the Beijing strains. Based on the frequency of these 21 LSPs, three classes were identified: four common LSPs (RD105, RD149, RD152 and RD207) were found deleted in all Beijing strains. RD207 (also detected by spoligotyping), RD149 and RD152 are associated with mobile genetic elements and may not be UEPs. Three LSPs (RD181, RD142 and RD150) were found variably deleted and these are not associated with either repetitive DNA or mobile genetic elements therefore likely to represent UEPs and the remaining 14 LSPs were unique to a single strain. Based on this, Beijing lineage is principally defined by deletion RD105, which is a useful marker for the identification of this family. RD181, RD142 and RD150 further divided this family into four monophyletic subgroups. RD142 and RD150 are only observed in strains with RD181 being this the more ancestral event (47). After that, the RD207 was included within the monophyletic subgroups of the East-Asian lineage (19) (Figure 11).

A



B

Region of Diference (RD)	Affected genes
105	Rv0071 to Rv0074
142	Rv1189 to Rv1192
149*	Rv1572c to Rv1586c
150	Rv1671 to Rv1674
152*	Rv1754c to Rv1765
181	Rv2262c to Rv2263
207	Rv2814c to Rv2820c

Figure 11: Subfamilies Beijing. (A) Schematic phylogenetic tree of the *M. tuberculosis* Beijing family. The name of the LSPs or regions of difference are shown in rectangles. The evolutionary process of the lineage (ancient to modern) is also indicated. **(B)** Genes affected for the deletion of the RD. The genes affected for the deletion of RD149 and RD152 (common in all Beijing strains) are included (*). Adapted from (19, 31, 47).

Some authors have proposed an evolutionary pathway of Beijing lineage based on RD deletions and on the IS*6110* insertions in the NTF region indicating that ancient strains have neither deletion RD181 nor insertion of IS in NTF region (16). However, recent studies have shown that not all ancient/atypical strains are RD181 intact which indicates that the RD181 deletion occurred before the insertion of an IS element in the NTF locus (24, 31) (Figure 11). Moreover, it has been demonstrated that the 3' common truncation in the gene Rv2820c (RD207, figure 11B) of Beijing strains could enhance mycobacterial virulence *ex vivo* and *in vivo*. This enhancement was not observed in non-Beijing strains with intact gene (29).

IS*6110* in Beijing family

Three features which characterize this family are related to IS*6110* insertion sequence: The presence of one copy of this element in the *ori*C region, the deletion of the right-site DR spacers (the so-called deletion RD207) and similar RFLP multiband pattern profile (22). Members of this genotype usually are high copy number of IS*6110* (between 15-25 copies per genome) suggesting the relevance of this element in the variability of their genomes. IS*6110* insertions, genetic reorganizations and deletions (gene deletion can occur by homologous recombination between 2 flanking copies of IS*6110*) are some of the mechanisms proposed to be responsible for differences in the virulence phenotypes of *M. tuberculosis*. Supporting this possibility, sublineages of this family were identified to carry an important genome duplication that involves up to 8% of the genome (corresponding to more than 300 genes). Copies of IS*6110* were identified flanking that duplication, thus suggesting the occurrence of homologous recombination event mediated by this IS (15). It has been demonstrated that IS*6110* can upregulate downstream genes through an outward-directed promoter in its 3' end (41, 44). The fact that the Beijing lineage contains a larger number of IS*6110* copies than other lineages (21) could be related to the special characteristics of this family in terms of virulence and capacity for rapid dissemination.

The clinical isolate *M. tuberculosis* GC1237

M. tuberculosis GC1237, which belongs to the Beijing family, has been responsible for different epidemic outbreaks in the Gran Canary Island. The first case recognized on the island with this strain was a refugee from Liberia (Africa) diagnosed with smear-positive pulmonary tuberculosis in 1993. Its explosive spread in this community over the next few years was reported in 2001 (11). Nowadays, this strain continues being predominant in the area due to its rapid and successful dissemination within the community (Table 1).

Table 1. Percentage of Beijing isolates in Canary Island from 1991 to 2008. Adapted from (11, 33).

Period	Method	Total isolates	Beijing percent (%)
1991-1992	RFLP	85	0
1993	RFLP, spoligotyping	179	5.5
1994	RFLP, spoligotyping	148	8.1
1995	RFLP, spoligotyping	110	16.4
1996	RFLP, spoligotyping	129	27.1
1999	Spoligotyping	40	22.5
2002	Spoligotyping	154	28.57
2003	Spoligotyping	49	28.57
2004	Spoligotyping	120	29.16
2007-2008	Spoligotyping	214	25.23

It was reported that one isolate with the same RFLP was detected in Madrid. Although in that study the strain did not show increased intracellular replication (THP1-cells) and it was not distinguishable from other isolates (3), the increased capacity of infection and the high success rate of this strain to spread rapidly could be a consequence of genetic advantages and unidentified mechanisms of virulence not yet thoroughly investigated.

Objectives

- To classify the clinical isolate *M. tuberculosis* GC1237 within the Beijing family with the analysis of several Region of Differences (RD).
- To localize all the copies of IS*6110* insertion sequence in the genome of *M. tuberculosis* GC1237.
- To study the direct repeats flanking each copy of IS*6110* of GC1237.
- To analyze the orientation and distance to neighbouring genes of each copy of IS*6110* present in *M. tuberculosis* GC1237.

Material and Methods

Bacterial strains, culture media, and growth conditions

BACs library of *M. tuberculosis* GC1237 (Figure 14A) was used in this study. This library, which represents around 75% of the chromosome of GC1237, was constructed in a previous work using pBeloBAC11 plasmid.

E. coli DH10B cultures were used in order to isolate the BACs. Strains were grown at 37ºC in Lurina-Bertani (LB) broth supplemented with chloramphenicol (CM) (12.5μg/ml).

The clinical isolate *M. tuberculosis* GC1237, *M. bovis* BCG and the reference *M. tuberculosis* H37Rv strains were used in this work. Mycobacterial strains were grown at 37ºC in Middlebrook 7H9 broth supplemented with albumin-dextrose-catalase (ADC) and 0.05% Tween 80 or in Middlebrook 7H10 medium Bacto agar supplemented with oleic acid-albumin-dextrose-catalase (OADC) (Difco Laboratories, Detroit, Mich.) and 0.05% Tween 80 (23). Liquid cultures were grown to logarithmic phase for mycobacterial DNA extraction.

M. tuberculosis manipulation was carried out in a biosafety level 3 (BSL3) laboratory (facilities notification A/ES/04/I-05; activity A/ES/06/03).

Mycobacterial DNA extraction

Genomic DNA of mycobacterial strains was isolated using the CTAB method (48). Briefly, mycobacteria were resuspended in 400μl TE (100mM Tris/HCl, 10mM EDTA, pH 8.0) and heated for 20min at 80ºC. Samples were slightly cooled at room temperature before adding 0.5mg lysozyme and were then incubated for at least 1h at 37ºC. Subsequently, 0.05mg proteinase K dissolved in 75μl 10% SDS were added and the suspension warmed for 10min at 65ºC. Hereafter, 100μl 5M NaCl and 100μl CTAB/NaCl (10% CTAB in 0.7M NaCl) pre-warmed at 65ºC were added and samples incubated for further 10min at 65ºC. Genomic DNA was extracted by adding 750μl of chloroform:isoamylalcohol 24:1 (v/v). Samples were mixed by vortexing for 10s before centrifugation (13,000rpm for 5min). The upper (aqueous) phase was transferred to a fresh tube containing 450μl isopropanol and samples incubated overnight at -20ºC. Precipitated nucleic acids were collected by centrifugation (13,000rpm for 10min at 4ºC). The pellets were dissolved in 50μl double-distilled water and extracted DNA was quantified by Abs_{260} readings using a ND-1000 spectrophotometer (NanoDrop Technologies).

BAC DNA extraction

BAC DNA extraction was done as previously described by Birnboim *et al*. (8) with minor modifications. Briefly, 100ml of BAC-transformed *E. coli* was prepared in LB medium containing 12.5μg/ml CM, and the cultures were grown overnight at 37ºC with vigorous and continuous agitation. Then, the bacterial cells were collected by centrifugation. The bacterial pellet was softly resuspended in a solution of 5ml of 50mM glucose 10mM EDTA, 25mM Tris pH 8, 200mg lysozime was added to the tube and a solution of 4ml of ice-cold NaAc pH 4.8 was added to the mixture. The tube was placed on ice and the precipitated debris was removed by centrifugation. Then, 14ml of chloroform/isoamilic alcohol (24/1) were added to the supernatant and after centrifugation the aqueous phase was transferred to a new microfuge tube. DNA of the BACs was precipitated adding isopropanol and finally obtained by centrifugation.

Identification of genomic deletions and analysis of *pks*15/1 region in *M. tuberculosis* GC1237 strain

The study of the genomic deletions of the regions of difference 105, RD142, RD150, RD181 and RD207 in *M. tuberculosis* GC1237 strain, which identify and phylogenetically sub-classify the Beijing lineage, was performed by PCR. Other regions of difference (RD108, RD110a, RD127, RD129, RD139BW, RD149, RD152, RD165, RD166 and RD182a) were also analysed. The primers used in these amplifications were as described elsewhere (47) and are included in Table 1 of Annex I. The PCR was carried out in a total volume of 50μl, containing 0.5μg of DNA, 5μl of 10x PCR buffer, 200μM dNTPs, 12.5pmol of each primer and 1U of Taq Gold polymerase (Roche). Before the amplification, the template was initially denatured by incubation at 94ºC for 9min then the amplification was performed in 35 cycles of 94ºC for 30s, corresponding annealing temperature for 30s, and 72ºC for 2 to 3min depending on the amplified product. After the last cycle, the samples were incubated at 72ºC for 10min. The RD deletions were confirmed by DNA sequencing using H37Rv as reference genome.

The *pks15/1* polymorphism in this clinical isolate was determined by PCR, as above described, with the primers pks1I and pks1J (Table 2 of Annex I) and sequencing and analyzed as previously described (14).

Location of the copies of IS*6110* insertion sequence in *M. tuberculosis* GC1237

- BACs with IS*6110*

The study of the presence of IS*6110* insertion sequence in GC1237 BACs library was carried out by PCR with the specific primers of this sequence, Gab 1 and Gab 2 (42) (Table 5 of Annex I) and was performed as explained before in this chapter changing the final volume to 25µl. After that, PCR products were analyzed by 0.8% agarose gel electrophoresis and visualized by ethidium bromide staining.

- Ligation mediated PCR (LMPCR)

LMPCR was used to locate the copies of IS*6110* as previously described by Prod'hom *et al*. (38). This technique amplifies both ends of each copy of IS*6110*. Briefly, BACs containing IS*6110* and genomic DNA of *M. tuberculosis* GC1237 were digested with *Sal*I or with *Sma*I and ligated to a linker containing a *Sal*I restriction site or a *Sma*I restriction site, respectively. The resulting template was then digested by *Sal*I or with *Sma*I. PCR was performed using ISA1 and ISA3 (Table 2 Annex I), specific primers for IS*6110* directed outwards (32) and the common linker primer Salgd (38) (Table 2 Annex I) (Figure 12). PCR products were purified using GFX PCR DNA gel band purification kit (Amersham Pharmacia Biotech) and the restriction enzyme ExoSAP-IT® (Affymetrix).

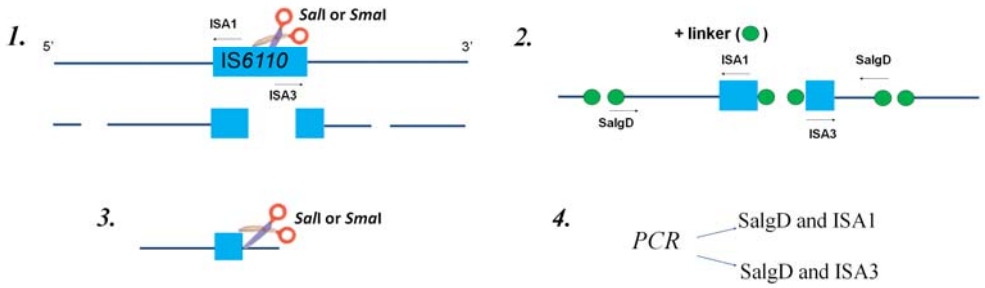


Figure 12: Ligation Mediated PCR (LMPCR). **(1)** Total DNA or BACs containing IS*6110* are restriction digested, **(2)** linkers are ligated, **(3)** unspecified ligations are removed and **(4)** transposon junctions are PCR amplified.

The amplified products were sequenced using CNIO service with the corresponding oligonucleotides and when a match was found for a flanking region in the databases, additional primers were designed to verify the point of insertion. PCR amplification was carried out as previously described in this chapter with genomic DNA of GC1237. PCR

products included the complete sequence of IS*6110* and approximately 100-200 bp of both flanking sequences.

Each amplified PCR product was sequenced and analysed for homology with Tuberculist (http://genolist.pasteur.fr/TubercuList), Bovilist (http://genolist.pasteur.fr/BoviList) and NCBI (http://www.ncbi.nlm.nih.gov/) database Blast analysis.

- PCR with specific primers

In addition, we designed primers to amplify the different regions that might include IS*6110* based in the published genome locations of IS*6110* in 210 and W Beijing strains (4). As controls H37Rv and BCG genomes were used.

The primers used to localize IS6110 are collected in Table 5 of Annex I.

Determination of Direct Repeats (DR) of each copy of IS*6110* in GC1237 genome

The DR generated by the mechanism of the transposition of IS*6110* was determined with the sequence analysis of the flanking regions of each copy of IS*6110* in GC1237 genome.

Design of a multiplex PCR to detect the clinical isolate *M. tuberculosis* GC1237

To detect the clinical isolate *M. tuberculosis* GC1237 faster than by RFLP and spolygotyping, a multiplex PCR based on the detection of two of its IS-locations was designed in this study. The first target (fragment of 550bp) is the IS*6110* located between *dnaA:dnaN* genes which is common to all Beijing strains (26) and it is amplified with the primers dnaII and dnaAIII (Table 3 of Annex I). This last oligonucleotide is more specific to Beijing genotype as it anneals in the 3' end of the IS*6110* and the following nucleotides of that region. The second target, amplified with primers Rv2179cfw and Rv2179crv (Table 3 of Annex I), is specific of GC1237 (see results of this chapter). In this case, the GC1237-PCR product includes the entire sequence of IS*6110* (1626 bp) and the PCR fragment of the rest of the strains correspond to 261 bp.

The multiplex PCR was carried out as describe above with minor modifications. Briefly, PCR was performed in a final volume of 25µl, containing 200µM dNTPs, 50mM KCl, 1.5mM $MgCl_2$, 2.5U of puReTaq DNA polymerase (puReTaq Ready-to-go, GE Healthcare), reaction buffer, 25µM of each primer and 0.5µg of DNA. Temperature cycling conditions included 94°C for 10 min, followed by 35 cycles of 94°C for 30s, 70°C for 30s, and 72°C for 1.5min, and 10min for a final extension at 72°C. PCR

products were analyzed by 1% agarose gel electrophoresis and visualized by ethidium bromide staining.

Results and discussion

Characterization of *M. tuberculosis* GC1237: study of Regions of difference (RD) and *pks15/1* region

Comparative whole-genome hybridization of Beijing strains showed LSPs also known as RDs, which subdivide the Beijing family into at least four phylogenetic subgroups (47), raising the possibility that there are phenotypic differences within the Beijing family. The subdivisions are made on the basis of RD105 and RD207, and the variable appearance of RD181, RD150 and RD142 deletions (47). With the aim to classify GC1237 strain within this genotype, these regions of difference were studied and the results indicated that GC1237 has deleted RD105 and RD207, as all Beijing strains, and RD181. This strain also has deleted RD149 and RD152 which is common to find in all Beijing strains (25).

A copy of IS*6110* without flanking DR was detected in RD152 and in RD207 confirming a recombination event of two previous copies of IS*6110* resulting in the loss of these regions. This finding is in agreement with different studies indicating that strains with a high number of IS*6110* copies (>14) have lost genomic regions more often than strains with only few copies (9), as in the case of GC1237. RD149 has also been associated with recombinations between mobile genetic elements (47) but IS*6110* is not implicated in the deletion of this region as we show the absence of IS*6110* in this region in GC1237 (Figure 14B-14C).

The deletion of other RD observed in clinical strains belonging to Beijing family from East Asia clade was also studied (RD110a, RD127, RD129, RD139BW, RD165, RD166 and RD182a) (47); GC1237 has these RD intact. However, the possibility of the presence of specific deleted regions in GC1237 is not ruled out. The high number of different combinations of deleted RD that we can find in Beijing strains shows that although the Beijing genotype is considered as a homogenous group, genetic variability can be found these strains.

The study of *pks15/1* region revealed an intact gene in GC1237 indicating that this strain could retain the ability to produce functional glycolipid PGL in contrast to the reference strain H37Rv.

Another feature of GC1237 is the presence of two extra regions, Mb2047c-Mb2048c and Mb3356-Mb3357, with respect to H37Rv. These regions, present in other sequenced strains of *M. tuberculosis* and *M. bovis,* do not correspond exactly to the described RvD1 and RvD5, respectively. The two IS*6110* localized in these regions suggest possible implication in new genomic reorganizations with loss of these regions.

By studying the flanking regions of IS*6110* compared with H37Rv, we have detected as many new regions as loss of other ones. These findings confirm the implication of IS*6110* in genomic reorganizations and its capacity to generate genomic plasticity.

Different strategies to localize all copies of IS*6110* insertion sequence in *M. tuberculosis* GC1237

In this study, nineteen different copies of IS*6110* were localized in GC1237 applying two main strategies: LMPCR technique with BAC library and genomic DNA of the studied strain and PCR with designed primers based on the known points of the insertion sequence in the two reference Beijing strains 210 and W (4). The presence of IS*6110* in the representative GC1237 BAC library (Figure 14A) was detected using PCR with specific primers of this element and 10 of total of 59 BACs were positive for IS*6110* (Table 2). Using LMPCR the ten different copies were localized in GC1237. This technique was further utilized with genomic DNA (Figure 13A) to detect the rest of the copies present in GC1237 by amplifying one or both ends of each copy of IS*6110* and its flanquing sequences.

As a second strategy, based on genome known locations of IS*6110* in 210 and W strains (4), we designed primers to amplify the different regions that might include IS*6110* in the studied strain (Figure 13B). This design was realized using *M. tuberculosis* H37Rv as genome reference.

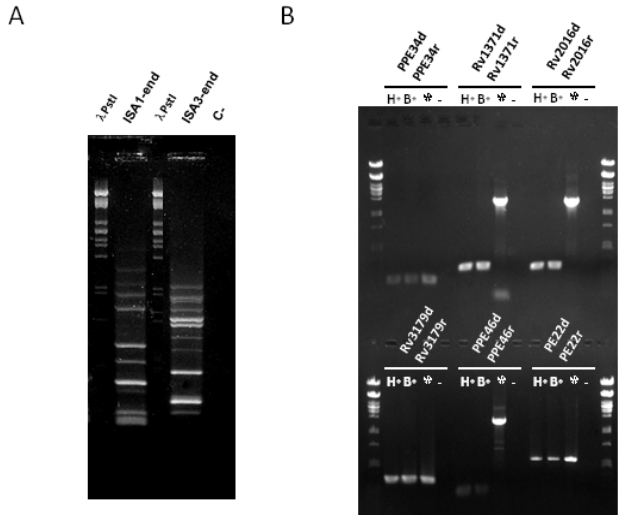


Figure 13: Example of PCR analysis. (A) LMPCR of GC1237 genome. The flanking regions of IS*6110* were amplified with the oligonucleotides ISA1 or ISA3 and SalgD. **(B)** PCR analysis of different regions that might include IS*6110* based in genome known locations of IS*6110* in 210 and W Beijing strains. Each set of four reactions consists of two positive controls PCR assay with genomic DNA of H37Rv and BCG (H^+ and B^+, respectively), PCR with genomic DNA of GC1237 and a negative control assay.

LMPCR technique was useful to determine the points of insertion of this sequence in *M. bovis* strains (35), which contain very minor number of IS*6110* (1 to 5). However, strains of *M. tuberculosis* present a high level of copies of IS*6110* and the probability to localize all the copies by LMPCR technique lowers significantly. In this regard, the use of BAC library of GC1237 strain has been a successful tool.

By RFLP, GC1237 strain presents 15-16 copies of IS*6110*. However, using two different molecular strategies it was possible to determine the exact location of the 19 copies of IS*6110* in GC1237 genome (Table 2). Except for some sequenced strains, to our knowledge this was the first time to localize all IS*6110* copies in a strain of *M. tuberculosis* containing a high number of this element.

The three Beijing strains share a high number of locations of IS*6110*

Comparing the three Beijing strains (210, W and GC1237) it was observed that they share nine IS*6110* positions, these include the two insertions characteristic of the Beijing family, the insertion A1, and in the Direct Repeat Region (Figure 4C and Table 2). GC1237 strain contains eight copies of IS*6110* in intergenic regions and eleven in intragenic regions (Table 2). Of the eleven interrupted genes, *pip*, *ctpD, idsB* and Rv2180c are involved in bacterial metabolism and respiration, four correspond to genes with no predicted function and three are members of the PE and PPE family. The interruption of these ORFs by the IS*6110* probably causes their no-transcription and consequently, these genes would be non-essential for *M. tuberculosis*. Interestingly, it was found that the two copies of IS*6110* within the *PPE34* were consecutive and in the same orientation but the second copy lacked 60 bp at the 5' end. Moreover, the copy located between *PPE39* and *PPE40* genes, is also present in H37Rv but the region is longer in the studied strain.

Numerous studies have indicated that IS*6110* has some preferential points of insertion (hot-spots) in the genome. Both IS*1547* (*ipl* loci) is an example of a site of intensive insertion of this element (17, 18). The studied strain corroborates this hot-spot with the insertion of a copy of IS*6110* in each *ipl* region: one between Rv0794c and Rv0797, which is in the opposite orientation in H37Rv and other studied Beijing strains and the second one was located between Rv3324A:Rv3327 in the same orientation as H37Rv.

With the exception of the copy of IS*6110* located between *dnaA*:*dnaN* genes, we have not found any copy of IS*6110* in the quarter of *M. tuberculosis* GC1237 circular genome surrounding these genes (Figure 14C). This fact was also observed in the two Beijing strains 210 and W. This result is in agreement with other authors when they have characterized IS*6110* integration loci (43). This finding could indicate that in this part of the genome there are no specific sequences or regions recognised by the

sequence or that these genes have an important role in *M. tuberculosis* and the interruption or modulation of their expression could result in no benefit or even be a disadvantage for the bacteria.

It is frequent to find at least a copy of IS*6110* in the NTF region in strains belonging to Beijing family (37) and some members of this family may have a second insertion within this locus such as the MDR strain W and its progenies (5, 27). However, the absence of IS*6110* in NTF locus is characteristic of ancestral sublineages within the Beijing genotype (26). Due to the course of evolution, some strains ("modern" sublineages) have acquired the insertion of IS*6110* in this region (27). In addition to this, some authors have proposed an evolutionary pathway of the Beijing lineages based on the RD deletions and on the IS*6110* insertions in the NTF region indicating that ancient strains could have the deletion of RD181 and modern strains have RD181 and at least one insertion of IS*6110* in NTF region indicating that the RD181 deletion occurred before the insertion of an IS*6110* in the NTF locus (24, 31). Our results are in agreement with this classification because GC1237 presents a deletion in RD181 but lacks IS*6110* in NTF locus.

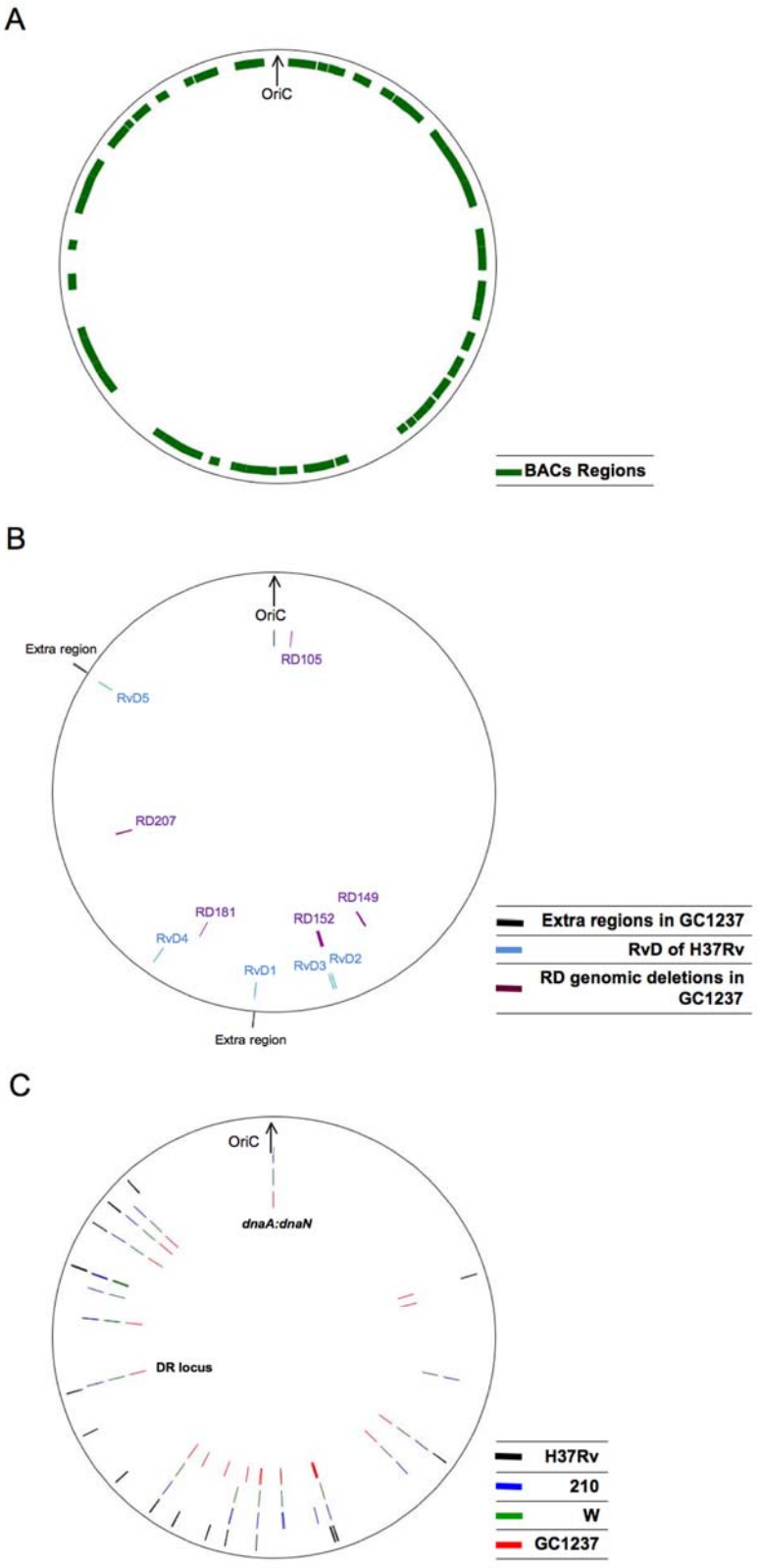


Figure 14: Circular plots of GC1237 genome prepared with DNAplotter. (A) BACs regions of the representative library of *M. tuberculosis* GC1237. **(B)** Genomic regions absent in *M. tuberculosis* H37Rv (RvDs), lacking regions of difference (RD) and extra regions in GC1237 strain. **(C)** Locations of IS*6110* in the Beijing strains GC1237, 210 and W. The Direct Repeat region (or DR locus) and the *dnaA:dnaN* region are indicated.

Table 2: Insertion sites of IS*6110* in *M. tuberculosis* GC1237, 210 and W.

GENES [a]	GC1237		210 [d]	W [e]
	GC1237 genome [b]	BACs [c]		
Rv0001 (*dnaA*):Rv0002 (*dnaN*)	+	-	+	+
Rv0794*c*:Rv0797 (IS*1547*)	+	B4		
Rv0840c (*pip*)	+	-		
Rv1135c		B11	+	+
Rv1371	+	C7	+	+
***Rv1469 (*ctpD*)**	+	C10	+	+
Rv1754c		D4	+	+
Rv1754c-Rv1765c (RD152)	+	D4		
Rv1798:Rv1799 (*lppT*)		D6	+	
Rv1917c (*PPE34*)		D9	+	
Rv1917c (*PPE34*)		D9	+	+
Rv1917c (*PPE34*)	++	D9		
Rv2016	+	D12	+	+
MT2080 (Mb2047c):MT2081 (Mb2048c)	+	D12		
Rv2077A:Rv2078	+	E1		
Rv2104c:Rv2107 (*PE22*)		E1	+	
Rv2104c:Rv2107 (*PE22*)		E1		+
Rv2107 (*PE22*):Rv2108 (*PPE36*)		E1	+	
***Rv2180c**	+	-		
Rv2286c	+	E7		
Rv2352c (*PPE38*)		E8	+	+
Rv2353c (*PPE39*):Rv2356c (*PPE40*)	+	E8		
Rv2813-Rv2820c (RD207)	+	-	+	+
***Rv3018A (*PE27A*):Rv3019c (*esxR*)**	+	-	+	+
Rv3019c (*esxR*):Rv3020c (*esxS*)		-	+	+
Rv3128c		G1	+	+
Rv3128c:Rv3129		G1		+
Rv3178:Rv3179		-	+	
Rv3179:Rv3180c		-	+	+
Rv3184:Rv3187		-	+	+
Rv3324A:Rv3327 (IS*1547*)	+	-	+	+
Rv3383c (*idsB*)	+	G7	+	+
***Rv3427c (IS*1532*):Rv3428c (IS*1532*)**	+	G8	+	+

[a] Gene names come from the *M. tuberculosis* H37Rv genome map. One gene is listed when the location of insertion is in that gene and two genes are listed when the insertion of IS*6110* is between both.
[b] Sites are in order of appearance on the *M. tuberculosis* H37Rv genome map.
[c] Name of BAC.
[d, e] The plus sign indicates that the *M. tuberculosis* 210 and W contain a copy of IS*6110* in those regions (4).
[*] *Candidate locations to act as a promoter.*

The open reading frames represent 91% of *M. tuberculosis* genome (13) but in GC1237 the insertion of IS*6110* into coding regions occurred in 42% of the cases then, apparently the transposition is relatively more frequent in intergenic regions. Our results agreed with other studies that found that 58% of discrete IS*6110* insertion sites occurred within coding regions in *M. tuberculosis* (43) and in *M. bovis* strains (35). In this context, because of the insertion of IS*6110* in possible promoter regions the probability of IS*6110* influencing the expression of the neighbouring genes is increased. It is necessary to take into account that the locations of IS*6110* observed in the different genomes are the result of the transposition and selection of viable infecting mycobacteria because when insertion occurs in a gene necessary for virulence, we cannot observe it.

Analysis of Direct Repeats (DR) of each copy of IS*6110* in GC1237 genome

The flanking regions of each of the nineteen copies of IS*6110* in GC1237 genome were analyzed and the presence of DR of 3-4 nucleotides at the extremities of sixteen IS*6110* sequences were detected (Table 3). These DR indicate that the presence of IS*6110* was due to transposition events. The other three copies analysed were not flanked by DRs. Genomic regions loss was observed in two of these locations which correspond to Rv1754c-Rv1765c (RD152) and RD207 respectively, and this is probably a consequence of recombination between two adjacent copies of IS*6110* (47). The last of the three IS*6110* elements without DR, which was localized between Rv0794c:Rv0797, was in the opposite orientation with respect to the IS*6110* in the reference genome H37Rv. The IS*6110* in H37Rv contains DR and the lack of DR in the studied IS*6110* could be explained by reorganization of this region in GC1237 genome.

The two copies localized in *PPE34* are flanked by the direct repeat sequence TTA. These DRs were located in the 5' end of the first IS*6110* and the 3' end of the second one, probably due to simultaneous transposition of the two copies. The DR sequences differ among each other indicating the lack of transposition specificity.

There are very few studies where authors report DR flanking IS*6110* (32, 35). In these studies the number of copies is low (<6) and all of them were flanked by DR, indicating that these IS*6110* are consequence of transposition. In our study, because of the high number of copies of IS*6110* we observed some copies flanked by DR and other copies without DR suggesting that the probability of rearrangement process between copies rises when the number of those increases producing more variability among strains. This finding is in agreement with different studies indicating that strains with a high number of IS*6110* copies (>14) have lost genomic regions more often than strains with only few copies (9), as is in the case of GC1237.

Table 3. Direct repeats of IS*6110* insertion sequence in *M. tuberculosis* GC1237.

Genes of IS*6110* in GC1237 genome	DRs[a]
Rv0001 (*dnaA*):Rv0002 (*dnaN*)	ATT
Rv0794c:Rv0797	-
Rv0840c (*pip*)	ACG
Rv1371	GAGG
Rv1469 (*ctpD*)	CGT
Rv1754c-Rv1765c (RD152)	-
Rv1917c (*PPE34*)	TTA
Rv1917c (*PPE34*)	TTA
Rv2016	AGG
MT2080 (Mb2047c):MT2081 (Mb2048c)	GAA
Rv2077A:Rv2078	AGG
Rv2180c	AGC
Rv2286c	ATC
Rv2353 (*PPE39*):Rv2356c (*PPE40*)	CCG
Rv2813-Rv2820c (RD207)	-
Rv3018A (*PE27A*):Rv3019c (*esxR*)	GCC
Rv3324A:Rv3327 (IS*1547*)	GGC
Rv3383c (*idsB*)	ATC
Rv3427c (IS*1532*):Rv3428c (IS*1532*)	CCCG

[a]The absence of DR is indicated by minus.

Study of the orientation and the distance of IS*6110* to the neighbouring genes

Different studies have indicated that IS*6110* could up-regulate the expression of downstream genes. Previous studies have shown that when IS*6110* is inserted in the same orientation as, and close enough to, a downstream gene, IS*6110* could potentially function as a promoter (41). We analysed the orientation of each copy of the IS*6110* in GC1237 strain and the distance to the close gene in order to test the promoter function of IS*6110.* We obtained 8 genes with an IS*6110* inserted upstream in the right orientation but in 4 locations the distance is reasonable to act as a mobile promoter (Table 2). One of them is located 297 bp upstream of the Rv1468c gene, another 138 bp upstream of *PE27A* gene, the third one 41 bp upstream Rv3427c gene and the last one is located 31 bp upstream of the essential gene Rv2179c.

The analysis of the natural insertion site of IS*6110* upstream the gene *ctpD* in the *M. tuberculosis* Beijing strain 210 was carried out both in broth and in monocytes and compared with that in H37Rv (41). In broth, transcriptional levels of this gene were

similar in the two strains. However, the expression of *ctpD* was three to fivefold higher levels inside monocytes when IS was upstream (41). As GC1237 presents a copy of IS*6110* in this same point and in the same orientation, the expression of *ctpD* in this strain was analysed (by qRT-PCR) and compared with its expression in H37Rv. Similar results to those reported were obtained both in broth and inside the cells corroborating the promoter activity of IS*6110* in this location.

The insertion of IS*6110* in Rv2180c gene stands out as it is located 31 bp upstream of the essential gene Rv2179c and close to the essential gene *AroG* which is involved in the common aromatic amino acid biosynthesis in *M. tuberculosis* (36). This location was also investigated in other *M. tuberculosis* strains, including Beijing and non-Beijing strains, and surprisingly, this point of insertion is unique to GC1237 strain. For that, the promoter activity in this location will be studied widely in Chapter 2.

Detection of the clinical isolate *M. tuberculosis* GC1237 by multiplex PCR

With the aim of facilitating the detection of this isolate faster than with traditional techniques, a multiplex PCR based on the detection of two locations of IS*6110* in GC1237 genome was developed in this study (Figure 15). The design of this assay was possible as we located all the copies of IS*6110* in this strain.

The first step of the design was to distinguish between Beijing and non-Beijing genotypes. For that, the IS*6110* located between *dnaA*:*dnaN* (Target 1 in Figure 15A) and the genomic deletion of the region of difference RD105 were chosen. These two targets were selected for its specificity of Beijing genotype. Given that the presence of a copy of IS*6110* between *dnaA:dnaN* is not exclusive of Beijing strains (28), one of the primers used for the amplification of this region was designed annealing in the 3' end of the IS*6110* and the following nucleotides (Target 1 of Figure 15A) becoming totally specific of Beijing genotype. The second step was to distinguish the GC1237 genotype among Beijing genotypes. To obtain this, screening of IS*6110* locations in Beijing strains and non-Beijing strains was carried out as mentioned above and finally, the IS*6110* located in Rv2180c was selected as specific of this strain (Target 2 of Figure 15A). The multiplex PCR was performed with these three targets and it was possible to differentiate GC1237 genotype from the rest of the strains used as control strains. However, this PCR was performed with pure DNA and when it was performed with DNA obtained from boiled samples (it is important to remember that the main objective was to obtain a rapid test for the identification of GC1237 genotype), the PCR results failed with high frequency. After this data, the detection of RD105 region was removed from the multiplex PCR and on this occasion, as Figure 15B shows, it was a success. Although Figure 15 only shows the PCR results obtained with *M. tuberculosis* (Beijing

and non-Beijing), this multiplex was also tested with different mycobacteria genomes as *M. vaccae, M. bovis* and *M. smegmatis* obtaining in all cases the same profile as H37Rv strain.

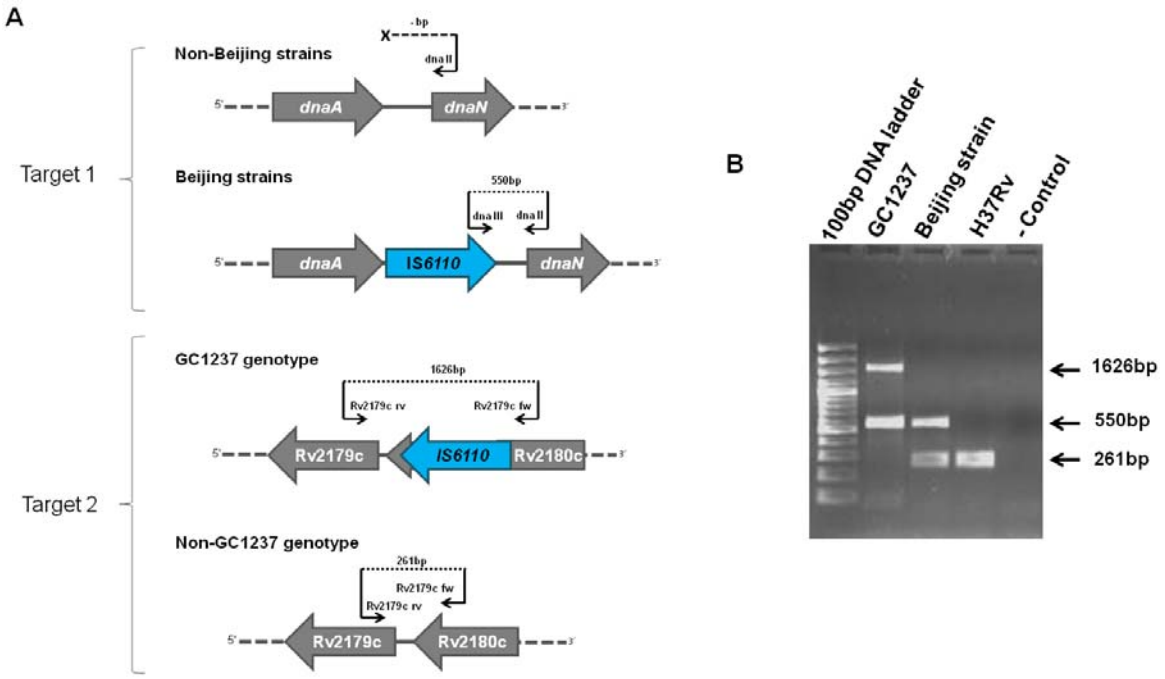


Figure 15: Detection of GC1237 genotype by multiplex PCR. **(A)** Schematic representation of the two targets selected for the multiplex PCR. Target 1 (primers dnaII and dnaIII) permits to differentiate between Beijing and non-Beijing genotypes. Target 2 (primers Rv2179c fw and Rv2179c rv) was selected to identify GC1237 genotype among Beijing strains. **(B)** PCR analysis of the multiplex PCR. Two fragments were obtained for isolates with GC1237 genotype (1626 bp and 550 bp). Two amplicons of 550 bp and the 261 bp were obtained as expected for other Beijing isolates and only one fragment of 261 bp in H37Rv used as non-Beijing strain.

Due to the high incidence of this strain in the Gran Canary Island, its rapid detection would be useful for better control of TB cases. Considering that the RFLP and spoligotyping methods are difficult to implement in the clinical setting, the described multiplex PCR could bring the possibility of *in situ* identification of GC1237 genotype. The incidence of this strain in this Island has been screened by spoligotyping until now. This technique is good distinguishing the Beijing lineage from the rest; however this method does not provided the information about clonality within this genotype. In addition, the better use of this method is for the study of 40 isolates at a time. The RFLP based on IS*6110* give the clonality information, but this technique requires higher quantity of DNA of high quality. In contrast to spoligotyping and RFLP, this multiplex PCR can be used to identify one or more positive isolates. It is quick and inexpensive, it has a simple interpretation and it can be completed in one day.

Conclusions

- Based on NTF classification, *M. tuberculosis* GC1237 is an ancestral strain.
- The LSPs RD105, RD149, RD152, RD181 and RD207 are deleted in GC1237 strain.
- The representative BAC library of *M. tuberculosis* GC1237 was a useful tool for locating IS*6110* in a high-copy strain.
- *M. tuberculosis* GC1237 has 19 copies of the insertion sequence IS*6110*.
- The complementation of different techniques guaranteed the success in the search of IS*6110* in GC1237 genome.
- Sixteen copies of IS*6110* have DR indicating that its presence was due to a transposition event. In two of the other 3 copies recombination between two adjacent copies of IS*6110* was observed and the presence of the last one is implicated in genomic reorganization.
- According to the locations, four of the nineteen copies could act as a mobile promoter: one of them is located 297 bp upstream of the Rv1468c gene, another 138 bp upstream of *PE27A* gene, the third one 41 bp upstream Rv3427c gene and the last one is located 31 bp upstream of the essential gene Rv2179c.
- The copy of IS*6110* locates 31 bp upstream of the essential gene Rv2179c and is specific of GC1237 strain.
- The study of all the copies of IS*6110* in *M. tuberculosis* GC1237 allowed us to design an assay based on its specific location of IS*6110*.
- The multiplex PCR facilitates the identification of GC1237 quickly and easily.

References

1. 2006. European Concerted Action on new generation genetic markers and techniques for the epidemiology and control of Tuberculosis. Beijing/W genotype Mycobacterium tuberculosis and drug resistance. Emerg Infect Dis **12:**736-743.
2. **Agerton, T. B., S. E. Valway, R. J. Blinkhorn, K. L. Shilkret, R. Reves, W. W. Schluter, B. Gore, C. J. Pozsik, B. B. Plikaytis, C. Woodley, and I. M. Onorato.** 1999. Spread of strain W, a highly drug-resistant strain of Mycobacterium tuberculosis, across the United States. Clin Infect Dis **29:**85-92; discussion 93-5.
3. **Alonso, M., N. Alonso Rodriguez, C. Garzelli, M. Martinez Lirola, M. Herranz, S. Samper, M. J. Ruiz Serrano, E. Bouza, and D. Garcia de Viedma.** Characterization of Mycobacterium tuberculosis Beijing isolates from the Mediterranean area. BMC Microbiol **10:**151.
4. **Beggs, M. L., K. D. Eisenach, and M. D. Cave.** 2000. Mapping of IS6110 insertion sites in two epidemic strains of Mycobacterium tuberculosis. J Clin Microbiol **38:**2923-8.
5. **Bifani, P. J., B. Mathema, N. E. Kurepina, and B. N. Kreiswirth.** 2002. Global dissemination of the Mycobacterium tuberculosis W-Beijing family strains. Trends Microbiol **10:**45-52.
6. **Bifani, P. J., B. Mathema, Z. Liu, S. L. Moghazeh, B. Shopsin, B. Tempalski, J. Driscol, R. Frothingham, J. M. Musser, P. Alcabes, and B. N. Kreiswirth.** 1999. Identification of a W variant outbreak of Mycobacterium tuberculosis via population-based molecular epidemiology. Jama **282:**2321-7.
7. **Bifani, P. J., B. B. Plikaytis, V. Kapur, K. Stockbauer, X. Pan, M. L. Lutfey, S. L. Moghazeh, W. Eisner, T. M. Daniel, M. H. Kaplan, J. T. Crawford, J. M. Musser, and B. N. Kreiswirth.** 1996. Origin and interstate spread of a New York City multidrug-resistant Mycobacterium tuberculosis clone family. JAMA **275:**452-7.
8. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res **7:**1513-23.
9. **Brosch, R., W. J. Philipp, E. Stavropoulos, M. J. Colston, S. T. Cole, and S. V. Gordon.** 1999. Genomic analysis reveals variation between Mycobacterium tuberculosis H37Rv and the attenuated M. tuberculosis H37Ra strain. Infect Immun **67:**5768-74.
10. **Brudey, K., J. R. Driscoll, L. Rigouts, W. M. Prodinger, A. Gori, S. A. Al-Hajoj, C. Allix, L. Aristimuno, J. Arora, V. Baumanis, L. Binder, P. Cafrune, A. Cataldi, S. Cheong, R. Diel, C. Ellermeier, J. T. Evans, M. Fauville-Dufaux, S. Ferdinand, D. Garcia de Viedma, C. Garzelli, L. Gazzola, H. M. Gomes, M. C. Guttierez, P. M. Hawkey, P. D. van Helden, G. V. Kadival, B. N. Kreiswirth, K. Kremer, M. Kubin, S. P. Kulkarni, B. Liens, T. Lillebaek, M. L. Ho, C. Martin, I. Mokrousov, O. Narvskaia, Y. F. Ngeow, L. Naumann, S. Niemann, I. Parwati, Z. Rahim, V. Rasolofo-Razanamparany, T. Rasolonavalona, M. L. Rossetti, S. Rusch-Gerdes, A. Sajduda, S. Samper, I. G. Shemyakin, U. B. Singh, A. Somoskovi, R. A. Skuce, D. van Soolingen, E. M. Streicher, P. N. Suffys, E. Tortoli, T. Tracevska, V. Vincent, T. C. Victor, R. M. Warren, S. F. Yap, K. Zaman, F. Portaels, N. Rastogi, and C. Sola.** 2006. Mycobacterium tuberculosis complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol **6:**23.
11. **Caminero, J. A., M. J. Pena, M. I. Campos-Herrero, J. C. Rodriguez, I. Garcia, P. Cabrera, C. Lafoz, S. Samper, H. Takiff, O. Afonso, J. M. Pavon, M. J. Torres, D. van Soolingen, D. A. Enarson, and C. Martin.** 2001. Epidemiological evidence of the spread of a Mycobacterium tuberculosis strain of the Beijing genotype on Gran Canaria Island. Am J Respir Crit Care Med **164:**1165-70.
12. **Caws, M., G. Thwaites, S. Dunstan, T. R. Hawn, N. T. Lan, N. T. Thuong, K. Stepniewska, M. N. Huyen, N. D. Bang, T. H. Loc, S. Gagneux, D. van Soolingen, K. Kremer, M. van der Sande, P. Small, P. T. Anh, N. T. Chinh, H. T. Quy, N. T.**

Duyen, D. Q. Tho, N. T. Hieu, E. Torok, T. T. Hien, N. H. Dung, N. T. Nhu, P. M. Duy, N. van Vinh Chau, and J. Farrar. 2008. The influence of host and bacterial genotype on the development of disseminated disease with Mycobacterium tuberculosis. PLoS Pathog **4:**e1000034.

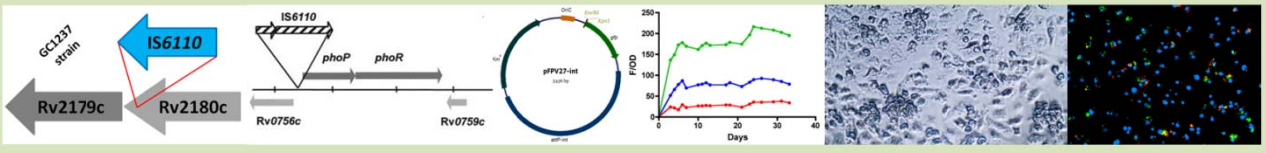
13. **Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, 3rd, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell.** 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature **393:**537-44.
14. **Constant, P., E. Perez, W. Malaga, M. A. Laneelle, O. Saurel, M. Daffe, and C. Guilhot.** 2002. Role of the pks15/1 gene in the biosynthesis of phenolglycolipids in the Mycobacterium tuberculosis complex. Evidence that all strains synthesize glycosylated p-hydroxybenzoic methyl esters and that strains devoid of phenolglycolipids harbor a frameshift mutation in the pks15/1 gene. J Biol Chem **277:**38148-58.
15. **Domenech, P., G. S. Kolly, L. Leon-Solis, A. Fallow, and M. B. Reed.** Massive gene duplication event among clinical isolates of the Mycobacterium tuberculosis W/Beijing family. J Bacteriol **192:**4562-70.
16. **Dou, H. Y., F. C. Tseng, C. W. Lin, J. R. Chang, J. R. Sun, W. S. Tsai, S. Y. Lee, I. J. Su, and J. J. Lu.** 2008. Molecular epidemiology and evolutionary genetics of Mycobacterium tuberculosis in Taipei. BMC Infect Dis **8:**170.
17. **Fang, Z., C. Doig, D. T. Kenna, N. Smittipat, P. Palittapongarnpim, B. Watt, and K. J. Forbes.** 1999. IS6110-mediated deletions of wild-type chromosomes of Mycobacterium tuberculosis. J Bacteriol **181:**1014-20.
18. **Fang, Z., and K. J. Forbes.** 1997. A Mycobacterium tuberculosis IS6110 preferential locus (ipl) for insertion into the genome. J Clin Microbiol **35:**479-81.
19. **Gagneux, S., K. DeRiemer, T. Van, M. Kato-Maeda, B. C. de Jong, S. Narayanan, M. Nicol, S. Niemann, K. Kremer, M. C. Gutierrez, M. Hilty, P. C. Hopewell, and P. M. Small.** 2006. Variable host-pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A **103:**2869-73.
20. **Glynn, J. R., J. Whiteley, P. J. Bifani, K. Kremer, and D. van Soolingen.** 2002. Worldwide occurrence of Beijing/W strains of Mycobacterium tuberculosis: a systematic review. Emerg Infect Dis **8:**843-9.
21. **Gutacker, M. M., B. Mathema, H. Soini, E. Shashkina, B. N. Kreiswirth, E. A. Graviss, and J. M. Musser.** 2006. Single-nucleotide polymorphism-based population genetic analysis of Mycobacterium tuberculosis strains from 4 geographic sites. J Infect Dis **193:**121-8.
22. **Hanekom, M., N. C. Gey van Pittius, C. McEvoy, T. C. Victor, P. D. Van Helden, and R. M. Warren.** Mycobacterium tuberculosis Beijing genotype: a template for success. Tuberculosis (Edinb) **91:**510-23.
23. **Jacobs, W. R., Jr., G. V. Kalpana, J. D. Cirillo, L. Pascopella, S. B. Snapper, R. A. Udani, W. Jones, R. G. Barletta, and B. R. Bloom.** 1991. Genetic systems for mycobacteria. Methods Enzymol **204:**537-55.
24. **Kang, H. Y., T. Wada, T. Iwamoto, S. Maeda, Y. Murase, S. Kato, H. J. Kim, and Y. K. Park.** Phylogeographical particularity of the Mycobacterium tuberculosis Beijing family in South Korea based on international comparison among surrounding countries. J Med Microbiol.
25. **Kanji, A., Z. Hasan, M. Tanveer, R. Laiq, and R. Hasan.** Occurrence of RD149 and RD152 deletions in Mycobacterium tuberculosis strains from Pakistan. J Infect Dev Ctries **5:**106-13.
26. **Kremer, K., J. R. Glynn, T. Lillebaek, S. Niemann, N. E. Kurepina, B. N. Kreiswirth, P. J. Bifani, and D. van Soolingen.** 2004. Definition of the Beijing/W lineage of

Mycobacterium tuberculosis on the basis of genetic markers. J Clin Microbiol **42:**4040-9.

27. **Kurepina, N., E. Likhoshvay, E. Shashkina, B. Mathema, K. Kremer, D. van Soolingen, P. Bifani, and B. N. Kreiswirth.** 2005. Targeted hybridization of IS6110 fingerprints identifies the W-Beijing Mycobacterium tuberculosis strains among clinical isolates. J Clin Microbiol **43:**2148-54.
28. **Kurepina, N. E., S. Sreevatsan, B. B. Plikaytis, P. J. Bifani, N. D. Connell, R. J. Donnelly, D. van Sooligen, J. M. Musser, and B. N. Kreiswirth.** 1998. Characterization of the phylogenetic distribution and chromosomal insertion sites of five IS6110 elements in Mycobacterium tuberculosis: non-random integration in the dnaA-dnaN region. Tuber Lung Dis **79:**31-42.
29. **Lam, J. T., K. Y. Yuen, P. L. Ho, X. H. Weng, W. H. Zhang, S. Chen, and W. C. Yam.** Truncated Rv2820c enhances mycobacterial virulence ex vivo and in vivo. Microb Pathog **50:**331-5.
30. **Lopez, B., D. Aguilar, H. Orozco, M. Burger, C. Espitia, V. Ritacco, L. Barrera, K. Kremer, R. Hernandez-Pando, K. Huygen, and D. van Soolingen.** 2003. A marked difference in pathogenesis and immune response induced by different Mycobacterium tuberculosis genotypes. Clin Exp Immunol **133:**30-7.
31. **Maeda, S., T. Wada, T. Iwamoto, Y. Murase, S. Mitarai, I. Sugawara, and S. Kato.** Beijing family Mycobacterium tuberculosis isolated from throughout Japan: phylogeny and genetic features. Int J Tuberc Lung Dis **14:**1201-4.
32. **Mendiola, M. V., C. Martin, I. Otal, and B. Gicquel.** 1992. Analysis of the regions responsible for IS6110 RFLP in a single Mycobacterium tuberculosis strain. Res Microbiol **143:**767-72.
33. **Millan-Lou, M. I., H. Alonso, P. Gavin, M. Hernandez-Febles, M. I. Campos-Herrero, R. Copado, F. Canas, K. Kremer, J. A. Caminero, C. Martin, and S. Samper.** Rapid test for identification of a highly transmissible Mycobacterium tuberculosis Beijing strain of sub-Saharan origin. J Clin Microbiol **50:**516-8.
34. **Mokrousov, I., H. M. Ly, T. Otten, N. N. Lan, B. Vyshnevskyi, S. Hoffner, and O. Narvskaya.** 2005. Origin and primary dispersal of the Mycobacterium tuberculosis Beijing genotype: clues from human phylogeography. Genome Res **15:**1357-64.
35. **Otal, I., A. B. Gomez, K. Kremer, P. de Haas, M. J. Garcia, C. Martin, and D. van Soolingen.** 2008. Mapping of IS6110 insertion sites in Mycobacterium bovis isolates in relation to adaptation from the animal to human host. Vet Microbiol **129:**333-41.
36. **Parish, T., and N. G. Stoker.** 2002. The common aromatic amino acid biosynthesis pathway is essential in Mycobacterium tuberculosis. Microbiology **148:**3069-77.
37. **Plikaytis, B. B., J. L. Marden, J. T. Crawford, C. L. Woodley, W. R. Butler, and T. M. Shinnick.** 1994. Multiplex PCR assay specific for the multidrug-resistant strain W of Mycobacterium tuberculosis. J Clin Microbiol **32:**1542-6.
38. **Prod'hom, G., B. Lagier, V. Pelicic, A. J. Hance, B. Gicquel, and C. Guilhot.** 1998. A reliable amplification technique for the characterization of genomic DNA sequences flanking insertion sequences. FEMS Microbiol Lett **158:**75-81.
39. **Reed, M. B., P. Domenech, C. Manca, H. Su, A. K. Barczak, B. N. Kreiswirth, G. Kaplan, and C. E. Barry, 3rd.** 2004. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. Nature **431:**84-7.
40. **Reed, M. B., S. Gagneux, K. Deriemer, P. M. Small, and C. E. Barry, 3rd.** 2007. The W-Beijing lineage of Mycobacterium tuberculosis overproduces triglycerides and has the DosR dormancy regulon constitutively upregulated. J Bacteriol **189:**2583-9.
41. **Safi, H., P. F. Barnes, D. L. Lakey, H. Shams, B. Samten, R. Vankayalapati, and S. T. Howard.** 2004. IS6110 functions as a mobile, monocyte-activated promoter in Mycobacterium tuberculosis. Mol Microbiol **52:**999-1012.
42. **Samper, S., I. Otal, M. C. Rubio, M. A. Vitoria, R. Gomez-Lus, and C. Martin.** 1993. [Application of RFLP to typing strains of Mycobacterium tuberculosis]. Enferm Infecc Microbiol Clin **11:**547-51.

43. **Sampson, S., R. Warren, M. Richardson, G. van der Spuy, and P. van Helden.** 2001. IS6110 insertions in Mycobacterium tuberculosis: predominantly into coding regions. J Clin Microbiol **39:**3423-4.
44. **Soto, C. Y., M. C. Menendez, E. Perez, S. Samper, A. B. Gomez, M. J. Garcia, and C. Martin.** 2004. IS6110 mediates increased transcription of the phoP virulence gene in a multidrug-resistant clinical isolate responsible for tuberculosis outbreaks. J Clin Microbiol **42:**212-9.
45. **Staden, R.** 1996. The Staden sequence analysis package. Mol Biotechnol **5:**233-41.
46. **Theus, S., K. Eisenach, N. Fomukong, R. F. Silver, and M. D. Cave.** 2007. Beijing family Mycobacterium tuberculosis strains differ in their intracellular growth in THP-1 macrophages. Int J Tuberc Lung Dis **11:**1087-93.
47. **Tsolaki, A. G., S. Gagneux, A. S. Pym, Y. O. Goguet de la Salmoniere, B. N. Kreiswirth, D. Van Soolingen, and P. M. Small.** 2005. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of Mycobacterium tuberculosis. J Clin Microbiol **43:**3185-91.
48. **van Soolingen, D., P. E. de Haas, P. W. Hermans, and J. D. van Embden.** 1994. DNA fingerprinting of Mycobacterium tuberculosis. Methods Enzymol **235:**196-205.
49. **Zhang, M., J. Gong, Z. Yang, B. Samten, M. D. Cave, and P. F. Barnes.** 1999. Enhanced capacity of a widespread strain of Mycobacterium tuberculosis to grow in human macrophages. J Infect Dis **179:**1213-7.

CHAPTER 2: Study of the promoter activity of IS*6110* in different locations of MTBC



Introduction

Promoter activity of IS*6110* insertion sequence

In spite of studies that show that IS*6110* transposes in microaerobic atmosphere (7), in patient to patient transmission (17) and in the adaptation to a particular host (13), the physiological role and impact of specific IS*6110* insertions on the biology of bacilli are not well known. However, we can affirm that this sequence induces duplications, inversions, deletions, and rearrangements in the bacteria genome, all of them essentials changes for the genome plasticity of the members of MTBC (10).

Moreover, this element has been related with a promoter activity inducing transcription of downstream genes, during infection (16) or not (2, 16, 18). IS*6110* can upregulate downstream genes through an outward-directed promoter in its 3'end. The promoter region of IS*6110*, called OP6110, is localized within a 110 bp fragment adjacent to the right terminal IR (16). This promoter activity was demonstrated in several natural insertions of IS*6110* such as the located upstream Rv2280 gene in H37Rv strain (16), the located upstream Rv1468c gene in 210 Beijing strain (16) and the copy located 75 bp upstream *phoP* gene in the MDR *M. bovis* strain MBZ (18). These three locations of IS*6110* are widely explained in next section.

Studied locations of IS*6110* where is acting as a mobile promoter

The genomic data of the strains H37Rv and 210, which is member of Beijing family, were examined for locations where IS*6110* was inserted in the same orientation as, and close enough to, a downstream gene potentially to function as a promoter (16). The 210 strain has a copy of IS*6110* inserted 297 bp upstream the Rv1468c (*PE_PGRS29*) gene and in H37Rv a copy of IS*6110* is located 191 bp of Rv2280 gene (Figure 16A). The effect of IS*6110* (OP6110) in these two genes was studied both in broth and inside monocytes (16). The expression of the two genes was analyzed by qRT-PCR and compared between both strains. During exponential growth in broth, transcription levels of Rv1468c and Rv2280 were similar in both strains indicating that, in that condition, an upstream copy of IS*6110* had minimal overall effects on gene expression. However, in monocytes, IS*6110* was associated with three to fivefold higher levels of transcription of both downstream genes (Figure 16B). Moreover, the transcription between IS*6110* and these downstream genes was demonstrated (16).

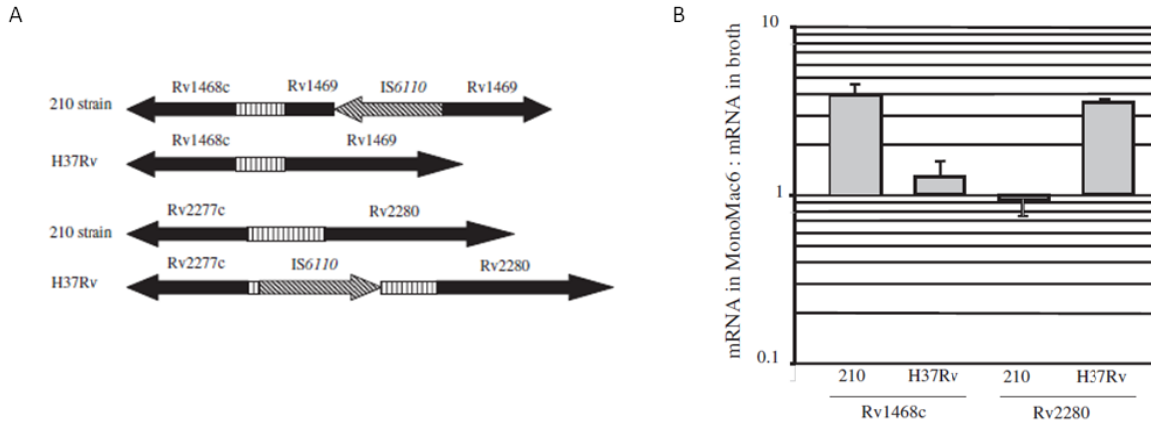


Figure 16: Promoter activity of IS*6110* in two natural insertion sites of IS*6110*. (A) Schematic representation of two IS*6110* insertion sites, one in *M. tuberculosis* 210 (upstream Rv1468c gene) and another one in *M. tuberculosis* H37Rv (upstream Rv2080 gene). The genomic arrangements are compared between both strains. Arrows indicate the direction of gene transcription. Intergenic regions are shown as bars with vertical stripes. **(B)** OP6110 activity at both genomic sites. Expression of mRNA during growth in MonoMac6 cells, compared with growth in broth. Adapted from (16).

The MDR strain *M. bovis* B (MBZ or B strain) responsible for large tuberculosis outbreaks in Spain has a copy of IS*6110* located in the promoter region of *phoP* gene (18) (Figure 17A). Due to the drug resistant condition of the strain, the effect of this IS*6110* on the transcription of *phoP* gene was studied with two recombinant plasmids: pSO5 (a mycobacterial replicative plasmid carrying the entire coding sequence of *phoP* gene and 1 kb of its promoter region) (15) and pSO7 (containing *phoP* gene with its promoter region which include the entire sequence of IS*6110*) (Figure 17B) inside *M. smegmatis* mc^2155 (18). The primer extension analysis was carried out with RNA isolated from both strains and in both cases two putative transcription start points were detected (Figure 17C and D). In *M.smegmatis* pSO5, tsp1 and tsp2 were located 121 and 57 bp upstream from the ATG start codon of *phoP*, respectively; and in *M. smegmatis* pSO7, tsp1 and tsp2 were located 136 and 57 bp upstream from the ATG starting codon of *phoP*, respectively; tsp1 from the MBZ strain (tsp1*) was located inside IS*6110* (18) (Figure 17C and D).

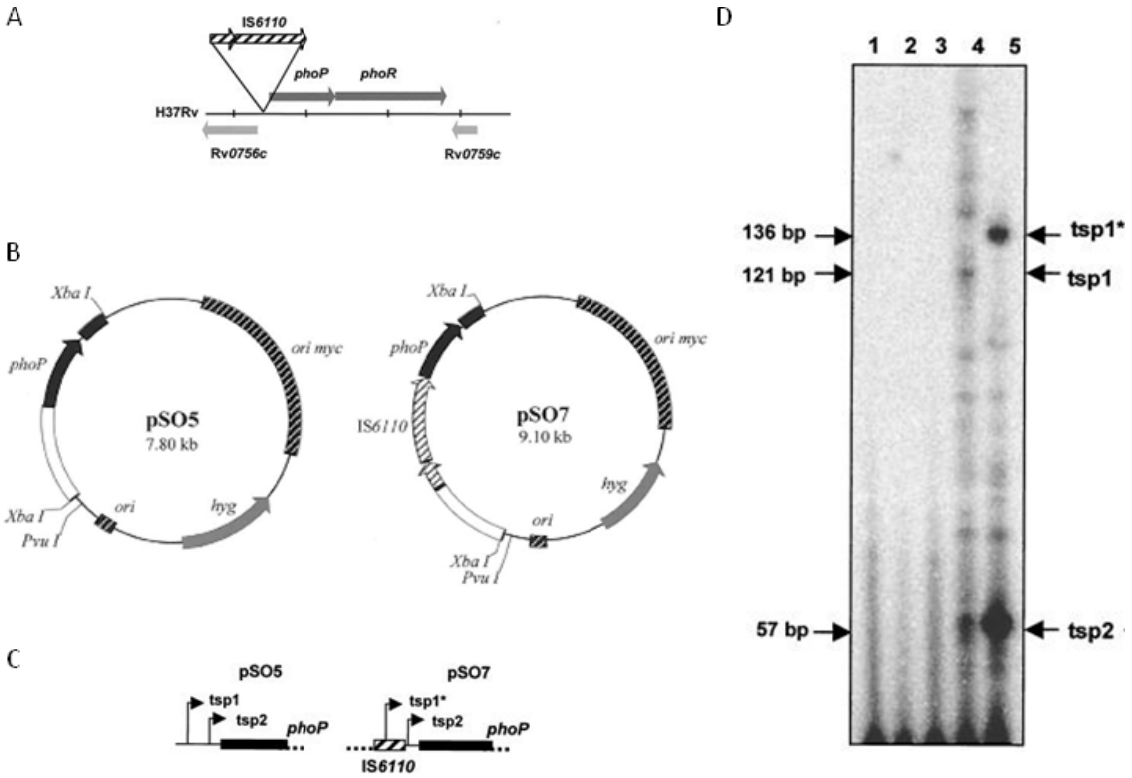


Figure 17: Schematic summary of the study of the IS*6110* located in the promoter region of *phoP* gene in MBZ strain (A) Schematic representation showing the position of IS*6110* in the *phoP* promoter region in the *M. bovis* B strain. **(B) Plasmids pSO5** (15) which contains the entire nucleotide sequence of *phoP* gene and 1 kb of its promoter region and **pSO7** which contains *phoP* gene with its promoter region including an entire sequence of IS*6110*. **(C)** Schematic figure showing the locations of the putative transcription start points of potential *phoP* promoters for *M. smegmatis* mc^2155 pSO5 (tsp1 and tsp2) and *M. smegmatis* mc^2155 pSO7 (tsp1*and tsp2). **(D)** Products of primer extension analysis using total RNA from the different *M. smegmatis* mc^2155 strains. No *phoP transcripts* were detected with the control strains (lane1, lane2 and lane3), *M. smegmatis* mc^2155:pSO5 (lane 4), *M. smegmatis* mc^2155:pSO7 (lane 5). The locations of the transcription starting points (tsp1, tsp1* and tsp2) are indicated. Adapted from (18).

Furthermore, it was observed that the MBZ strain grew slowly and displays peculiar colony morphology (smaller colonies than other *M. bovis* strains) and it was suggested that the insertion of IS*6110* at this particular locus could be responsible for these changes (18). With these premises, a better understanding of the possible relation between the promoter activity of this copy of IS*6110* and the phenotype of this strain became one of the major objectives of our study.

The selected copies of IS*6110* to be studied: The copy upstream Rv2179c of GC1237 strain and the insertion A1 common of all Beijing strains

Another of our major objective was to study the promoter activity of IS*6110* in the clinical isolate *M. tuberculosis* GC1237. As it was explained in Chapter 1, four of the 19 copies of IS*6110* localized in GC1237 could act as a mobile promoter as the orientation is the same as and are close enough to the neighboring gene. We focused the study in the copy of IS*6110* located 31 bp upstream of the essential gene Rv2179c and close to the essential gene *aroG* which is involved in the common aromatic amino acid biosynthesis in *M. tuberculosis* (14).

In addition to this, we decided to analyze the copy of IS*6110* located between *dnaA:dnaN* (insertion A1), inserted 460 bp upstream *dnaN*, as this copy could be relevant in the phenotype of Beijing strains as it is one of the common locations; for this reason, this copy was also study in detail.

Objectives

- To evaluate the promoter activity of three natural insertions of IS*6110* (before *dnaN* gene which is characteristic of all Beijing strains, before the gene Rv2179c in *M. tuberculosis* GC1237 Beijing strain and before *phoP* gene in *M. bovis* B strain) by using several GFP-constructions containing promoter regions.

- To study the promoter activity inside three different mycobacteria, the reference strain H37Rv, the clinical isolate GC1237 and *M. bovis* BCG, to demonstrate that this activity does not depend on the genetic background in MTBC.

- To verify the promoter activity of these three locations in the corresponding wild-type strain.

- To analyze these promoter activity inside macrophages with the GFP-constructions and with the wild-type strains.

Material and methods

Bacterial strains, culture media, and growth conditions

The clinical isolate *M. tuberculosis* GC1237, the reference *M. tuberculosis* H37Rv, *M. bovis* BCG and the *M. tuberculosis* H37Rv ΔphoP strains were used. Mycobacterial strains were grown in Middlebrook 7H9 broth supplemented with ADC and 0.05% Tween 80 or in Middlebrook 7H10 mediumBacto agar supplemented with OADC (Difco Laboratories, Detroit, Mich.) and 0.05% Tween 80 (9). Liquid cultures were grown to logarithmic and stationary phases to be used for macrophage infection in vitro and for mycobacterial RNA extraction.

E. coli XL1 was used for cloning experiments. Strains were grown in LB broth or LB agar plates. When necessary, media were supplemented with kanamycin (Km) (20μg/ml) or hygromicing (Hyg) (50μg/ml). The plasmid pFPV27-int (Figure 18A), derived from pFPV27 (1), was used for the construction of the strains expressing GFP.

Electrotransformation of *E. coli* and of mycobacteria

To prepare *E. coli* competent cells, bacteria were grown to an OD_{600nm} of 0.4 to 0.6. Then the growth was stopped for 30min on ice, and bacteria were washed twice in chilled-cold water, and once in chilled-cold 10% glycerol. Aliquots of 50µl were directly used or frozen at -80ºC. Aliquots of 50µl were electroporated with plasmid DNA in 0.2cm gap cuvettes (Bio-Rad) with a single pulse (2.5kV, 25µF, 200Ω) in a GenePulser Xcell™ (Bio-Rad). Cells were resuspended in LB to a final volume of 1ml and incubated for 1h at 37ºC if required before plating.

M. tuberculosis competent cells were prepared as described by Wards *et al.* (20). Bacteria were grown until an OD_{600nm} of 0.6 to 0.8. After two washes in 0.05% Tween-80 and one wash in 10% glycerol-0.05% Tween-80, cells were resuspended in 2ml of 10% glycerol-0.05% Tween-80. Aliquots of 400µl were electroporated with plasmid DNA in 0.2cm gap cuvettes (Bio-Rad) with a single pulse (2.5kV, 25µF, 1000Ω) in a GenePulser Xcell™ (Bio-Rad). Cells were recovered with 1ml of 7H9-ADC-0.05% Tween-80 and incubated for 24h at 37ºC, to express the antibiotic resistance, before plating several dilutions on plates containing the relevant antibiotic.

Construction of GFP-strains

- ***From the copy of IS6110 located upstream dnaN in GC1237 strain:***

To construct pFPVdnaN-int plasmid, primers dnaN-EcoRI and dnaN-kpnI (Table 6 Annex I) were used to amplify a 913 bp fragment from H37Rv strain. To construct pFPVISdnaN-int plasmid, ISA4-Eco and dnaN-kpnI (Table 6 Annex I) were used to amplified 713 bp segment from GC1237 strain. The PCR products were inserted in pFPV27-int, which had been cut with *EcoR*I and *Kpn*I enzymes (Figure 18B). The plasmids were sequenced and electroporated into H37Rv strain obtaining HFPVdnaN and HFPVISdnaN strains.

- ***From the IS6110 located upstream phoP gene in MBZ strain:***

Plasmids pFPVSO5-int and pFPVSO7-int (Figure 18C) were constructed in a previous work (not published). Briefly, to construct pFPVSO5-int plasmid, primers BCG2B-Eco and PhoBO3.1-Kpn (Table 6 Annex I) were used to amplify a 351 bp fragment from H37Rv strain. To construct pFPVSO7-int plasmid, primers ISA4-Eco and PhoBO3.1-Kpn (Table 6 Annex I) were used to amplify a 350 bp segment from MBZ strain. The PCR products were inserted in pFPV27-int that had been cut with *EcoR*I and *Kpn*I enzymes. Both plasmids were sequenced to verify the constructions.

In this work the plasmids were electroporated into H37Rv, GC1237 and BCG strains obtaining HFPVSO5, HFPVSO7, GFPVSO5, GFPVSO7, BFPVSO5 and BFPVSO7 strains respectively.

- ***From the IS6110 located upstream Rv2179c in GC1237 strain:***

To construct pFPV79c-int plasmid, primers Rv2179c-Eco and Rv2179c-Kpn (Table 6 Annex I) were used to amplify a 670 bp fragment from H37Rv strain. To construct pFPVIS79c-int plasmid, primers ISA4-Eco and Rv2179c-Kpn (Table 6 Annex I) were used to amplify a 210 bp segment from GC1237 strain. The PCR products were inserted in pFPV27-int, which had been cut with *EcoR*I and *Kpn*I enzymes (Figure 18D). To verify the constructions, both plasmids were sequenced with the corresponding primers. After that, they were electroporated into H37Rv, GC1237 and BCG strains obtaining HFPV79c, HFPVIS79c, GFPV79c, GFPVIS79c BFPV79c and BFPVIS79c strains.

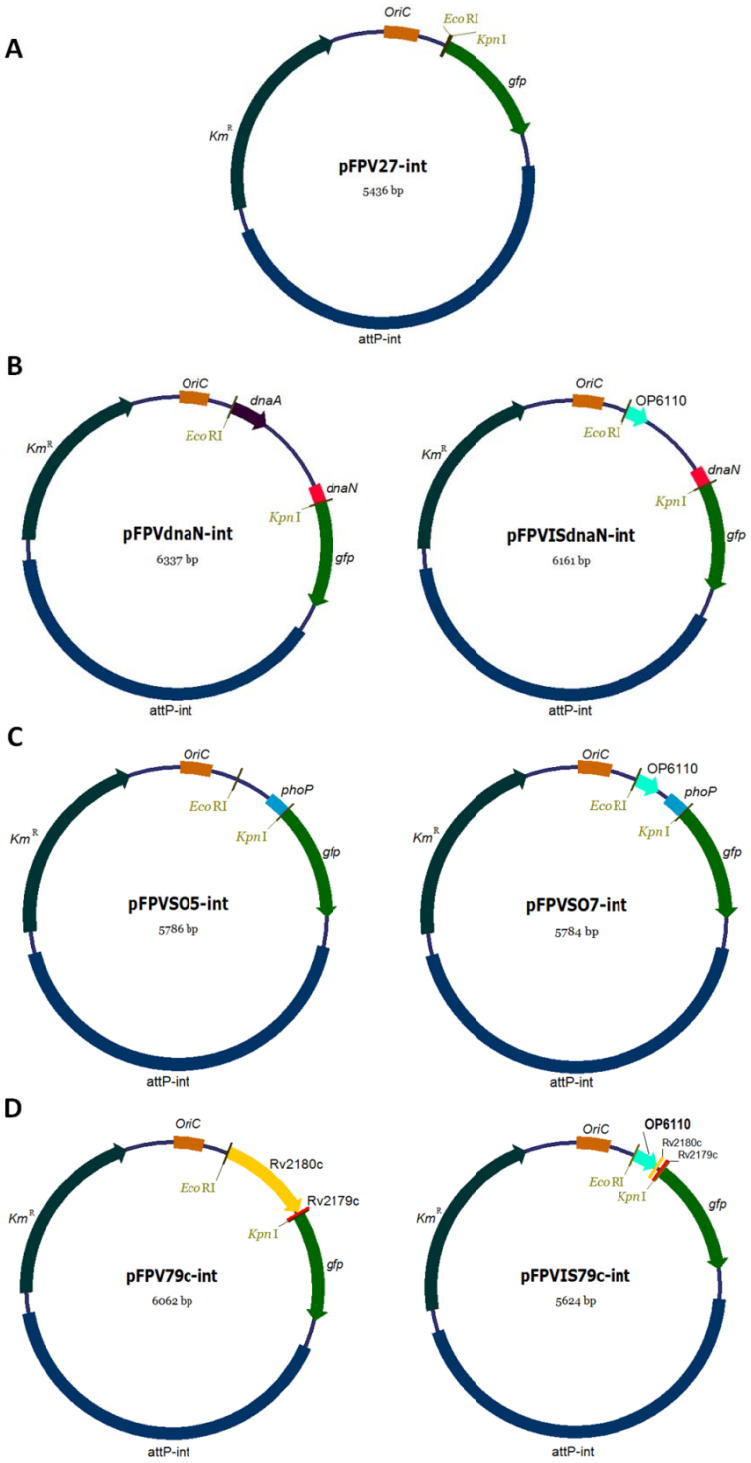


Figure 18: *gfp*-plasmids used in this work. (A) pFPV27-int which contains the promoterless *gfp* gene. **(B) pFPVdnaN-int** which contains a 913 bp from H37Rv strain upstream the *dnaN* gene and **pFPVISdanN-int** which contains a 713 bp fragment upstream the *dnaN* gene from GC1237 strain containing the promoter region of IS*6110* (OP6110). **(C) pFPVSO5-int** which contains 351 bp fragment upstream *phoP* gene of H37Rv strain and **pFPVSO7-int** which contains 350 bp fragment of MBZ strain upstream *phoP* gene containing OP6110. **(D) pFPV79c-int** which contains a 670 bp fragment from H37Rv strain upstream the Rv2179c gene and **pFPVIS79c-int** which contains a 210 bp segment from GC1237 strain upstream the Rv2179c gene containing OP6110.

Construction of pSO5k and PSO7k plasmids derived from pSO5 (15) and pSO7 (18)

The replicative plasmid pSO5k (Figure 19A) derived from pSO5 (15) (Figure 17B), explained previously in this chapter, was constructed in a previous work (8). Briefly, the kanamycin resistance marker (Km^R) from the plasmid pUC4K (Amersham Biosciences) was cloned into the *Pst*I restriction site of pSO5. This plasmid was used for the complementation of the *phoP* gene in H37Rv ΔphoP obtaining H37Rv ΔphoP pSO5k strain.

In this work pSO7k (Figure 19B), also replicative, derived from pSO7 (18) (described previously in this chapter) (Figure 17B) was constructed exactly as pSO5k. The plasmid was totally sequenced and electroporated inside H37Rv ΔphoP mutant obtaining H37Rv ΔphoP pSO7k strain.

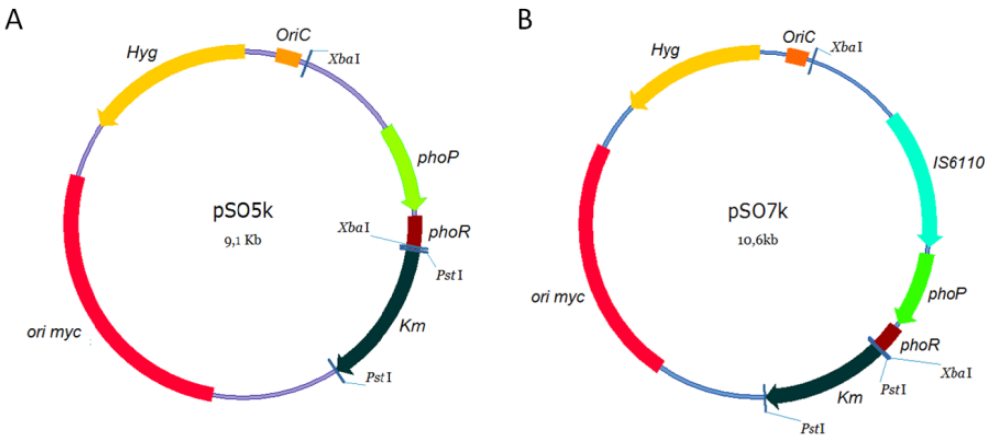


Figure 19: Replicative plasmids pSO5k and pSO7k. (A) pSO5k (adapted from (8)) is derived from pSO5 (15) and **(B)** pSO7k was constructed from pSO7 (18), the Km^R marker from the plasmid pUC4K was cloned into the *Pst*I restriction site of pSO7 (18).

Cell culture and infections

MH-S murine alveolar macrophages were obtained from HPA culture collections. Cells were grown in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 4mM L-glutamine. Infections were performed during 4h at a multiplicity of infection (MOI) of 10 bacteria per cell. After incubating with bacteria, cells were washed three times with PBS, and cultured in complete medium during the time indicated for each experiment.

Raw 264.7 cells were grown in RPMI 1640 medium supplemented with 10% of FBS. Infections were performed during 2h at MOI 3. After incubating with bacteria, cells were washed three times with phosphate buffered saline (PBS), and cultured in complete medium during the time indicated for each experiment.

- ***Experiments of replication:***

The studies of intracellular replication of *M. tuberculosis* were carried out in 24-well plate. Briefly, the infections were performed at a MOI of 1 bacterium per 20 cells and 2 wells were used per strain and per day of processing. After 4h, infection was terminated by removing the overlaying medium and washing three times with PBS before adding 1ml of fresh complete culture medium per well. At 0 (4h post-infection), 4 and 7 days, the number of intracellular bacteria was evaluated by plating appropriate dilutions of lysed macrophages (0.1% triton X-100) on 7H10-ADC solid medium. After 2-4 weeks, CFUs were evaluated.

- ***M. tuberculosis infection in vivo in mice:***

Intratracheal infection was performed with 100 CFU of bacteria in 50μl of PBS per mouse and 5 mice were used per strain. To deliver bacterial suspension, isoflurane anesthetized mice were orally intubated with a lachrymal olive luer-lock (UNIMED), 30mm in length and 0.6mm in diameter. Four weeks post-infection, lungs from each animal were harvested and placed in PBS for bacterial burden evaluation. To analyze bacterial replication, lungs were homogenized using GentleMacs homogeneizer (Miltenyi Biotec) and CFU counted by plating serial dilutions on 7H11-ADC solid medium.

The protocol for animal handling was previously approved by University of Zaragoza Animal Ethics Committee.

Isolation of RNA from mycobacteria

- ***Extracellular conditions:***

M. tuberculosis strains were grown at 37ºC until the desired OD_{600} under aerobic conditions. The RNA from bacterial pellet was stabilized using the RNAprotect Bacteria Reagent (QIAGEN) following manufacturer's recommendations. Cells were resuspended in 1ml acid phenol:chloroform (5:1) and 0.4ml lysis buffer (0.5% SDS, 20mM NaAc, 0.1mM EDTA) and transferred to 2ml Lysing Matrix B screw-cap tubes containing 0.1mm silica spheres (Q-BIOgene). Cells were disrupted by three 30s pulses in a FastPrep homogenizer (Q-BIOgene). After centrifugation, RNA from the supernatant was further extracted with 0.9ml chloroform:isoamyl alcohol (24:1). Total

RNA was precipitated with NaAc/isopropanol and washed with 70% ethanol. RNA pellet was treated with RNase-free DNase (Ambion), and the DNA free RNA was then further purified using and RNeasy kit (Qiagen). DNA contamination was ruled out by lack of amplification products after 35 cycles of PCR and the integrity of the RNA from the different strains were checked by gel electrophoresis on a 1% agarose gel. Purified RNA was kept at -80ºC until further use.

- ***Intracellular conditions:***

Intracellular mycobacteria RNA extraction was performed with a modified protocol based on the one described by Fontan *et al.* (5). Briefly, 12.10^6 MH-S cells per flask were seeded in $150cm^2$ flasks. An appropriate volume of liquid mycobacteria culture was suspended in complete medium and added to each flask containing cells. After incubation, cells were washed with PBS to remove extracellular bacteria, and cultured in complete medium for 0 and 48h post-infection, when they were treated to extract intracellular mycobacteria. With this aim, cells were lysed and homogeneized during 5min adding 10ml per flask of GTC buffer (25mM sodium citrate, 4M guanidine thiocyanate, 0.5% N-lauryl sarcosine, 0.125M 2-mercaptoethanol and 0.5% Tween 80, pH 7.0). Next, samples were collected in 15ml centrifuge tubes and centrifuged for one hour at 4000rpm. Bacterial pellets were washed with GTC buffer and centrifuged again for one hour at 4000rpm. After that, dry bacterial pellets were treated as described above to extract mycobacterial RNA.

RT-PCR

Reverse transcription-PCR (RT-PCR) was carried out in two steps. RT was carried out with Expand Reverse Transcriptase (Roche) using 1μg RNA as the template and the appropriate reverse primer. Reaction mixtures were incubated at 42ºC for 90min. RT products were the subjected to PCR amplification, using TaqGold polymerase (Roche) as we previously described with the appropriate primers. Samples were analyzed by electrophoresis on a 1% agarose gel.

qRT-PCR

cDNA libraries from mycobacteria were constructed as follows. One μm of RNA was mixed with 25pmol of random hexanucleotides primers (Sigma) and 50U of Expand Reverse Trancriptase (Roche) in a final volume of 20μl. Reaction mixtures were incubated at 65ºC for 10min and then at 42ºC for 90min. The gene expressions were measured and normalized with respect to the levels of *rrnaP1* mRNA by quantitative real-time PCR (qRT-PCR). qRT-PCR was carried out in a StepOne Plus (Applied Biosystems) instrument, using the cDNA generated by RT from 25ng of RNA as a

template, 1X Power SYBER green PCR master mix (Applied Biosystems), and the appropriated primers, each at a concentration of 250nM. The PCR program involved an initial denaturation step for 10min at 95ºC, followed by 40cycles at 95ºC for 15s and 60ºC for 1min. The specificity of the PCR products was confirmed by the loss of fluorescence at a single temperature, when the double-stranded DNA melted to single-stranded DNA.

Protein extraction from mycobacteria

- ***Extraction of cellular proteins:***

Cell-free protein extracts of mycobacteria were prepared from early log-phase cultures grown in Middlebrook 7H9-ADC-0.05% Tween-80. Cells from 50ml of culture were pelleted by centrifugation (3500rpm for 15min at 4ºC). Mycobacterial cell pellets were washed twice with PBS and then resuspended in cold PBS. Mycobacterial suspensions were transferred to tubes containing glass beads (Qbiogene) and were disrupted by mechanical traction (Fast-prep instrument) in two cycles (15s at speed 6.5m/s and 15s at 4.0m/s, respectively) cooling the samples on ice 5min between the pulses. The proteins were centrifuged at 5000rpm for 15min at 16ºC and the supernatant containing whole-cell protein extracts was filtered through a 0.22µm-pore-size low protein binding filter (Pall).

- ***Extraction of culture filtrate proteins:***

Mycobacterial strains were cultured in 7H9 Middlebrook medium supplemented with dextrose in the absence of Albumin. Cultures were grown 4 weeks and then centrifuged (3500rpm for 15min at 37ºC). The supernatant was filtered through 0.22µm filter Stericup systems (Millipore). The extracellular proteins were precipitated during 1h at 4ºC with the 10% v/v of Trichlore acetic acid. After that, pellets were collected by centrifugation (4000rpm 1h at 4ºC), washed twice with cold acetone and resuspended in 250µl of PBS 1x.

In both cases, proteins were quantified using the RC DC protein assay (BioRad) and stored at -80ºC.

Western blot analysis

Western blot analysis was carried out using standard procedures. Equivalent quantities of protein extracts were boiled (for 5min) in presence of 5µl of 150mM Tris/HCl pH 7.4, 3% SDS, 0.3mM sodium molybdate, 30mM sodium pyrophosphate, 30mM NaF, 30% glycerol, 30% mercaptoethanol and 0.06% bromophenol blue. Cellular proteins were separated by electrophoresis through a 15% polyacrylmide gel containing 0.1% SDS in

running buffer (25mM Tris, 192mM glycine, 3.4mM SDS) at constant amperage of 20mA/gel. Proteins were transferred to PVDF membranes (pre-activated with methanol) in transfer buffer (48mM Tris/HCl pH 8.3, 39mM glycine, 0.037% SDS, 20% methanol) for 1h at 20V with a semi-dry electrophoretic transfer cell (BioRad).

Blotted proteins were analyzed using specific antibodies: polyclonal antibodies against phoP protein (ZEU-immunotec) (12) and monoclonal antibody against ESAT-6 protein (Abcam).

The membrane was blocked with 5% skim powdered milk in buffer B (0.12M NaCl, 10mM Tris/HCl pH 8.0, 0.5% Tween-20 in PBS pH 7.4) for 30min at room temperature and mild shaking. Hereafter, the membrane was incubated with primary antibody anti-PhoP (1:2000) or anti-ESAT-6 (1:5000) in buffer A (2.5% skim powdered milk in buffer B) over night. Three washes with buffer B were performed previous to the incubation with commercial anti-rabbit (in the case of PhoP) or anti-mouse (for ESAT-6) secondary antibodies (1:20,000 in buffer A) conjugated with horseradish peroxidase, for 1h at room temperature and mild shaking. Then, the membrane was washed three times with buffer B. Immunocomplex detection was performed by incubation (for 90s) with a chemiluminiscent AP substrate (ImmobilonTM western, Millipore) and subsequent exposure (for 2min) to a photographic film. Molecular weights of the detected proteins were estimated relative to the standard protein ladder.

Analysis of GFP expression

- ***Extracellular conditions:***

To study the expression of the GFP protein in GFP strains, the fluorescence was measured by fluorometric method (Synergy_HT, BioTek) during the bacterial growth and normalized with OD_{600nm}. Briefly, 1ml of the mycobacterial broth cultures was collected and 200µl of the total volume were added to a 96-well plate and the GFP expression was measured at λ 485nm. After that, the OD_{600nm} of the broth culture was measured using the rest of the volume (800µl). The obtained value of fluorescence was normalized with the OD of the culture.

Fluorescence median intensity (FMI) of the different GFP strains was measured by Fluorescence activated cell sorting (FACS). Thus, 800µl of mycobacterial logarithmic cultures were centrifugated at 14000rpm 5min and the pellets were resuspended with 1ml of paraformaldehide (PFA) 4% and incubate 30min. After this time, the mixtures were centrifugated and resuspended in PBS. The cytometer used in this work was a FACSaria (BDBiosciences) and the analysis was performed with the software Weasel.

- ***Intracellular conditions:***

The GFP expression under intracellular conditions was studied by two methods. The first one, MH-S cells were infected with different GFP strains and H37Rv or GC1237 were used as reference. Infections were carried out in 24-well plates, where cells were seeded at 10^5 cells per well. Cells were collected at 0, 24, 48 and 72h post-infection and they were analyzed by FACS. Previously, cells were labeled with annexinV to discern between live and dead cells. Briefly, after washing collected cells once with AnnexinV binding buffer (ABB) (BD Biosciences), they were incubated for 15min at room temperature with annexinV APC-conjugated diluted in ABB, and then fixed with PFA 4% in calcium-containing buffer.

The second one was carried out as previously describe by Brodin *et al* (3). Briefly, Raw cells were seeded at a density of $3x10^5$ cells per well in 384-well plates in 50μl of medium. Adherent cells were then infected with bacterial suspensions at MOI of 3 and incubated for 2h. Cells were then washed three times with PBS and fixed with 1.5% formaldehyde for 10min, washed twice and stained with 5μg/ml DAPI dilactate (Sigma) in 0.1% triton X-100 (Sigma) in PBS. After that, confocal images were recorded on an automated fluorescent confocal microscope Opera (Evotec) using a 20X water immersion objective (NA 0.7). Subsequently, 405nm and 488nm lasers were used for excitation of DAPI and GFP, in two sequential exposures. Because the 20X magnification does not allow the acquisition of the entire 384-well plate, 6 fields within each well were recorded. For each field, two channels were recorded using two different cameras: the blue image for nuclei channel (Figure 20A) and the green image for GFP (Figure 20B). Each image was then processed using dedicated in-house image analysis software (IM 3.0). To determine the number of macrophages and delineate their outlines, cell centers were positioned using an algorithm that identifies the pixels with a maximum local intensity in the blue image. Cells that were too small based on cell radius area enclosed were discarded. The result of this step was the acquisition of the average cell surface that is the number of cells and their spatial localization (Figure 20D). To quantify the bacterial load, a similar procedure was carried out with green image (Figure 20B). In this case, the GFP-intensity of the different strains was one of the important parameters and for that, a threshold for GFP signal was determined with GFP-control strains. As DAPI stain did not allow us to define the contour of the cell, a macrophage was considered infected if labeled bacterium objects were distant of less than 5 pixels from the cell nucleus (Figure 20F). Proximity is a parameter manually set-up by the user given the fact that one cell has only one nucleus and a macrophage was labeled as infected if there were at least three contiguous green pixels overlapping the cell surface. Once images were processes

results are: the nuclei number, the percentage of infected cells (determined by the ratio of infected cells to the total number of cells in each field) and the intensity of infection (or GFP signal inside the cell).

When necessary, HFPV27, GFPV27 and BFPV27 strains containing the plasmid pFPV27-int. were used as control GFP-signal.

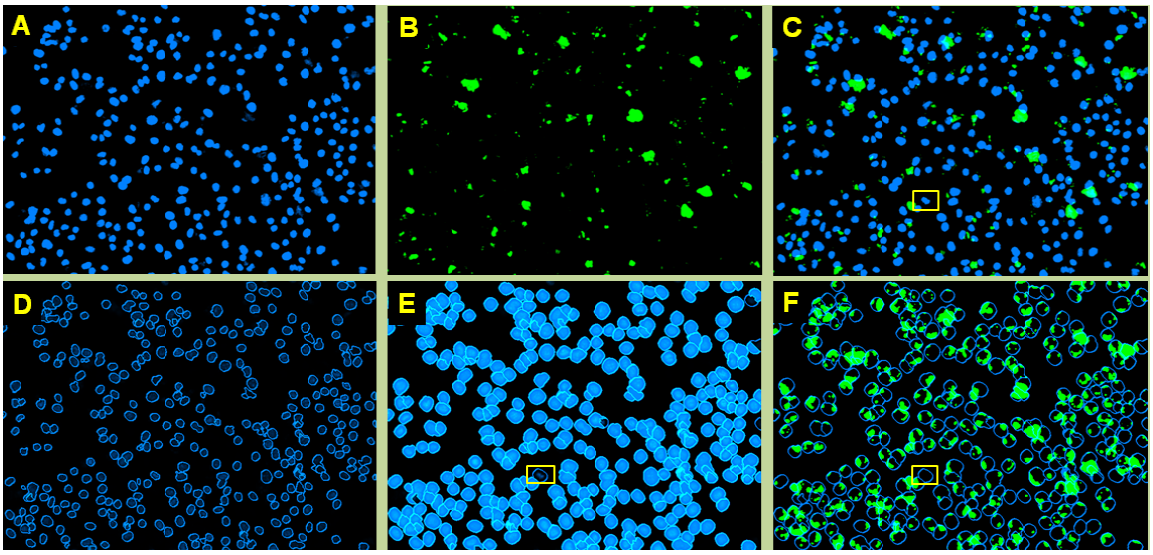


Figure 20: Representative images of infected raw cells (0h post-infection at MOI of 3) recorded by the automated confocal microscope Opera and its analysis in software IM 3.0. The first line of images (A, B and C) corresponds to opera confocal acquisition: **(A)** DAPI stain (nuclei channel). **(B)** GFP signal of bacteria **(C)** Two-color image (DAPI and GFP). Second line of images (D, E and F) corresponds to an example of image analysis carried out with IM 3.0: **(D)** Blue circles correspond to detected cell nuclei from the blue channel image, **(E)** blue-filled nuclei correspond to infected cells and **(F)** green-filled surfaces correspond to GFP bacteria detected as being proximal to blue-circled cell nuclei. An example of non-infected cell is squared in yellow in images C, E and F. Images span 0.450 x 0.340mm^2.

This part of the study was performed in a three-month stay in Institute Pasteur of Korea under supervision of Dr. Priscille Brodin, head of the Department of Biology of Intracellular Pathogens.

Results and Discussion

The interest on the different functions of IS*6110* in MTBC strains has been gaining force and importance. It has been related with genomic reorganizations, deletions and it is known that its insertion in an ORF causes loss of gene activity. This phenomenon can affect the regulation of the surrounding genes or can alter the regulation of the mycobacteria in general. There are several studies which defend that IS*6110* is maintained within the genome because it may confer some selective advantage. It is considered that the influence of the promoter activity of IS*6110* on the downstream genes is related to the distance among the gene and the 3'-end of this sequence. Thus, a promoter influence is possible within the range of 31 to 300 bp of distance among them.

With the aim of analyzing in depth this activity, three insertions of IS*6110* were selected. The first one is located 460 bp upstream *dnaN* gene. Although this distance is longer than the mentioned above, the study of this region was of interest as being common in Beijing genotype. The second insertion, located in Rv2180c was selected for being unique to *M. tuberculosis* GC1237 and for its proximity to the essential genes Rv2179c and *aroG*. The third one, before *phoP* gene in *M. bovis* B strain, was selected due to the possible relation of IS*6110* with the virulence of this strain as it has been demonstrated that phoP is related with the virulence of the mycobacteria.

1. The effect of the IS*6110* in *dnaN* gene in GC1237 strain

The copy of IS*6110* located upstream *dnaN* gene in GC1237 had minimal overall effects on the expression of this gene in broth

It is characteristic to find a copy of IS*6110* insertion sequence in the nucleotide 1594 (referred to H37Rv genome) in all Beijing strains (insertion A1). The presence of IS*6110* between *dnaA:dnaN* is expected to have some influence on the synchronization of the bacterial cell division (4). This region is currently considered a preferential locus, and multiple transposition events were described in several clinical isolates (19). Moreover this element could be inserted in both directions in this region (4, 19) having thus putatively a variable influence on the bacterial cell division.

As this locations of IS*6110* is common in Beijing genotype, its influence in this neighboring gene was studied. The expression of this gene was measured by qRT-PCR and compared between GC1237 and H37Rv strains. Thus, RNAs of both strains were extracted at logarithmic phase and the expression of *dnaN* was normalized to *rrnAP1* expression levels. Results demonstrated that in both strains the *dnaN* expression was similar (slightly greater in GC1237) (Figure 21), suggesting that IS*6110* in this location is not altering gene expression.

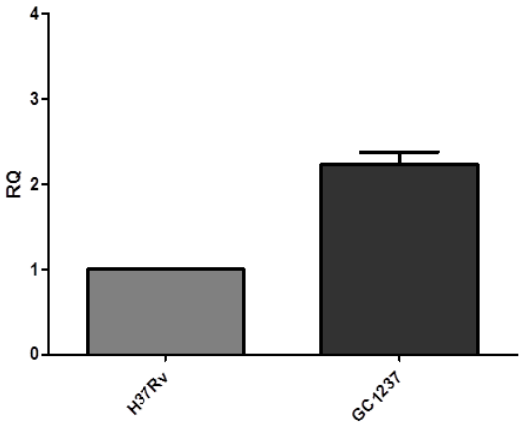


Figure 21: Relative quantification (RQ) of *dnaN* gene in GC1237 and H37Rv strains in broth. The expression of *dnaN* gene of GC1237 strain was compared with that in H37Rv strain at logarithmic phase. The expression of this gene was normalized to the levels of *rrnAP1* mRNA and the results are the means of three independent experiments; error bars indicate the standard deviations of the means.

To check that the data obtained with mRNA could be related with the IS*6110* located upstream *dnaN* gene, we constructed the integrative plasmids pFPVdnaN and pFPVISdnaN containing the promoterless *gfp* gene preceded by the region upstream *dnaN* gene in H37Rv strain (913 bp) or in GC1237 strain (713 bp), respectively. They were transformed into H37Rv obtaining HFPVdnaN and HFPVISdnaN strains. As control, we used H37Rv transformed with pFPV27-int which contains *gfp* gene without promoter (HFPV27 strain).

The GFP expression of these strains was measured by fluorometry during the growth in broth and as Figure 22 shows, similar GFP levels were observed in the two strains (HFPVdnaN and HFPVISdnaN), indicating that, an upstream copy of IS*6110* had minimal overall effects on this gene expression in broth. This result could be due to the IS*6110* is not located close enough to *dnaN* gene to act as a promoter.

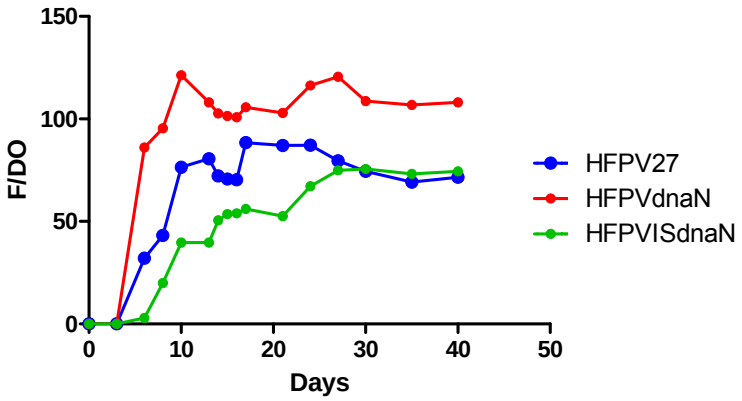


Figure 22: GFP expression of HFPVdnaN and HFPVISdnaN strains. Time course the evolution of the GFP-fluorescence of HFPV27, HFPVdnaN and HFPVISdnaN strains during the bacterial growth normalized with OD_{600nm}.

dnaN gene expression of H37Rv and GC1237 strains was also similar in intracellular conditions

To determine the effect of IS*6110* on *dnaN* gene expression during infection, bacterial RNA was obtained both from intracellular and broth growth conditions. The results showed (Figure 23) that *dnaN* gene expression is greatly increased in both strains when they are inside macrophages. Moreover, as Figure 23 shows the expression of this gene rises with the time of infection. This result corroborates the data obtained with GFP-strains and it could be indicating that in these conditions, the copy of IS*6110* located upstream *dnaN* gene in GC1237 and consequently in Beijing strains, causes minimal effect in *dnaN* gene expression.

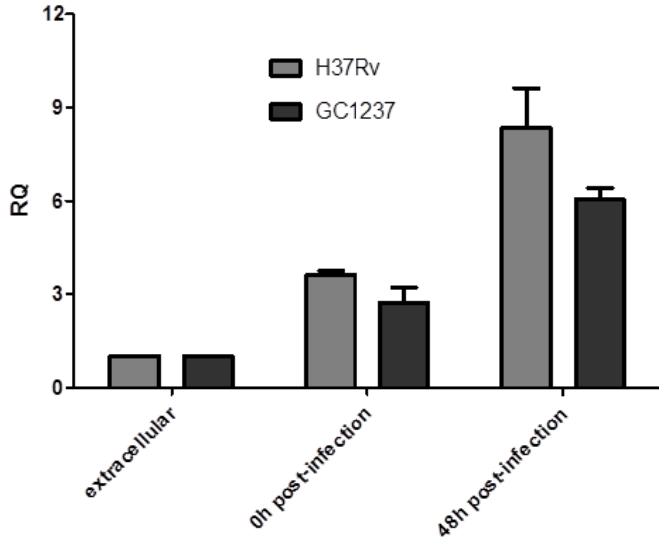


Figure 23: Relative quantification (RQ) of *dnaN* gene in H37Rv and GC1237 strains inside macrophages. The RQ of *dnaN* in intracellular conditions was analyzed at 0h and 48h post-infection in H37Rv and GC1237 strains and compared to extracellular conditions of both strains. The expression of *dnaN* was normalized to the levels of *rrnAP1* mRNA and the results are the means of three independent experiments; error bars indicate the standard deviations of the means.

These results do not rule out the possibility that this IS*6110* could act as a mobile promoter in other conditions such as inside human host.

2. The effect of the IS*6110* located upstream Rv2179c gene in GC1237 strain

Rv2179c gene expression is increased in GC1237 strain both in broth and intracellular conditions

As it has been describe before in this study, GC1237 strain presents a copy of IS*6110* upstream the Rv2179c gene. To analyze the effect of this element in Rv2179c, the expression of this gene was compared between GC1237 and H37Rv strains. Thus, both strains were cultured to mid-exponential growth phase and RNAs were extracted to generate cDNA libraries. The expression of Rv2179c was measured by qRT-PCR and normalized to *rrnAP1* expression levels. Results demonstrated that Rv2179c expression was about 6-fold higher in GC1237 than in H37Rv (Figure 24A), which suggests that IS*6110* in this location is up-regulating gene expression.

To determine the effect of IS*6110* on Rv2179c expression during infection, bacterial RNA was obtained both from intracellular and broth growth conditions. The results showed that Rv2179c expression rose dramatically in GC1237 infected macrophages, being 6 and 10-fold higher at 0h and 48h post-infection, respectively (Figure 24B). On the contrary, lower increase was observed in the case of H37Rv, confirming that the presence of IS*6110* correlated with higher expression of Rv2179c in the studied strain both under intracellular and extracellular conditions. These results confirm the promoter activity of IS*6110* in this location and suggest that this activity is upregulated inside macrophages.

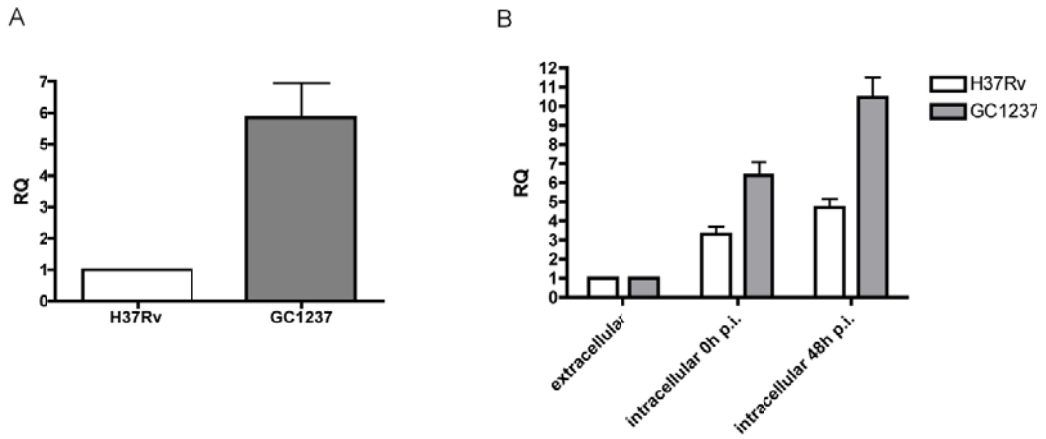


Figure 24: Quantification of Rv2179c in *M. tuberculosis* GC1237 and H37Rv by qRT-PCR. **(A)** Relative quantification (RQ) of Rv2179c expression in GC1237 with respect to H37Rv at logarithmic phase. **(B)** RQ of Rv2179c expression in intracellular conditions at 0h and 48h post-infection in GC1237 or H37Rv with respect to extracellular conditions. In both cases, the expression of Rv2179c was normalized to the levels of *rrnAP1* mRNA. Results are the means of three independent experiments; error bars indicate the standard deviations of the means.

Transcripts from IS*6110* to *aroG* and from Rv2180c to *aroG* were detected in GC1237 strain and in H37Rv strain, respectively

To determine if a transcript extended from IS*6110* or Rv2180c into downstream genes Rv2179c and *aroG*, RNA from broth cultures of GC1237 and H37Rv strains was extracted and RT-PCR was performed as previously described in this chapter. We obtained amplification products from Rv2180c to *aroG* in H37Rv strain (Figure 25A and C). These results indicate that the three genes are cotranscribed in an operon. In the case of GC1237 strain, amplification products were obtained from 3' end of IS*6110* to *aroG* suggesting that the 3' end of IS*6110*, Rv2179c and *aroG* genes are cotranscribed (Figure 25B and D). All the RT-PCR products were sequenced and the blast results verified their specificity.

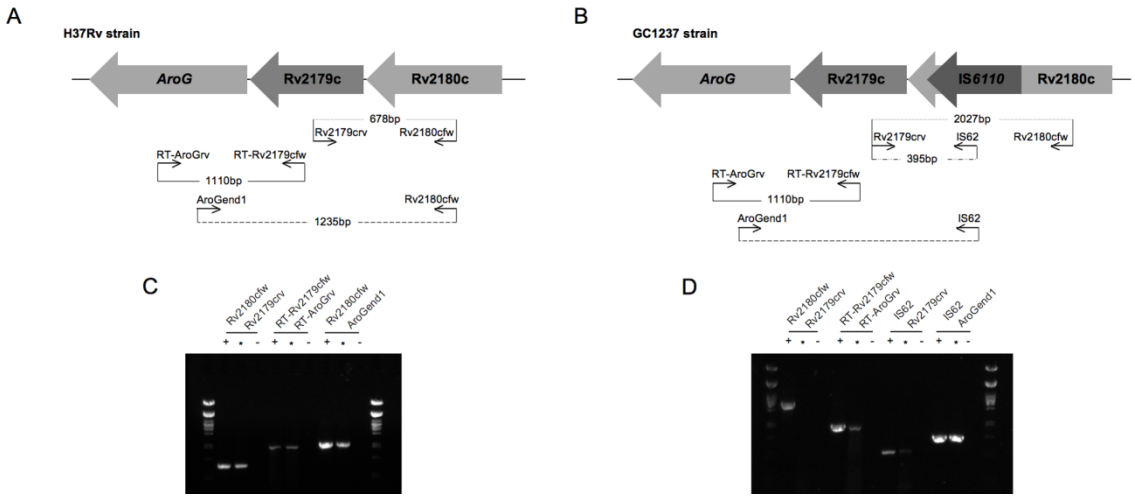


Figure 25: RT-PCR analysis of Rv2179c region from *M. tuberculosis* H37Rv and GC1237 strains. **(A).** Schematic diagram of Rv2179c region in H37Rv strain. **(B)** Schematic diagram of Rv2179c region in GC1237 strain. In both diagrams, the primers used for RT-PCR and the sizes of the fragments obtained with each pair of primers are indicated. The direction of transcription for Rv2179c is indicated by arrows. **(C)** RT-PCR analysis of this region in H37Rv. **(D)** RT-PCR analysis of this region in GC1237. In both RT-PCR the combination of primers is indicated above each set of reaction. Each set of three reaction consist of a positive control PCR assay with genomic DNA as the template (+), an RT-PCR (*), and a negative control assay without reverse transcriptase (-).

The *aroG* gene is an essential gene involved in the biosynthesis of chorismate precursor of the three aromatic amino acids in *M. tuberculosis* (14). There are no available data about the function of the protein encoded by Rv2179c gene. In fact, according to Tuberculist database, the product of this gene is a conserved essential hypothetical protein and it is conserved in different mycobacterial strains (11). From our results we can hypothesize that the overexpression of this gene would be

advantageous for GC1237 at least in certain environments such as infecting macrophages.

These results indicated that the three genes are cotranscribed in an operon, but these results do not exclude the possibility of independent promoters for each gene.

IS*6110* is acting as a promoter of Rv2179c gene both under extracellular and intracellular conditions

- ***Extracellular conditions***

To discern that Rv2179c upregulation observed in GC1237 strain was really due to the presence of IS*6110* and not to the different genetic background, the recombinant plasmids FPV79c and FPVIS79c containing the promoterless *gfp* gene preceded by the region upstream Rv2179c in H37Rv strain or in GC1237 strain, respectively were constructed and transformed into H37Rv obtaining HFPV79c and HFPVIS79c strains. HFPV27 was used as GFP-control strain.

GFP expression was measured both by fluorometry and flow cytometry. Results of both techniques showed that the fluorescence of HFPVIS79c strain is about 5-fold higher than the observed in HFPV79c strain (Figure 26A-B). Moreover, data obtained by fluorometry during the growth curve indicated that the effect of IS*6110* on the regulation of GFP in this case was independent of bacterial growth phase (Figure 26A). Therefore, this result suggests that in GC1237 strain, IS*6110* is acting as a promoter of Rv2179c gene.

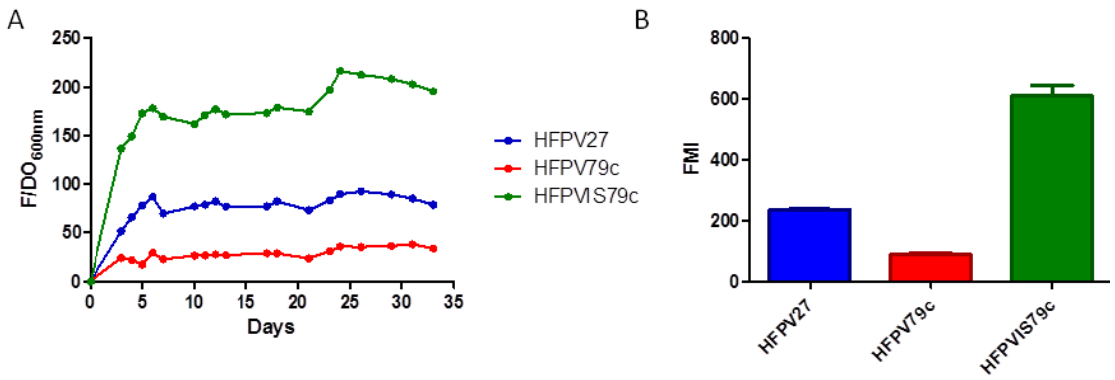


Figure 26: GFP expression in HFPV79c and HFPVIS79c strains. **(A)** Time course the evolution of the fluorescence of HFPV27, HFPV79c and HFPVIS79c strains during the bacterial growth normalized with OD_{600nm}. **(B)** Fluorescence median intensity (FMI) of HFPV27, HFPV79c and HFPVIS79c strains was measured by flow cytometry at logarithmic phase.

To confirm that this promoter activity was not exclusive inside H37Rv strain, these two plasmids were transformed in GC1237 and BCG strains and the GFP expression was studied during the growth curve both by fluorometry and flow cytometry. Figures 27A-B and 28A-B show that the GFP expression of the mutants with IS*6110* is about 2-fold higher than that in the mutants without IS*6110*. This result corroborates the no dependence on the genetic background of IS*6110*.

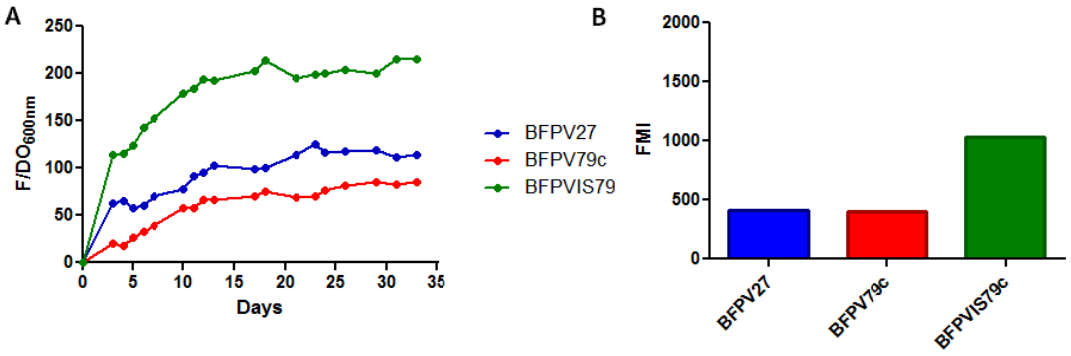


Figure 27: GFP expression in BFPV79c and BFPVIS79c strains. **(A)** GFP fluorescence of BFPV27, BFPV79c and BFPVIS79c strains during the bacterial growth normalized with OD_{600nm}. **(B)** FMI of the three strains measured by flow cytometry at logarithmic phase.

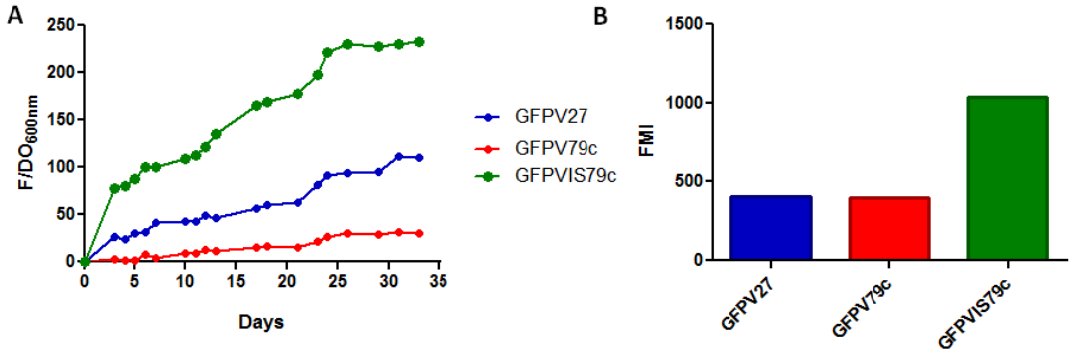


Figure 28: GFP expression in GFPV79c and GFPVIS79c strains. **(A)** Evolution of the fluorescence of GFPV27, BFPV79c and BFPVIS79c strains during the bacterial growth normalized with OD_{600nm}. **(B)** FMI of the three strains measured by flow cytometry at logarithmic phase.

- ***Intracellular conditions***

Murine immortalized alveolar macrophages (MH-S cell line) were infected with the different GFP expressing strains and infected cells FMI was measured by FACS. Thus, to exclude non-infected cells, it was also infected with non fluorescent H37Rv, considering their FMI as negative fluorescence level (R1 in Figure 29A). On the other hand, cells contained in the region with a higher fluorescence level were established as the infected ones (R2 in Figure 29A). Moreover, data were referred to annexinV-negative cells, in order to analyze only live cells. As it is observed in Figure 29B cells infected with HFPVIS79c strain showed a clear increase of their FMI values regarding the control at all times studied. These results are in agreement with the obtained data under extracellular conditions. Furthermore, GFP expression of HFPVIS79c-infected cells tended to increase during the experiment, being this rise more dramatic at 72h. In contrast, in the absence of IS*6110* this event failed to happen (Figure 29B).

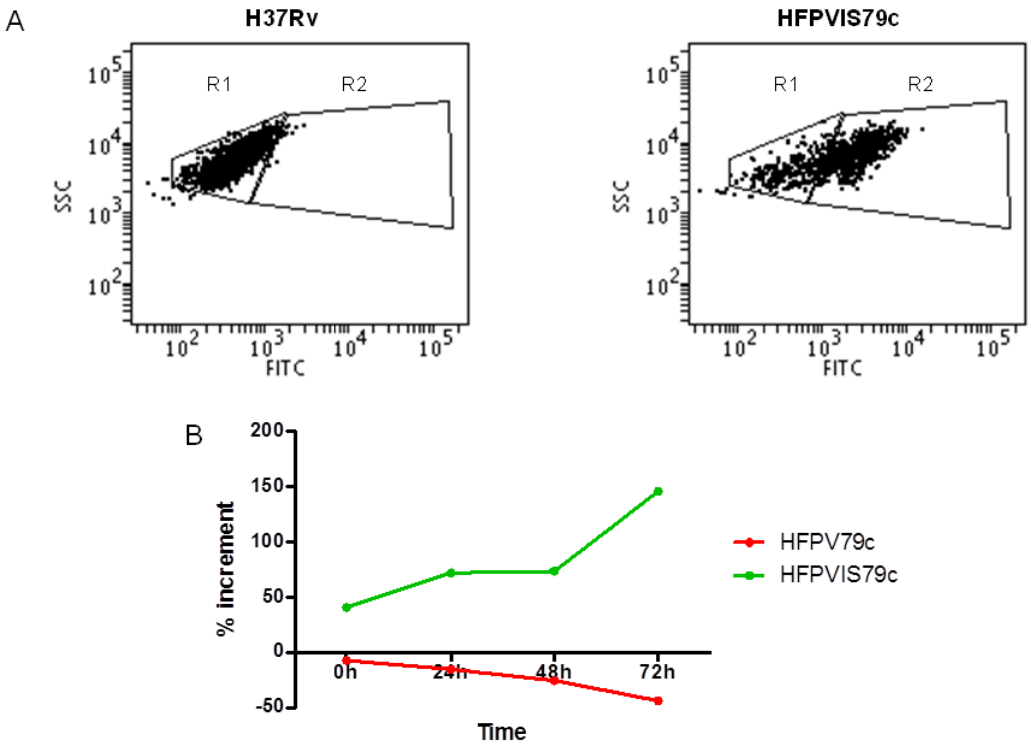


Figure 29: GFP expression in HFPV79c and HFPVIS79c strains in murine macrophages. (A) Histograms of H37Rv and HFPVIS79c infected cells. R1 indicates the non-fluorescent cells and R2 represents the fluorescent cells. **(B)** Time course of FMI of cells gated in R2, infected with HFPV79c and HFPVIS79c strains. The percentage of increment was calculated with respect to FMI values of HFPV27 strain. Figure shows a representative experiment of three performed.

A previous work showed a reinforcement of IS*6110* promoter activity under intracellular conditions (16). However, in that work authors studied the expression of different genes and the presence or not of IS*6110* upstream in two different strains. Nevertheless, authors did not discern the possibility of a regulation due to the different genetic background of the studied strains. Our results clearly show that the promoter activity of IS*6110* on Rv2179c gene is only related to disposition of the sequence with respect to this gene.

As part of the three genes forming an operon we analyzed the expression of *aroG* gene by qRT-PCR, but we could not observe significant increment of this gene in GC1237 strain compared with H37Rv strain. This result does not exclude the possibility of another regulation of *aroG* gene.

3. The effect of IS*6110* in *phoP* of the location of MBZ strain

Soto *et al*. demonstrated that the presence of IS*6110* insertion sequence 75 bp upstream the *phoP* gene in the MDR *M. bovis* B strain causes an increment in its transcription (18). This fact could have an important consequence due to this gene is an important transcriptional regulator.

To depth in the study of the promoter activity of IS*6110* in *phoP* of MBZ strain the plasmids pFPVSO5, pFPVSO7 containing the promoterless *gfp* gene preceded by the region upstream *phoP* in H37Rv strain or in MBZ strain, respectively were used. The plasmids were then transformed in H37Rv strain obtaining HFPVSO5 and HFPVSO7 strains and the RNA at logarithmic and stationary phase were obtained. The GFP expression of HFPVSO7 strain was measured by qRT-PCR and compared with that in HFPVSO5 strain. Figure 30 shows that the GFP expression of the mutant which contains IS*6110* is 3.5 (logarithmic phase) to 8 (stationary phase) times higher than that in HFPVSO5 strain.

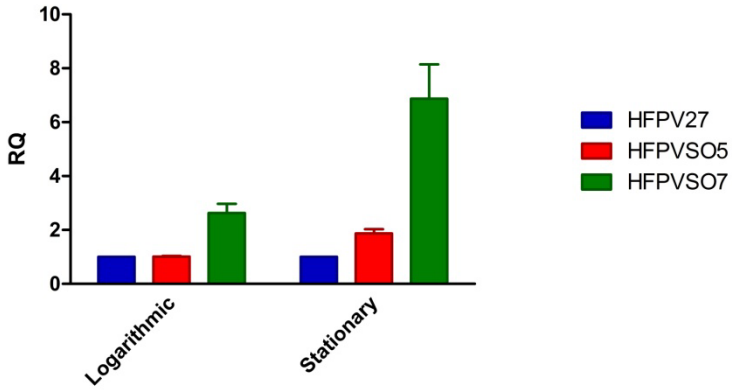


Figure 30: Relative quantification (RQ) of GFP in HFPVSO5 and HFPVSO7 strains. The GFP expression (normalized with *rrnaP1* levels) of both strains was measured by qRT-PCR at logarithmic and stationary phases. In both cases the expression was referred to GFP levels of HFPV27 strain. The results are a mean of three independent experiments.

Once the overexpression of *gfp* gene in HFPVSO7 strain was confirmed by qRT-PCR (Figure 30), this overexpression was analyzed in three different genetic backgrounds. For that issue, these two *gfp*-plasmids were transformed into two more strains: the clinical isolate GC1237 and *M. bovis* BCG and the GFP expression during the growth curve was studied in the three groups of strains both by fluorometry and flow cytometry. As controls, HFPV27, GFPV27 and BFPV27 strains were used. Similar high increment in the GFP expression was observed when IS*6110* was located upstream of

gfp gene. Results of both techniques showed that the fluorescence of the FPVSO7 strains is about 4 to 5 fold higher than the observed in FPVSO5 strains (Figure 31).

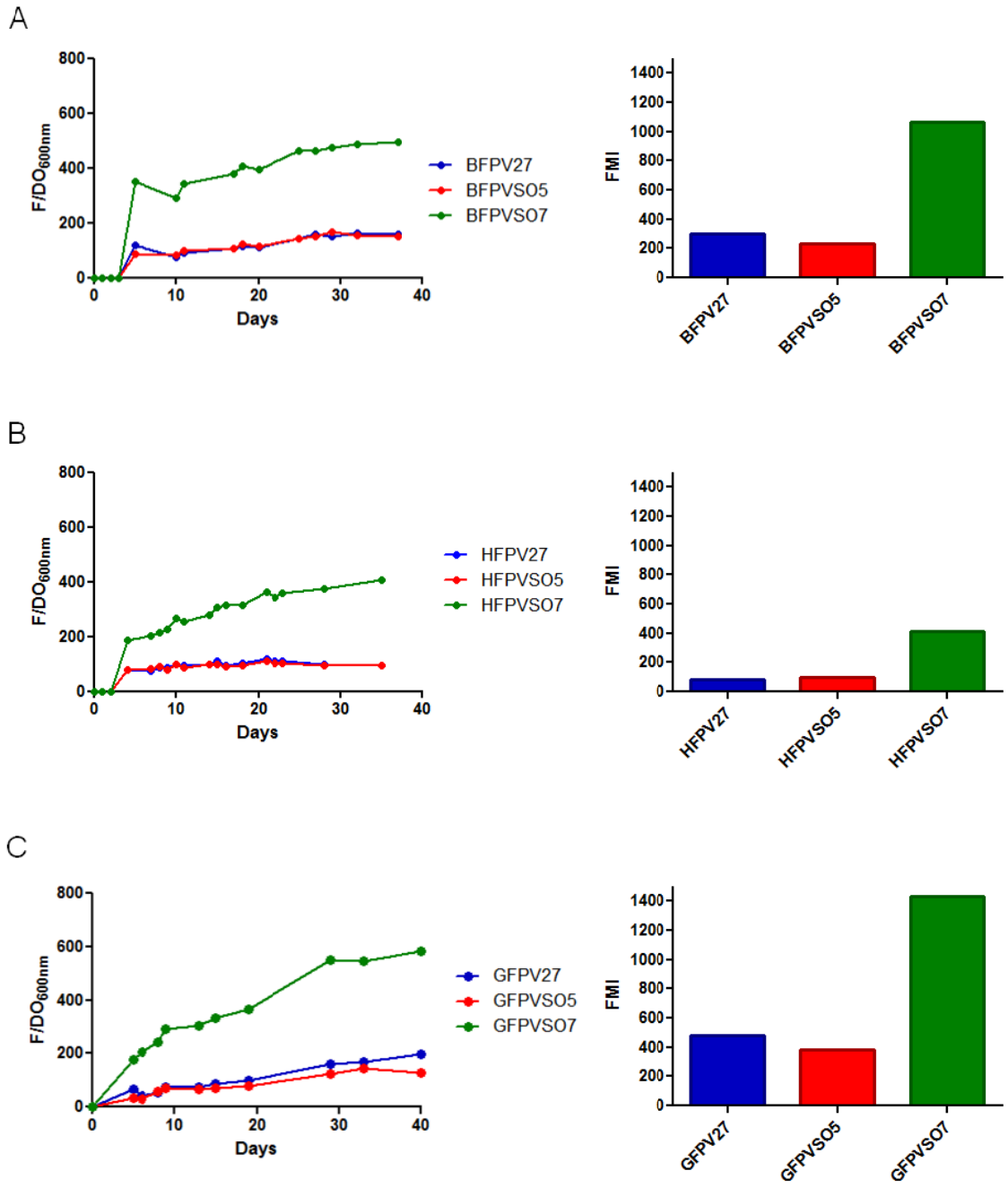


Figure 31: Comparison of GFP expression of the three groups of GFP-strains. Evolution of the GFP expression during the growth curve and FMI value of the three groups of strains measured by flow cytometry at logarithmic phase. (A) BFPVSO5, BFPVSO7 and BFPV27 (as a control) strains. **(B)** HFPVSO5, HFPVSO7 and HFPV27 (as a control) strains and **(C)** GFPVSO5, GFPVSO7 and GFPV27 (as a control) strains.

This result indicates that, independently of the genetic background, IS*6110* acts as a promoter in the studied region in different mycobacteria of *M. tuberculosis* complex causing a similar increment of expression of the upstream gene. Furthermore, this result corroborates the data obtained with the natural insertion of IS*6110* upstream the Rv2179c gene indicating that the increment of the expression caused for IS*6110* when it is acting as a promoter is regardless of the region.

The next step was to analyze whether the behavior of the promoter activity of this sequence in this location was also independent of the strain or genetic background inside the cell so, two different macrophages cell lines, MHS and Raw, were infected with the GFP-expressing H37Rv and GC1237 strains, and the fluorescence was analyzed at different times with two methods, by cytometry and by confocal microscope. In this case we do not infect with BCG-GFP strains due to the lack of replication inside the host.

MHS Infected cells FMI were measured by FACS. Thus, to exclude non-infected cells, we also infected with non-fluorescent H37Rv and GC1237, considering their FMI as negative fluorescence level. Data was referred to annexinV-negative cells, in order to analyze only live cells. Moreover, the FMI levels were normalized with respect to promoterless *gfp* strains. As it is observed in Figure 32A-B cells infected with HFPVSO7 and GFPVSO7 strains respectively, shows a clear increase of their FMI values regarding the control (HFPVSO5 and GFPVSO5 strains) at all times studied. These results are in agreement with the obtained data in extracellular conditions. The increment observed in HFPVSO7 strain was slightly higher than that in GFPVSO7 strain but the tendency of both mutants is exactly the same. This result confirms the promoter activity of IS*6110* in this location under intracellular conditions.

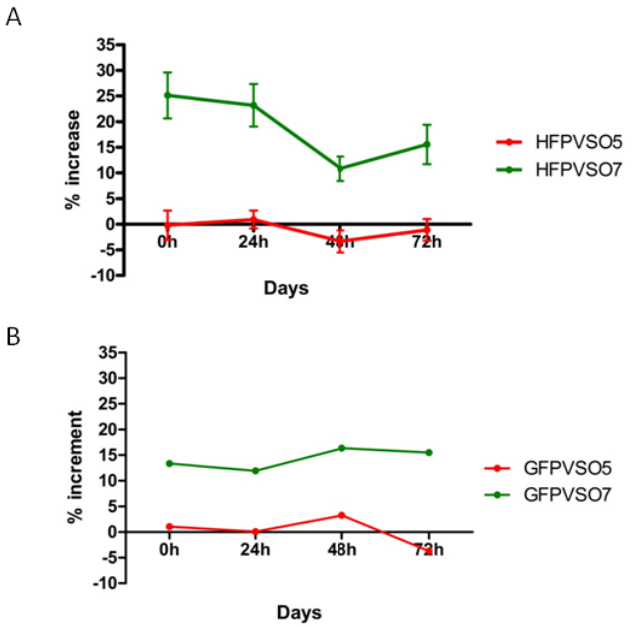


Figure 32: GFP expression in HFPVSO5/HFPVSO7 and GFPVSO5/GFPVSO7 strains inside murine macrophages (MHS). (A) Time course of FMI of cells, infected with HFPVSO5 and HFPVSO7 strains. The percentage of increment was calculated with respect to FMI values of HFPV27 strain. The result is a mean of three independent experiments. **(B)** Time course of FMI of cells, infected with GFPVSO5 and GFPVSO7 strains. The percentage of increment was calculated with respect to FMI values of GFPV27 strain. The graphic shows a representative experiment of three performed.

The promoter activity of IS*6110* in these GFP-strains was also studied by a second method. In this case, the intensity of infection of GFP strains was measured by a confocal microscope. For that issue, RAW cells were infected and after staining and fixation, images were collected with the confocal microscope Opera (Evotec) at the studied times (Figures 33A and 34A). After that, image analysis was done in IM 3.0 software. First, the nuclear stain DAPI was used to define macrophages number and spatial position in non infected wells. Because this stain did not allow us to define the contour of the cell, a macrophage was considered infected if labeled bacterium objects were distant of less than 5 pixels from the cell nucleus. Around 90% of the cells were found to be infected after 2h (Figures 33B and 34B). The intensity of infection means the GFP levels inside the macrophages (Figures 33C and 34C) and that was normalized with respect to the levels of promoterless *gfp* strains (HFPV27 and GFPV27).

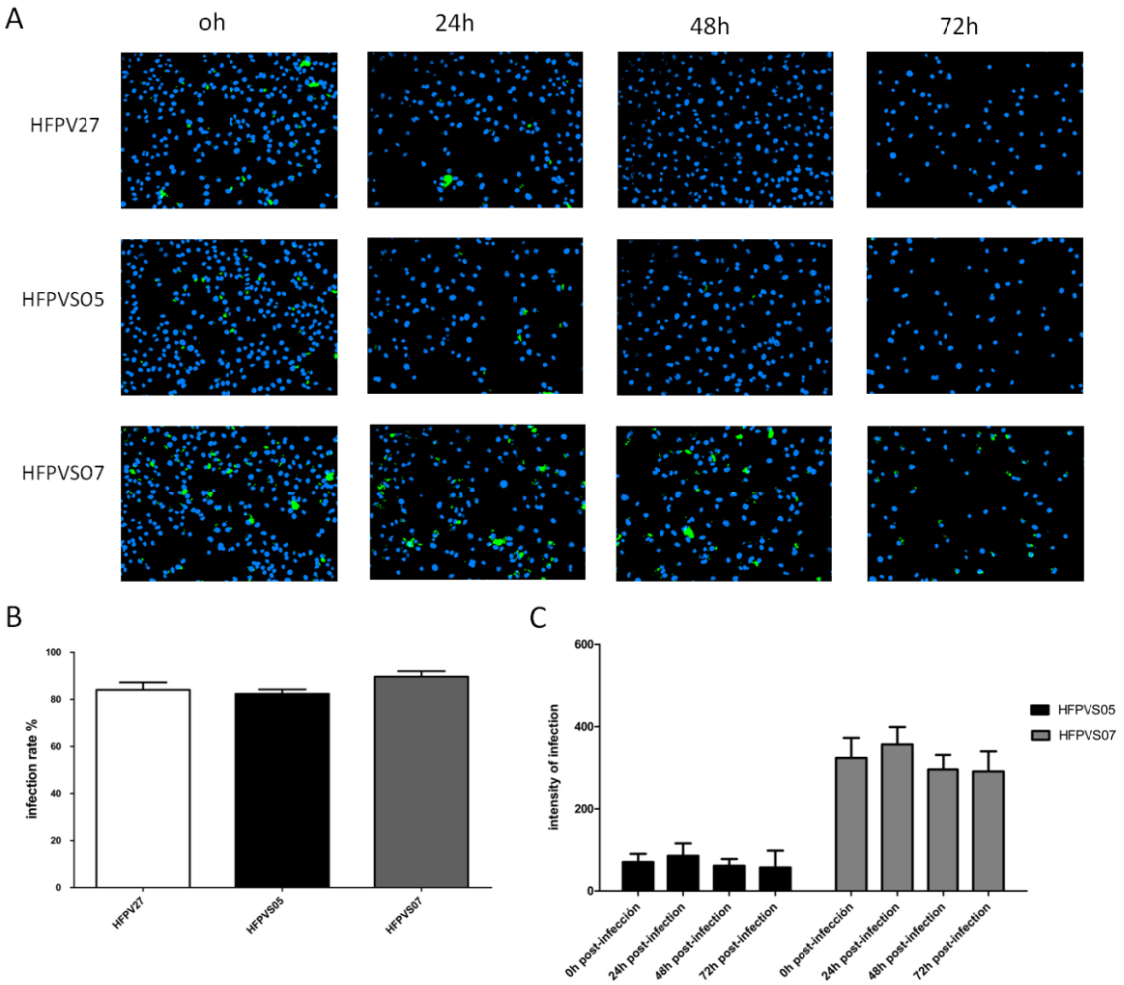


Figure 33: Intensity of infection of HFPV27, HFPVSO5 and HFPVSO7 strains in Raw cells: **(A)** Representative pictures of RAW cells infected with HFPV27, HFPVSO5 and HFPVSO7 strains and analyzed at: 0, 24, 48 and 72h post-infection. **(B)** Infection rate at 0h post-infection. **(C)** Intensity of infection of HFPVSO5 and HFPVSO7 strains normalized with intensity of infection of HFPV27 strain. Results are the means of four independent experiments; error bars indicate the standard deviations of the means.

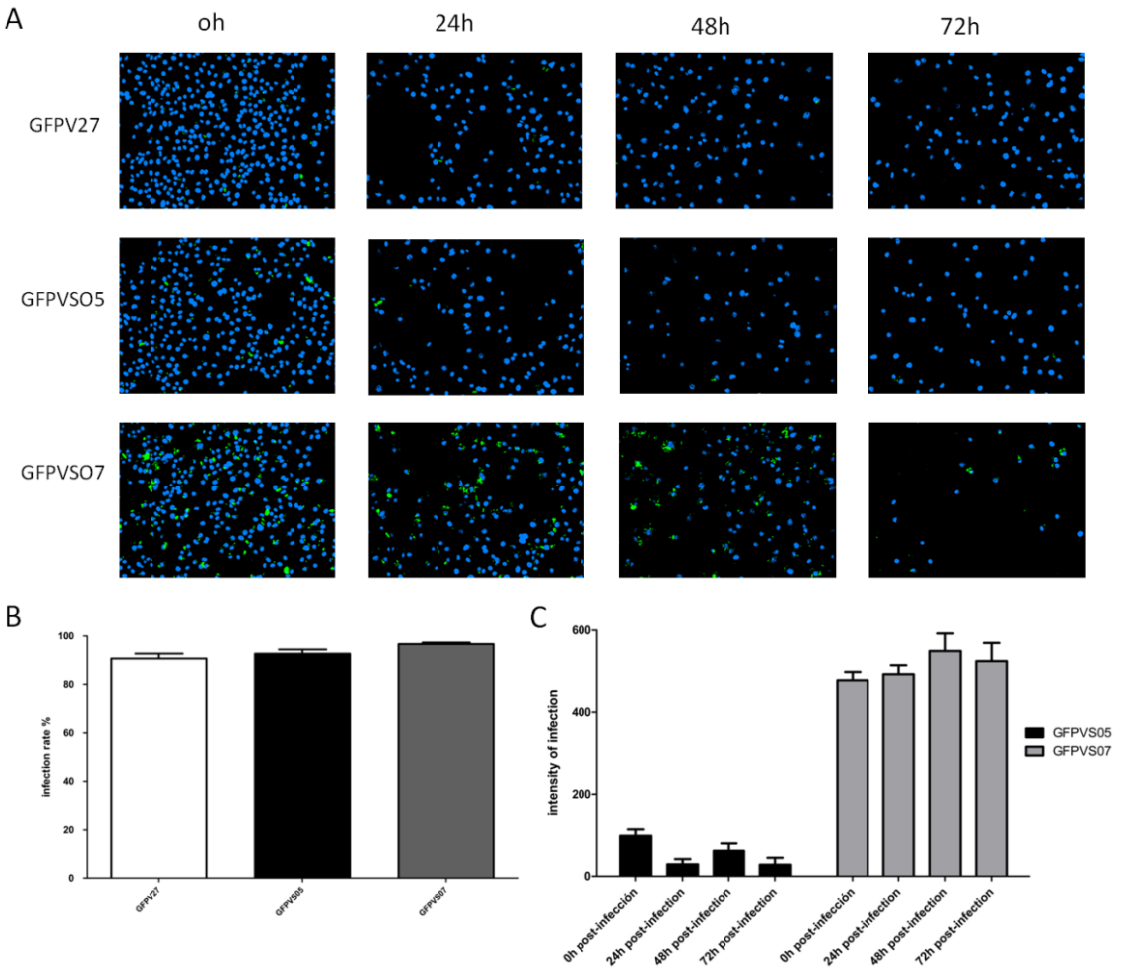


Figure 34: Intensity of infection of GFPV27, GFPVSO5 and GFPVSO7 strains in Raw cells: (A) Representative pictures of RAW cells infected with GFPV27, GFPVSO5 and GFPVSO7 strains and analyzed at: 0, 24, 48 and 72h post-infection. **(B)** Infection rate at 0h post-infection. **(C)** Intensity of infection of GFPVSO5 and GFPVSO7 strains normalized with intensity of infection of GFPV27 strain. Results are the means of four independent experiments; error bars indicate the standard deviations of the means.

The high increment of fluorescence from cells infected with HFPVSO7 or GFPVSO7 strains was quiet similar (Figures 33C and 34C). It is noteworthy that, data obtained by the two different methods as flow cytometry and confocal microscope indicated similar behavior across the experiment when we infected both cell lines with HFPVSO7 and GFPVSO7 strains, indicating the robustness of these results.

Approximation to the real transcription level of *phoP* in MBZ strain

As it has been described in the introduction of this chapter, a previous study analyzed the effect of IS*6110* on the transcription of *phoP* with the recombinant plasmids pSO5 and pSO7 in *M. smegmatis* mc^2155 as a surrogate host (because the manipulation of MBZ is difficult due to its resistance to most antituberculosis drugs) (18).

With the aim of coming close to the real effect of IS*6110* in *phoP* expression in MBZ strain; it was used the strain *M. tuberculosis* H37Rv ΔphoP which does not express *phoP* complemented with pSO5k (which contain a copy of *phoP* and its promoter region of H37Rv strain) and with pSO7k (which contain the *phoP* gene and the region upstream this gene in MBZ strain including the entire sequence of IS*6110*).

The first step was to analyze the RNA levels of *phoP* in these strains, and as Figure 35 shows, the transcriptional levels of this gene in H37Rv ΔphoP pSO5k and H37Rv ΔphoP pSO7k strains were extremely higher than that in H37Rv strain. However, the difference of *phoP* transcription levels was minimal between H37Rv ΔphoP pSO5k and H37Rv ΔphoP pSO7k strains.

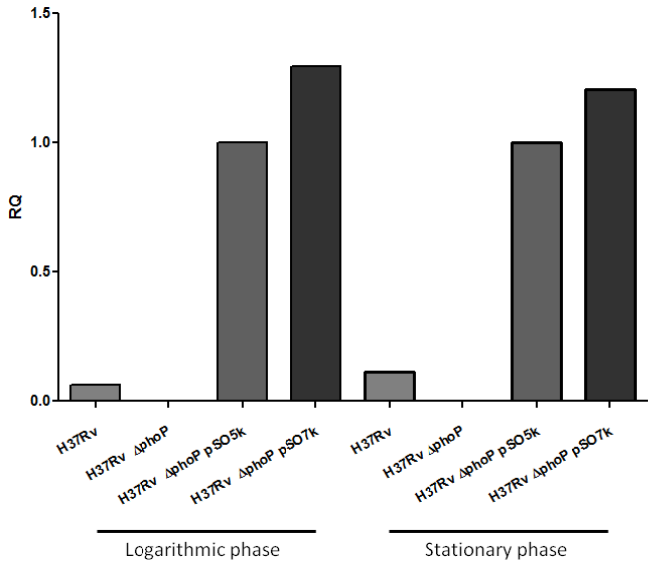


Figure 35: Transcriptional levels of *phoP* in the strains: H37Rv, H37Rv ΔphoP, H37Rv ΔphoP pSO5k and H37Rv ΔphoP pSO7k by qRT-PCR. RQ of *phoP* gene in H37Rv, H37Rv ΔphoP and H37Rv ΔphoP pSO7k strains with respect to H37Rv ΔphoP pSO5k strain at logarithmic and stationary phases. The RNA levels were normalized to the levels of *rrnAP1* mRNA. Results are the means of three independent experiments; error bars indicate the standard deviations of the means.

The high expression of *phoP* gene in H37Rv ΔphoP pSO5k strain could be due to pSO5k is a replicative plasmid and probably there is more than one copy of the plasmid inside H37Rv ΔphoP. The fact that no difference of transcriptional levels of *phoP* gen was observed between H37Rv ΔphoP pSO5k and H37Rv ΔphoP pSO7k strains could be due to the RNA generate from IS*6110* to *phoP* (from tsp1*, figure 17) is less stable than that in wile-type *phoP* (from tsp1).

Furthermore, it was studied the protein expression of phoP. Thus, western-blot analysis was carried out using H37Rv, H37Rv ΔphoP, H37Rv ΔphoP pSO5k and H37Rv ΔphoP pSO7k cell protein extracts and antibodies against PhoP and ESAT-6 proteins. The results clearly showed that PhoP is constitutively expressed in H37Rv strain while absent in H37Rv ΔphoP strain and the expression is recovered in the phoP-complemented strain H37Rv ΔphoP pSO5k and slightly more intense in H37Rv ΔphoP pSO7k strain (Figure 36A). This result is in agreement with the result obtained with RNA. However, the expected overexpression of *phoP* was not observed.

ESAT-6 is an immunodominant antigen whose gene (Rv3875) is absent in all *M. bovis* BCG strains, indicating a possible role in virulence. In previous studies, both PhoP and the ESAT-6 secretion system (ESX-1) have been identified independently as major virulence determinants of *M. tuberculosis*. The secretion of ESAT-6 was analyzed and it was observed that a strain with a no-functional PhoP did not secret ESAT-6 indicating that a direct link exists between the PhoP regulator and ESAT-6 secretion (6).

With the aim of checking whether an over-secretion of ESAT-6 is observed in H37Rv ΔphoP pSO7k strain, an extraction of culture filtrate proteins was carried out with H37Rv, H37Rv ΔphoP, H37Rv ΔphoP pSO5k and H37Rv ΔphoP pSO7k strains and a western-blot analysis was made comparing ESAT-6 in cell protein extracts (obtained previously) and filtrate protein extracts. As it is reflected in the Figure 36A-B, H37Rv ΔphoP pSO7k strain presents the least amount of ESAT-6 inside the cells and the highest expression in filtrate extracts. Due to the lack of a fully functional PhoP in H37Rv ΔphoP, a lack of ESAT-6 secretion is observed (Figure 36B) remaining this protein inside the cell (Figure 36A). The secretion of this protein gets back with the complementation of H37Rv ΔphoP with pSO5k and when the IS*6110* is acting as a promoter of *phoP* gene as in H37Rv ΔphoP pSO7k, an over-secretion of ESAT-6 is observed.

This result indicated that the presence of IS*6110* upstream *phoP* gene, in the particularly point of insertion of B strain, is changing the behavior of the strain, so it could be possible that one of the factors of the successful virulence of MBZ strain could be due to the over-expression and secretion of ESAT-6.

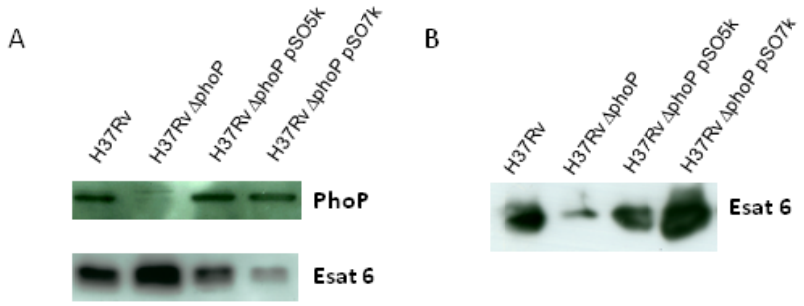


Figure 36: Western-blot analysis. (A) Western-blot image of H37Rv, H37Rv ΔphoP, H37Rv ΔphoP pSO5k and H37Rv ΔphoP pSO7k cell-protein extracts using polyclonal antibodies against phoP and monoclonal antibodies against ESAT-6 proteins. **(B)** Western-blot image of H37Rv, H37Rv ΔphoP, H37Rv ΔphoP pSO5k and H37Rv ΔphoP pSO7k filtrate proteins using monoclonal antibodies against ESAT-6 protein. 6μg of total protein were added in all cases.

What happen with H37Rv ΔphoP pSO7k inside the host?

To determine the contribution of IS*6110* in the phenotype of H37Rv ΔphoP pSO7k strain inside the host (this is an approximation to the behavior of B strain), MH-S cells were infected with: H37Rv, H37Rv ΔphoP, H37Rv ΔphoP pSO5k and H37Rv ΔphoP pSO7k strains and the mycobacterial replication inside the cells was analyzed at 0, 4 and 7 days post-infection (Figure 37). H37Rv presented a large increase in the number of CFUs (more than 2.5 logarithms of replication) within MH-S cells when is compared day 7 with day 0 post-infection. However, H37Rv ΔphoP showed a minimal replication (1 logarithm in total) within the cells. It was also observed that the replication is restored in the complemented strain, H37Rv ΔphoP pSO5k, indicating that *phoP gene* is essential for replication but not for persistence within this cell model. Although an increment of CFUs was observed in H37Rv ΔphoP pSO7k along the time of infection, this was minor than the observed in the wild-type strain H37Rv or in the *phoP*-complemented strain.

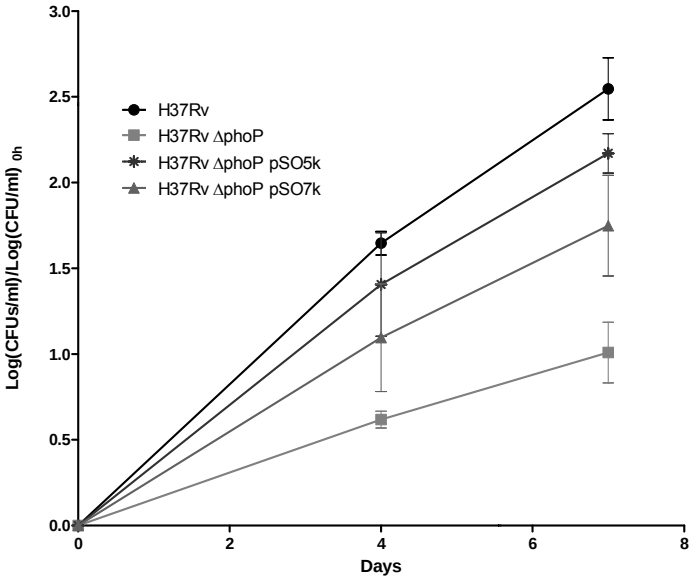


Figure 37: Replication of *M. tuberculosis* H37Rv, H37Rv ΔphoP, H37Rv ΔphoP pSO5k and H37Rv ΔphoP pSO7k strains in MHS-infected cells. Data are expressed as Log_{10} ± s.d. of CFU counts obtained and normalized with the CFUs values at 0h post-infection from two independent wells in one typical experiment.

As it has been explained before in this chapter, it was suggested that the insertion of IS*6110* at this particular locus could be responsible for the slowly grow of MBZ strain and for its peculiar colony morphology (smaller colonies than other *M. bovis* strains) (18). In order to check whether there were any differences in colony morphology of H37Rv ΔphoP pSO7k and H37Rv ΔphoP pSO5k, serial dilutions of both strains were plated in 7H10-ADC and incubate at 37ºC during 3 weeks. As figure 38 shows, H37Rv ΔphoP pSO7k has colony morphology smaller than H37Rv ΔphoP pSO5k; therefore, a similar behavior observed in MBZ strain is also observed in this strain. However, the bacterial growth in broth of H37Rv ΔphoP pSO7k was similar to that in H37Rv ΔphoP pSO5k.

The slow but constant replication of H37Rv ΔphoP pSO7k inside the cells could be due to a slower adaptation of this strain inside the host or a possible tendency of this strain to dormancy.

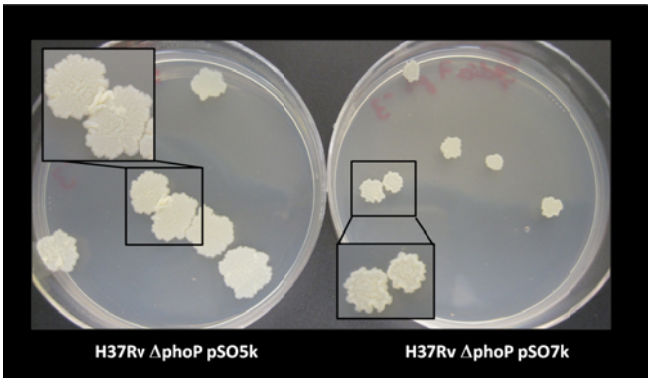


Figure 38: Colony morphology of H37Rv ΔphoP pSO5k and H37Rv ΔphoP pSO7k. Serial dilutions of both strains were plated in 7H10-ADC and incubated at 37ºC. After three week representative picture of the colony morphology was taken of the 10^{-3} dilution.

Finally, the phenotype of H37Rv ΔphoP pSO7k strain was analyzed after mouse infection and compared with those in H37Rv, H37Rv ΔphoP, H37Rv ΔphoP pSO5k strains. Thus, 5 mice were infected with each strain and after 4 week they were sacrificed and viable bacteria were counted in lungs. As figure 39 shows, for the four strains there was a large increment in CFU comparing inoculated-CFU with CFU after 4 week, however, no significant differences in these last CFU were observed among the studied strains.

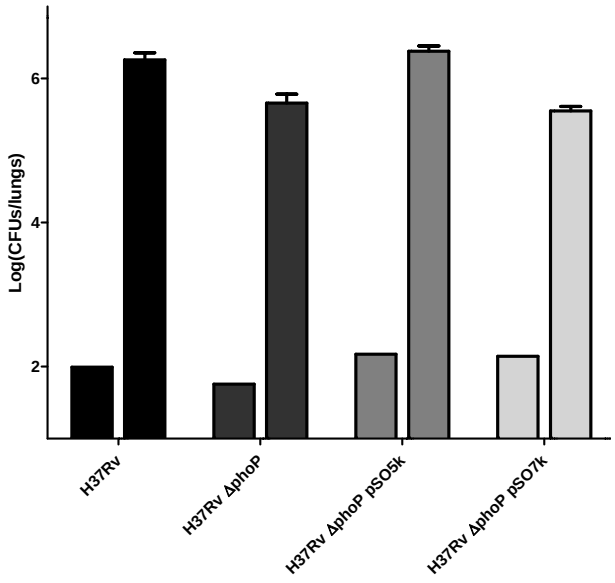


Figure 39: Replication of H37Rv, H37Rv ΔphoP, H37Rv ΔphoP pSO5k and H37Rv ΔphoP pSO7k strains in mice. Each group of two bars corresponds to the CFU of mycobacteria inoculated to the mouse and the CFU collected four week after inoculation in lungs.

Conclusions

- The insertion sequence IS*6110* located upstream *dnaN* gene in *M. tuberculosis* Beijing strain GC1237, has minimal effect on its gene expression both, in broth and in intracellular conditions.

- The copy of IS*6110* located upstream Rv2179c gene in *M. tuberculosis* GC1237 is acting as a mobile promoter both in broth and inside macrophages. Furthermore, this IS-promoter activity was checked in three different mycobacteria and the results indicated that this activity is independent of the genetic background.

- Rv2180c, Rv2179c and *aroG* genes are cotranscribed in H37Rv strain and the 3' end of IS*6110* (OP6110), Rv2179c and *aroG* genes are cotranscribed in GC1237 strain indicating the existence of an operon.

- The overexpression of *gfp* gene in HFPVSO7, which contain the region upstream phoP gene of MBZ (including the OP6110), was confirmed by qRT-PCR and the high GFP expression of this strain was confirmed by cytometry and fluorometry in broth. These results demonstrated the promoter activity of IS*6110* in this location.

- A high expression of GFP in HFPVSO7 and with similar profile was observed inside two different cell lines and by two different methods as flow cytometry and confocal microscope indicating the robustness of these results.

- H37Rv ΔphoP pSO5k and H37Rv ΔphoP pSO7k strains presented similar transcriptional levels of *phoP* gene and similar PhoP protein expression. Furthermore, both strains presented similar replication inside the cell.

- The colony morphology of H37Rv ΔphoP pSO7k strain is smaller than that in H37Rv ΔphoP pSO5k strain however; the bacterial growth in broth of H37Rv ΔphoP pSO7k strain was similar to that in H37Rv ΔphoP pSO5k strain.

- No difference on replication in mice was observed between H37Rv ΔphoP pSO5k and H37Rv ΔphoP pSO7k strain.

References

1. **Barker, L. P., D. M. Brooks, and P. L. Small.** 1998. The identification of Mycobacterium marinum genes differentially expressed in macrophage phagosomes using promoter fusions to green fluorescent protein. Mol Microbiol **29:**1167-77.
2. **Beggs, M. L., K. D. Eisenach, and M. D. Cave.** 2000. Mapping of IS6110 insertion sites in two epidemic strains of Mycobacterium tuberculosis. J Clin Microbiol **38:**2923-8.
3. **Brodin, P., Y. Poquet, F. Levillain, I. Peguillet, G. Larrouy-Maumus, M. Gilleron, F. Ewann, T. Christophe, D. Fenistein, J. Jang, M. S. Jang, S. J. Park, J. Rauzier, J. P. Carralot, R. Shrimpton, A. Genovesio, J. A. Gonzalo-Asensio, G. Puzo, C. Martin, R. Brosch, G. R. Stewart, B. Gicquel, and O. Neyrolles.** High content phenotypic cell-based visual screen identifies Mycobacterium tuberculosis acyltrehalose-containing glycolipids involved in phagosome remodeling. PLoS Pathog **6**.
4. **Casart, Y., L. Turcios, I. Florez, R. Jaspe, E. Guerrero, J. de Waard, D. Aguilar, R. Hernandez-Pando, and L. Salazar.** 2008. IS6110 in oriC affects the morphology and growth of Mycobacterium tuberculosis and attenuates virulence in mice. Tuberculosis (Edinb) **88:**545-52.
5. **Fontan, P., V. Aris, S. Ghanny, P. Soteropoulos, and I. Smith.** 2008. Global transcriptional profile of Mycobacterium tuberculosis during THP-1 human macrophage infection. Infect Immun **76:**717-25.
6. **Frigui, W., D. Bottai, L. Majlessi, M. Monot, E. Josselin, P. Brodin, T. Garnier, B. Gicquel, C. Martin, C. Leclerc, S. T. Cole, and R. Brosch.** 2008. Control of M. tuberculosis ESAT-6 secretion and specific T cell recognition by PhoP. PLoS Pathog **4:**e33.
7. **Ghanekar, K., A. McBride, O. Dellagostin, S. Thorne, R. Mooney, and J. McFadden.** 1999. Stimulation of transposition of the Mycobacterium tuberculosis insertion sequence IS6110 by exposure to a microaerobic environment. Mol Microbiol **33:**982-93.
8. **Gonzalo Asensio, J., C. Maia, N. L. Ferrer, N. Barilone, F. Laval, C. Y. Soto, N. Winter, M. Daffe, B. Gicquel, C. Martin, and M. Jackson.** 2006. The virulence-associated two-component PhoP-PhoR system controls the biosynthesis of polyketide-derived lipids in Mycobacterium tuberculosis. J Biol Chem **281:**1313-6.
9. **Jacobs, W. R., Jr., G. V. Kalpana, J. D. Cirillo, L. Pascopella, S. B. Snapper, R. A. Udani, W. Jones, R. G. Barletta, and B. R. Bloom.** 1991. Genetic systems for mycobacteria. Methods Enzymol **204:**537-55.
10. **Mahillon, J., and M. Chandler.** 1998. Insertion sequences. Microbiol Mol Biol Rev **62:**725-74.
11. **Marmiesse, M., P. Brodin, C. Buchrieser, C. Gutierrez, N. Simoes, V. Vincent, P. Glaser, S. T. Cole, and R. Brosch.** 2004. Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the Mycobacterium tuberculosis complex. Microbiology **150:**483-96.

12. **Martin, C., A. Williams, R. Hernandez-Pando, P. J. Cardona, E. Gormley, Y. Bordat, C. Y. Soto, S. O. Clark, G. J. Hatch, D. Aguilar, V. Ausina, and B. Gicquel.** 2006. The live Mycobacterium tuberculosis phoP mutant strain is more attenuated than BCG and confers protective immunity against tuberculosis in mice and guinea pigs. Vaccine **24:**3408-19.
13. **Otal, I., A. B. Gomez, K. Kremer, P. de Haas, M. J. Garcia, C. Martin, and D. van Soolingen.** 2008. Mapping of IS6110 insertion sites in Mycobacterium bovis isolates in relation to adaptation from the animal to human host. Vet Microbiol **129:**333-41.
14. **Parish, T., and N. G. Stoker.** 2002. The common aromatic amino acid biosynthesis pathway is essential in Mycobacterium tuberculosis. Microbiology **148:**3069-77.
15. **Perez, E., S. Samper, Y. Bordas, C. Guilhot, B. Gicquel, and C. Martin.** 2001. An essential role for phoP in Mycobacterium tuberculosis virulence. Mol Microbiol **41:**179-87.
16. **Safi, H., P. F. Barnes, D. L. Lakey, H. Shams, B. Samten, R. Vankayalapati, and S. T. Howard.** 2004. IS6110 functions as a mobile, monocyte-activated promoter in Mycobacterium tuberculosis. Mol Microbiol **52:**999-1012.
17. **Schurch, A. C., K. Kremer, A. Kiers, O. Daviena, M. J. Boeree, R. J. Siezen, N. H. Smith, and D. van Soolingen.** The tempo and mode of molecular evolution of Mycobacterium tuberculosis at patient-to-patient scale. Infect Genet Evol **10:**108-14.
18. **Soto, C. Y., M. C. Menendez, E. Perez, S. Samper, A. B. Gomez, M. J. Garcia, and C. Martin.** 2004. IS6110 mediates increased transcription of the phoP virulence gene in a multidrug-resistant clinical isolate responsible for tuberculosis outbreaks. J Clin Microbiol **42:**212-9.
19. **Turcios, L., Y. Casart, I. Florez, J. de Waard, and L. Salazar.** 2009. Characterization of IS6110 insertions in the dnaA-dnaN intergenic region of Mycobacterium tuberculosis clinical isolates. Clin Microbiol Infect **15:**200-3.
20. **Wards, B. J., and D. M. Collins.** 1996. Electroporation at elevated temperatures substantially improves transformation efficiency of slow-growing mycobacteria. FEMS Microbiol Lett **145:**101-5.

Chapter III

Introduction

The insertion sequence IS*6110* was isolated from a *Mycobacterium tuberculosis* (MTB) cosmid library as a repetitive sequence and was found to be specific to mycobacteria belonging to the MTBC (35). MTB strains typically contain multiple copies of this element (up to 25 copies per genome), although strains with only a single copy or no copies have been identified (14). In contrast, *M. bovis* usually have a low copy number such as *M. bovis* BCG which has either one or two copies (13). Although the genome of *M. tuberculosis* is highly homogeneous, strains are highly polymorphic with respect to copy number and location of IS*6110*. This sequence does not have a known target or consensus sequence; it has been found within ORFs and intergenic regions. However, *M. tuberculosis* and *M. bovis* strains with only one copy of this sequence show that it is inserted conservatively into one of array of 36-bp directly repeated sequences (16). Another study shows the possible conserved sites of insertion in strains of *M. tuberculosis* with less than seven copies (12) and also other hot-spots of IS*6110* have been shown such as the *ipl* loci (11), the phospholipase C regions (17, 39) and the *oriC* (20).

M. tuberculosis low copy number strains (LCS) and high copy number strains (HCS)

M. tuberculosis strains with less than six copies of IS*6110* are usually referred as **low copy number strains** (LCS) in the literature. Several studies reported the presence of LCS from regions as India, Vietnam or Tanzania (3, 27, 31). Furthermore, the 66% of the *M. tuberculosis* strains isolated in Tiruvallur, South India presented a single IS*6110* (32). **High copy number strains** (HCS), with six or more copies of IS*6110,* were reported by a greater number of papers. One study from Brazil, reported that 93.6% of *M. tuberculosis* strains had at least six copies ranging from 1 to 18 (34). In San Francisco, of 1,326 isolates investigated, 90% had six o more copies and only two isolates had no copies of IS*6110* (42). A majority (96.2%) of the 183 strains fingerprinted from Kampala were HCS, the number of IS*6110* copies ranged from 1 to 20 (2). Chauhan *et al.* analyzed 308 isolates of *M. tuberculosis*, from different parts of India and 56% of the isolates showed HCS of IS*6110* (9).

Both groups of strains, LCS and HCS, can originate outbreaks at similar proportion. The Beijing family is one of the lineages with the highest number of copies of IS*6110*. There are controversies among the factors that contribute to the success of this family. On the one hand, it is well known the wide distribution of Beijing strain 210 in the United States in 1995 (42) and its MDR clone, W, which caused disease in more than 350 patients in New York City (5, 6) and as we mention previously (Chapter 1), the

clinical isolate GC1237 has been responsible of epidemic outbreaks since it appeared in 1993 (7). However, one study conducted in Cape Town (South Africa) found no significant association between the *M. tuberculosis* genotype and transmissibility within the household (21). Besides, there are outbreaks reported caused by LCS, as was the extensive transmission of *M. tuberculosis* in a rural population with minimal risk factors for TB. This strain was designated as CDC1551 and the fingerprint showed 4 copies of IS*6110* (38).

LCS *versus* HCS and IS*6110* locations

Several reports have strongly suggested that the severity and clinical manifestations of tuberculosis depend on the immunogenicity and pathogenicity of the infecting *M. tuberculosis* strain. In this regard the IS*6110* sequence varies in number and position within the genome generating a high level of DNA polymorphism among strains. Interestingly, the location of IS*6110* in *M. bovis* isolates from endogenous reactivation cases from elderly people were studied in comparison to the bovine *M. bovis* strains, concluding that the presence of more copies in human strains could be related to the adaptation from the animal to the human host (25). On the other hand, a highly preferred site of insertion of IS*6110* was detected and designated as “DK1” (Rv0403c) in *M. tuberculosis* LCS but the prevalence of this site decreases sharply in HCS, suggesting a separate lineage for the HCS and the LCS (12). This contrasts with *M bovis* analysed without copy inserted at the same genomic position that *M tuberculosis* strains (25). This agreed with the idea that LCS of *M. tuberculosis* and *M. bovis* evolved separately after the progenitor acquired IS*6110* at the DR region. IS*6110* has been also checked as a tool to analyze the evolution of members of the MTBC. In such a case, transposition may have influence on the evolution of the strains, the parental strains should carry low copy number and the descendant, more evolved, would carry high copy number. One example that theoretically support that consideration are the Beijing lineage, members of this family are HCS of IS*6110* (between 15-25 copies per genome) and have shown high prevalence and high transmissibility (22). These characteristics could be seen as selective advantages of bacteria to its main purpose: infect humans (15).

Objectives

- To locate the common points of insertion of IS*6110* in Beijing strains.
- To carry out a scrutiny of points of insertion of IS*6110* in available literature and in the DNA sequence databases.
- To generate a representative primer collection of Beijing-IS*6110* points of insertion.
- To identify and analyze the common sites of IS*6110* in non-Beijing strains with respect to Beijing strains, being both groups HCS.

Material and methods

Bacterial selected strains, culture media, growth conditions and isolation of mycobacteria genomic DNA

Sixty-one *M. tuberculosis* clinical isolates were used in this work. The 61 isolates comprised 17 Beijing and 44 non-Beijing strains. Among the 17 Beijing strains 8 (NHN5, HM77, HM903, HM764, 990172, W4, N4 and CAM22) were previously selected as representative of this genotype using several typing methods from Europe. These 8 strains were selected as they share 80% or more of identity with Beijing strains from Shanghai area, China. The rest of the strains, each representing a different cluster, were selected for its high number copies of IS*6110* and were collected from Hospital Universitario Miguel Servet (HMS), Hospital Clínico Universitario Lozano-Blesa (HCU) from Zaragoza and Hospital General San Jorge from Huesca. Finally, *M. tuberculosis* H37Rv and *M. tuberculosis* GC1237 were used as control strains. An internal control, HMS 1301, was included in the study due to its RFLP is identical to GC1237 strain. Mycobacterial strains were grown at 37ºC in Middlebrook 7H9 broth supplemented with ADC and 0.05% Tween 80 and genomic DNA of all of them was isolated as it was described in Chapter 1. Table 4 summarizes the clinical isolates used in this work and figure 39 shows the RFLP, the spoligotyping and the family of all of them.

Location of the copies of IS*6110* insertion sequence in the eight Beijing strains selected as representative strains of this genotype

The study of the localization of the copies of IS*6110* in the eight representative Beijing strains (NHN5, HM77, HM903, HM764, 990172, W4, N4 and CAM22) was carried out by two methods, both based on PCR. A first research was conducted by LMPCR as previously described in Chapter 1. Briefly, LMPCR was performed with *Sal*I enzyme and the digestions were then subjected to PCR with the primers SalgD and ISA1 or ISA3 (Table 2 of Annex I). The amplified products were treated and sequenced and when a match was found, additional primers were designed and used with the eight strains to verify the point of insertion. These primers amplify the completed sequence of IS*6110* and approximately 300bp of both flanking sequences. Secondly, it was performed PCRs with specific primers, previously designed for amplifying the locations of IS*6110* in 210, W and GC1237 strains (Chapter 1). H37Rv and GC1237 were used as control strains in both cases. *All this primers are collected in Table 5 of Annex I.*

Table 4: Strains used in this study

Strain [a]	Family [b]	Year [c]	Strain [a]	Family [b]	Year [c]	Controls [d]
HMS 1356	Beijing	2003	**HMS 2584**	Beijing-like	2011	**H37Rv [e]**
HMS 1470	Beijing	2003	**HCU 2398**	Beijing	2002	**GC1237 [f]**
HMS 1480	Beijing	2004	**HCU 2613**	LAM 9	2004	**HMS 1301[g]**
HMS 1834	S	2005	**HCU 2617**	Beijing	2004	
HMS 2103	-	2007	**HCU 2633**	LAM 2	2004	
HMS 2193	LAM 9	2008	**HCU 2647**	LAM 5	2004	
HMS 2200	T1	2008	**HCU 2654**	LAM 3	2005	
HMS 2206	U	2008	**HCU 2663**	LAM 9	2005	
HMS 2208	LAM 3	2008	**HCU 2671**	T1	2005	
HMS 2224	H3	2008	**HCU 2703**	LAM12_MAD1	2005	
HMS 2227	-	2008	**HCU 2706**	-	2005	
HMS 2237	H3	2008	**HCU 2728**	T3	2005	
HMS 2241	LAM 10_CAM	2008	**HCU 2777**	T1	2006	
HMS 2258	T1	2008	**HCU 2821**	H3	2007	
HMS 2268	LAM 9	2008	**HCU 2864**	H1	2007	
HMS 2271	LAM 9-S	2008	**HCU 2886**	LAM 2	2007	
HMS 2280	LAM 9	2008	**HCU 2875**	T1	2007	
HMS 2281	LAM 9	2008	**HCU 2942**	S	2008	
HMS 2296	CAS	2008	**HCU 3151**	CAS	2010	
HMS 2310	T1	2008	**HSJ 025**	CAS	2008	
HMS 2355	-	2009	**HSJ 029**	U	2008	
HMS 2370	T1	2009	**HSJ 067**	Beijing	2010	
HMS 2391	U	2009	**NHN5**	Beijing	-	
HMS 2394	T2-T3	2009	**HM 77**	Beijing	-	
HMS 2400	Beijing	2009	**HM 764**	Beijing	-	
HMS 2405	T1	2009	**HM 903**	Beijing	-	
HMS 2414	H3	2009	**N4**	Beijing	-	
HMS 2513	T4_CEU1	2010	**W4**	Beijing	-	
HMS 2536	U	2010	**990172**	Beijing	-	
HMS 2580	Beijing	2011	**CAM 22**	Beijing	-	

[a] Names of the strains
[b] Family which belongs to by spoligotyping
[c] Year when they were isolated
[d] Control strains used in this study: [e and f] External controls of the study and [g] Internal control of the study

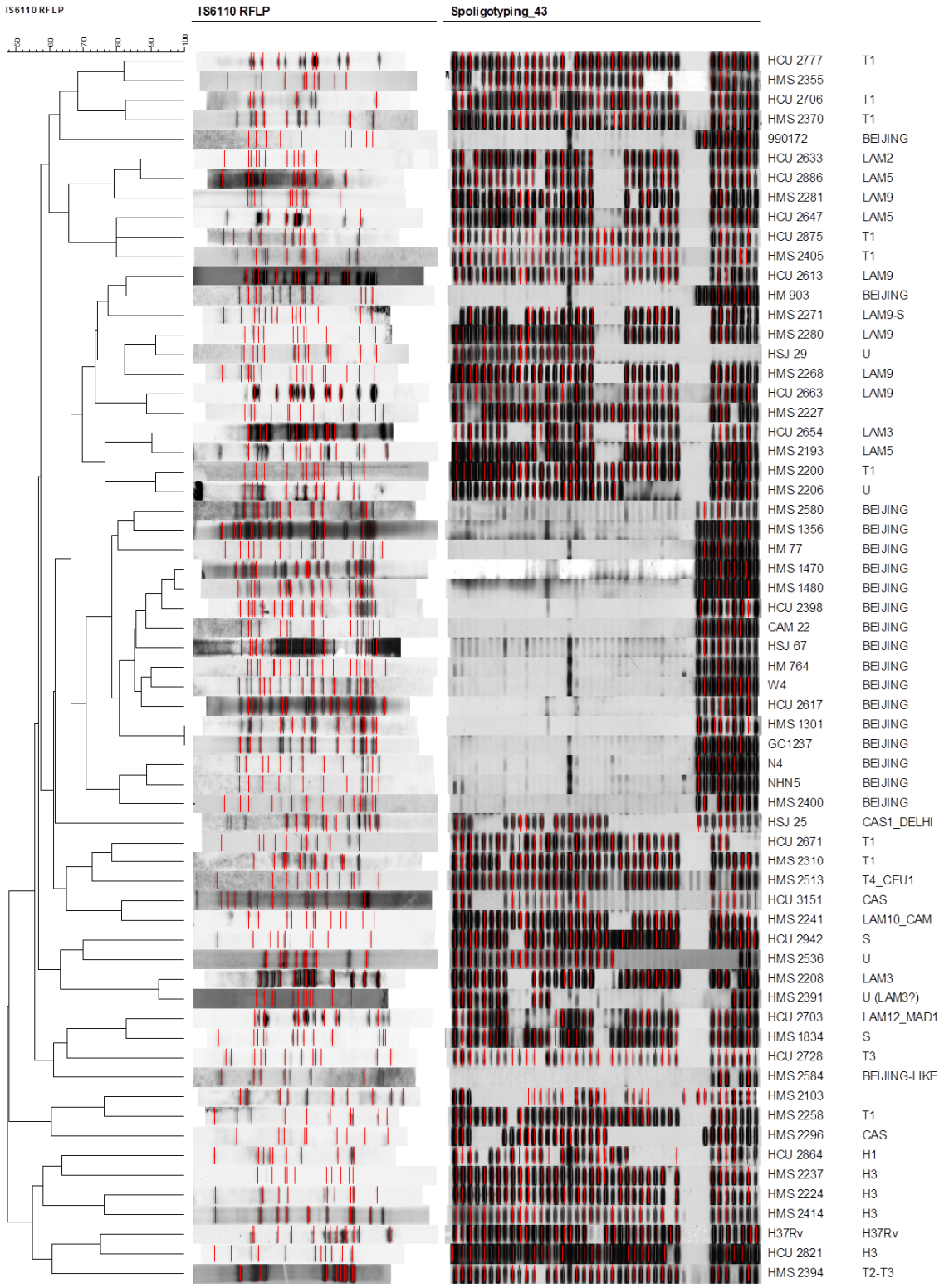


Figure 39: IS*6110*-RFLP and spoligotyping of the 61 strains used in this work. The RFLP and the spoligotyping of *M. tuberculosis* GC1237 and H37Rv were included.

Sequence analysis of points of insertion of IS*6110* in the DNA database and in the available literature

The points of insertion of IS*6110* of the reference sequenced strains: Beijing (210, 02_1987, 94_M4241A, HN878, R1207, T85, X-122, W and W-148) and non-Beijing (98-R604 INH-RIF-EM, BTB05-552, BTB05-559, C, CDC1551, CDC1551A, CPHL_A, EAS054, F11, GM 1503, K85, KZN 605, KZN R506, KZN V2475, KZN 1435, KZN 4207, NCGM 2209, str.Haarlem, S96-129, SUMu001, SUMu002, SUMu003, SUMu004, SUMu005, SUMu006, SUMu007, SUMu008, SUMu009, SUMu010, SUMu011, SUMu012, T17, T46 and T92), were obtained comparing the flanking regions of each IS*6110* in the genome sequences with the reference strain H37Rv using NCBI genetic sequence database (GeneBank) (http://www.ncbi.nlm.nih.gov/genome/166).

Moreover, available literature was analyzed to identify any further point of insertion of this sequence not described in DNA database (1, 4, 10-12, 18, 20, 23-25, 28-30, 33, 39-41).

After this sequence analysis, when a point of insertion of a Beijing strain (of GeneBank or available literature) was found outside of the amplified regions of our primer collection, additional primers were designed and included in the collection (Table 5 Annex I). In addition, flanking primers of DK regions (12) were designed to detect whether these frequent locations of IS*6110* in LCS were also present in the HCS selected in this study. Moreover, the primers BCG2A and BCG2B which amplify the promoter region of *phoP* (33) were included in the primer list to check whether another strain presented an IS*6110* in this region and finally, the primers designed for the study of locations of IS*6110* in clinical isolates of *M. bovis* with human host (25) were also included to study whether preferred locations of IS*6110* in these *M. bovis* strains were also present in the selected *M. tuberculosis* strains.

Location and analysis of copies of IS*6110* insertion sequence in the sixty-one selected clinical isolates

The localization of copies of IS*6110* in the 61 clinical isolates was carried out by PCR with all the pairs of oligonucleotides of the generated primer collection (Table 5 of Annex I). H37Rv and GC1237 were used as external controls and HMS1301 as internal control as this strain presents the same RFLP as GC1237. The 8 representative Beijing strains studied before by LMPCR and specific PCR were again included in this part of the study. The PCR products which might include an IS*6110* were sequenced with IS61 and IS62 primers (Table 5 of Annex I) and the flanking

regions were analyzed. The DRs were determined with the sequence analysis of the flanking regions of each copy of IS*6110*.

Results and Discussion

The eight Beijing strains selected as representatives of this genotype share IS*6110* points of insertion

A first research of points of insertion of IS*6110* in HCS was performed with the eight representative Beijing strains (NHN5, HM77, HM903, HM764, 990172, W4, N4 and CAM22) by LMPCR and PCR with the primers used previously in the study of IS-locations of GC1237 strain (Chapter 1). Seventeen new genomic points of insertion of this element were obtained by LMPCR. Additional primers were designed to amplify these new IS*6110* locations and used for checking whether they were also present in another strain of this group. As a result, a total of 45 different points of insertion of IS*6110* were obtained (Table 5).

With this first scrutiny, the IS*6110* common locations in these Beijing strains also shared with GC1237, 210 and W (Beijing control strains) were obtained (Table 5 and figure 40). All these strains presented the three characteristic locations of IS*6110* of Beijing strains: in the intergenic region of *dnaA:dnaN*, between Rv1754c-Rv1765c genes which correspond to the RD152 and in the Direct Repeat region. In the case of the strain CAM22, it was detected two copies of IS*6110* in this region, both in the same orientation.

Six more sites of insertion are stood out as IS*6110* was present with high frequency: Rv1371, *ctpD* (Rv1469c), Rv2016, *idsB* (Rv3383c), between the two IS*1532* and *esxR-esxS* region. Six of the eight strains presented a copy of IS*6110* in Rv2016 gene in the same point and orientation, whereas in the other two strains, IS*6110* was also present in this gene but in different points. *idsB* and Rv1371 genes were disrupted by this element in seven strains and a copy of IS*6110* was present in *esxR-esxS* region in six cases. Another highly preferred point of insertion was located in *ctpD* gene as five of the eight strains presented an IS*6110* disrupting this gene. Finally, between the two IS*1532*, a copy of IS*6110* was found in five strains (Table 5 and Figure 40). Due to the high frequency of all of these IS*6110* locations, these regions could be specific hot-spots for this element in Beijing strains (IS*6110* Beijing-hot-spots).

In some of these strains, a copy of IS*6110* was observed in the *ipl* loci (red arrow, Figure 40), however, its frequency was fewer than that in the other regions described above.

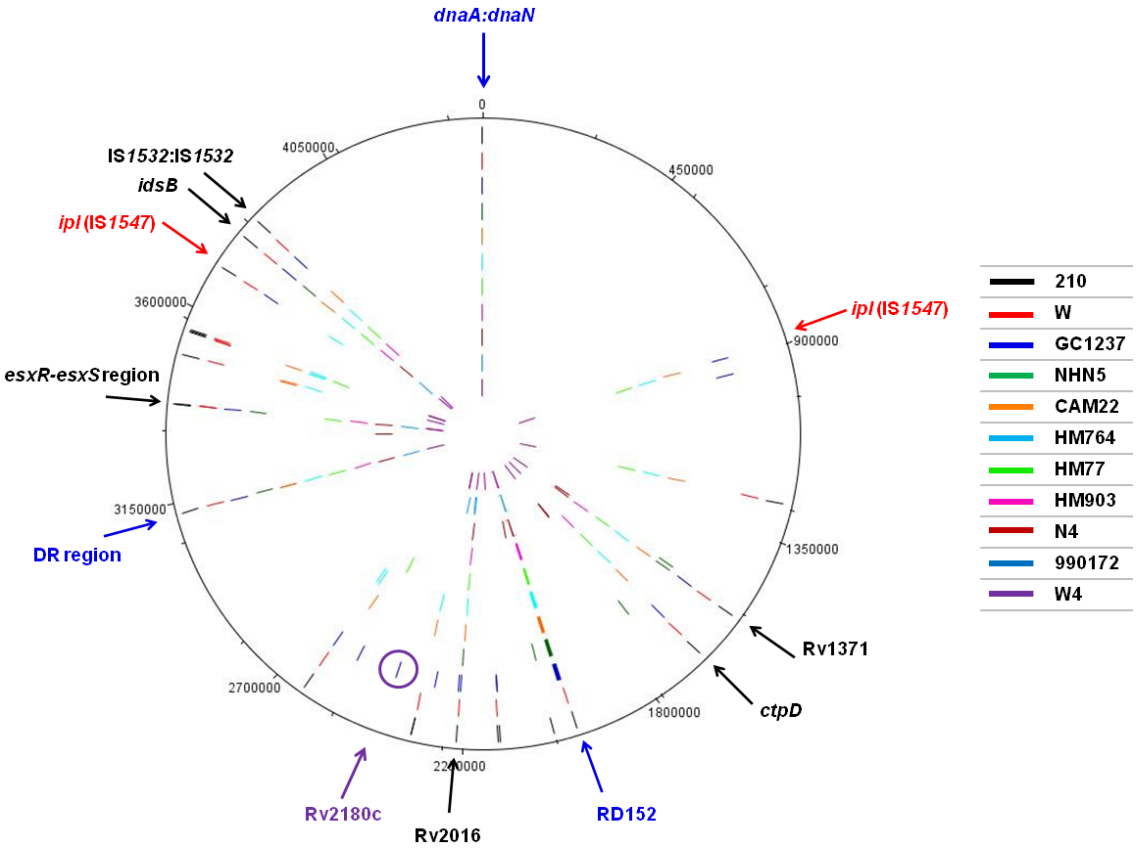


Figure 40: Distribution of IS*6110* insertion sequence throughout the genome of *M. tuberculosis* Beijing strains. The located copies of IS*6110* in NHN5, HM77, HM903, HM764, 990172, W4, N4 and CAM22 Beijing strains were plotted in H37Rv genome and represented in different concentric circles. IS*6110* locations present either in all strains (blue arrows) or with high frequency (black arrows) are indicated; *ipl* hot-spots are pointed with red arrows. In this DNA plotter, the IS*6110* locations of GC1237, W and 210 Beijing strains were also represented as IS*6110* reference points of insertion. The unique insertion of IS*6110* of GC1237 (within Rv2180c) is surrounded by a purple circle.

Comparing these strains, unique locations were also observed. Some examples are: within *papA*4 and Rv2957 genes (N4 strain), interrupting *mez* and *PPE49* genes (CAM22 strain) and in the intergenic region of Rv1542c:Rv1543 (W4 strain). It is noted that the location of IS*6110* within Rv2180c in GC1237 strain, described in Chapter 1 and analyzed in depth in Chapter 2 was unique (Figure 40 and Table 5). The high number of unique IS*6110* among these strains indicates that although IS*6110* has preferential genomic regions, its insertion is randomly and generates a high genomic plasticity among strains of the same family.

As it was explained in Chapter 1, it is frequent to find at least a copy of IS*6110* in the NTF region in Beijing family (26) and the absence of IS*6110* in this locus is

characteristic of ancestral sublineages within the Beijing genotype (19). Following this classification, these eight strains are ancestral as the absence of IS*6110* in this region.

The majority of the IS*6110* located copies were due to transposition mechanism

The flanking regions of each copy of IS*6110* of the eight Beijing strains were analyzed and with the exception of three locations of this element (indicated with minus in Table 5), the presence of direct repeats (DR) was detected (Table 5). The absence of DRs in two of the other three locations, RD152 and RD207, is probably a consequence of recombination between two adjacent copies of IS*6110* (36) as it was observed loss of these genomic regions. The last of the three IS*6110* elements without DR, which was localized between Rv0794c:Rv0797, was identical to GC1237 (explained in Chapter 1) and could be explained by reorganization of this region. This event could be characteristic of Beijing genotype.

As already mentioned in Chapter 1, due to the high number of copies of IS*6110* per strain, it was possible to observe copies flanked by DR and other copies without DR indicating that the probability of rearrangement process between copies rises when the number of those increases.

Table 5: Locations of IS*6110* of the eight representative Beijing strains, part 1

Genes	Points of insertion	Direct Repeats	210	W	GC1237
Rv0001(*dnaA*):Rv0002(*dnaN*)	1594	ATT	+	+	+
Rv0794*c*:Rv0797 (IS*1547*)	889072	-			+
Rv0840c (*pip*)	937116	ACG			+
Rv1135c (PPE16)	1262963	AGC	+	+	
Rv1358:Rv1359	1530161	GCG			
Rv1359	1530203	CTC			
Rv1371	1543972	GAGG	+	+	+
Rv1469 (*ctpD*)	1657016	CGT	+	+	+
Rv1527c (*pks5*)	1724406	CCGA			
Rv1528c (*papA4*)	1729409	CTT			
Rv1542c(*glbN*):Rv1543	1745059	ACC			
Rv1754c	1986638	/	+	+	
Rv1754c-Rv1765c (RD152)	1986638-1998625	-			+
Rv1798:Rv1799 (*lppT*)	2038898	/	+		
Rv1800 (PPE28):Rv1801 (PPE29)	2041741	GCG			
Rv1917c (PPE34)	2163392	TTA			++
	2163649	TCC	+		
	2167310	/	+	+	
Rv2016	2263619	TCC			
	2263627	AGG	+	+	+
	2263778	CCT			
Rv2019	2266167	GAC			
MT2080 (Mb2047c):MT2081 (Mb2048c)	2268726-2268693	GAA			+
Rv2077A:Rv2078	2334666	AGG			
	2334686	AGG			+
Rv2104c:Rv2107 (PE22)	2366897	CAC	+		
	-	/		+	
Rv2107 (PE22):Rv2108 (PPE36)	2367679	CGA	+		
Rv2180c	2442348	AGC			+
Rv2282c	2555718	AAT			
Rv2286c	2559506	ATC			+
Rv2332 (*mez*)	2605438	ATT			
Rv2349c (*plcC*)	2627829	CAG			
Rv2352c (PPE38)	2634022	/	+	+	
Rv2353c (PPE39):Rv2356c (PPE40)	2635592-2637688	CCG			+
Rv2356c (PPE40):Rv2357c(*glyS*)	2639612				
Rv2813-Rv2820c (RD207)	3127927	-	+	+	+
Rv2957	3310163	GTT			
Rv3018A (*PE27A*):Rv3019c (*esxR*)	3378553	GCC	+	+	+
Rv3019c (*esxR*): Rv3020c (*esxS*)	3379027	CTG	+	+	
Rv3125c (PPE49)	3491592	CTC			
Rv3128c	3493908	CGA	+	+	
Rv3128c:Rv3129	-	/		+	
Rv3178:Rv3179	3547342	ATG	+		
	3549199	CGG	+	+	
Rv3184:Rv3187	-	/	+	+	
Rv3190c	3555513	CGC			
Rv3324A:Rv3327 (IS*1547*)	3711737	GGC	+	+	+
Rv3382c (*lytB1*)	3796689	TGG			
Rv3383c (*idsB*)	3797823	ATC	+	+	+
Rv3427c (IS*1532*):Rv3428c (IS*1532*)	3844681	CGGG	+	+	+
	3844737	CCCG		+	

Table 5: Locations of IS*6110* of the eight representative Beijing strains, part 2

Strains							
NHN5	**HM764**	**990172**	**W4**	**N4**	**HM903**	**HM77**	**CAM22**
+	+	+	+	+	+	+	+
	+		+			+	+
	+		+			+	+
+							
				+			
+	+		+	+	+	+	+
	+		+		+	+	+
+				+			
				+			
			+				
+	+	+	+	+	+	+	+
+				+			
			+				
				+			
+	+		+		+	+	+
		+					
		+					
		+					
	+		+				+
			+				
						+	
							+
	+						
	+						
+	+	+	+	+	+	+	++
				+			
+		+	+	+	+	+	
			+				
							+
	+		+				+
	+		+				
	+		+			+	+
	+						
	+						
		+					
+	+		+	+	+	+	+
	+		+		+	+	+
	+						

Five of the eight Beijing strains have a possible IS-promoter

The orientation and distance of each copy of IS*6110* to the neighbouring genes was analyzed and in addition to the locations of IS*6110* described in Chapter 1 where could act as a promoter, 5 new copies were located with the same orientation and close enough to neighbouring genes to act as a mobile promoter (Table 6).

Table 6: New locations of IS*6110* of the eight representative Beijing strains where IS*6110* could act as a mobile promoter

Strain	Genes of IS*6110*	Distances to neighbouring genes
NHN5	Rv1358:Rv1359	12 bp upstream Rv1359
HM764	PPE40:*glyS*	77 bp upstream PPE40
990172	Rv2077A:Rv2078	92 bp upstream Rv2077A
W4	Rv1542c:Rv1543	5 bp upstream Rv1543
N4	Rv2016	370 bp upstream Rv2017

It is probably that the IS*6110* located upstream Rv1359 gene in NHN5 and the copy located upstream Rv1543 in W4 do not act as a mobile promoter due to the distance to these gene is too short however, to confirm this, the effect of this element in these genes should be studied. Finally, it is noteworthy that these IS*6110* were observed only in one strain and all of them were different from each indicating the genomic variability that IS*6110* generates among strains of a same genotype.

Randomly distribution of IS*6110* in reference sequenced strains of GeneBank

The analysis of the points of insertion of IS*6110* of 43 reference sequenced strains of GenBank was carried out comparing their sequenced fragments with H37Rv genome. All the genomic locations of IS*6110* of these strains were plotted on the H37Rv genome map (Figure 41). The IS-locations were grouped in two circles; the first one corresponds to Beijing strains and another one to non-Beijing strains. In this DNA plotter the locations of IS*6110* of GC1237 (red circle) and of H37Rv (green circle) were also included.

A randomly distribution of this element throughout the genome was observed among these strains (Figure 41). However, the presence of numerous preferential integration loci of IS*6110* were also detected (some examples are indicated with black arrows in Figure 41). Beijing and no-Beijing strains presented copies of IS*6110* in these hot-spots corroborating that preferential integration regions are common among *M. tuberculosis*

strains. In all the strains this element was found at an identical position within the DR region (with the exception of the Beijing strains which presented their characteristic deletion: RD207, so, characteristic point). Moreover, in all Beijing strains it was observed an identical copy of this element between *dnaA:dnaN* genes. In this region but in a different point, a copy of IS*6110* in a non-Beijing strain (SUMu003) was observed indicating that the presence of an IS*6110* in this intergenic region is characteristic of Beijing genotype but not exclusive of this family. This result is in agreement with studies which show different locations of IS*6110* between these two genes (8, 20, 37). The detection of Beijing strain amplifying this entire intergenic region is a useful tool but it is necessary another genomic feature of this genotype if not, the result could be a Beijing false positive.

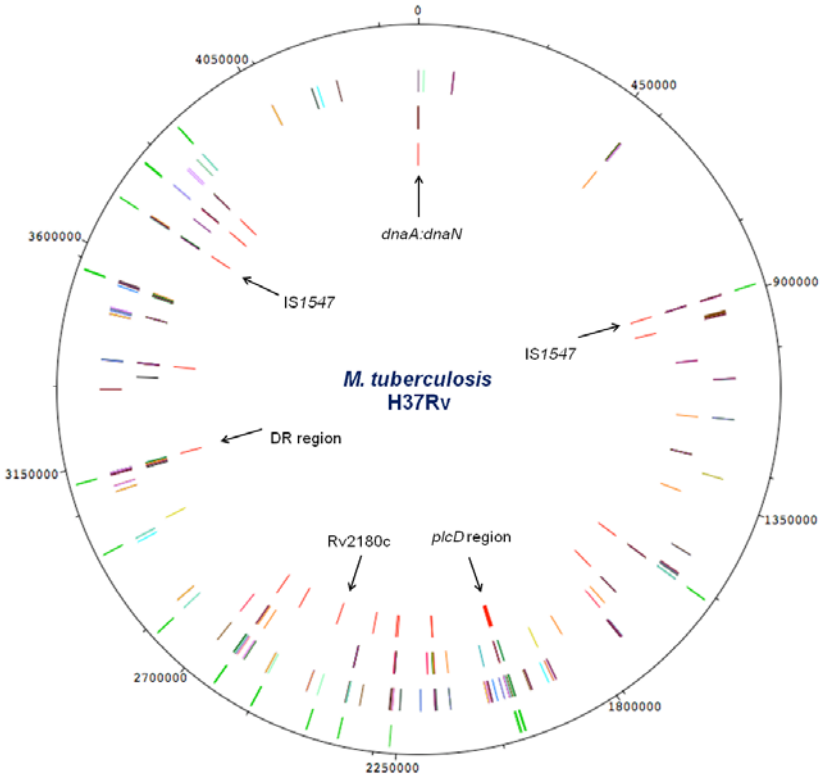


Figure 41: Graphic representation of point of insertion of IS*6110* in *M. tuberculosis* genome with Artemis DNA plotter. Distribution of insertion loci identified with the sequence analysis of genomic database in this work. The first circle (green) represents the genomic locations of IS*6110* in H37Rv. The second concentric circle represents the distribution of this sequence in non-Beijing strains. The third circle represents the locations of IS*6110* in Beijing strains and the fourth concentric circle (red) correspond to the genomic points of insertion of IS*6110* in GC1237 strain. All the IS-locations were plotted on H37Rv genome. The hot-spots: IS*1547*, *plcD* region, DR region and *dnaA:dnaN* region are indicated by arrows. In addition, the unique IS*6110* of GC1237 (upstream Rv2180c) is indicated.

The copy of IS*6110* located upstream Rv2180c gene in GC1237 was also indicated in the *M. tuberculosis* map with an arrow (Figure 41) and, as it is observed, there was no strain with this location confirming again that this location is unique of this strain.

It is noteworthy that this sequence was found inserted more often in some genomic regions, as the one from 1800000 to 2700000 (coordinates referred to H37Rv genomic map, Figure 41); whereas other regions lacked the presence of IS*6110* as happens in the region around the *OriC*. Probably, these areas could be more abundant in essential genes. In fact, although an IS*6110* transposes in these regions, if they are essential, the outcome of this event will not be observed. These findings are in agreement with previous studies of chromosomal distribution of IS*6110* (29, 30, 41).

It has been included a detailed table in Annex II with all the points of insertion of IS*6110* of the analyzed reference strains of GeneBank.

A primer collection which amplify locations of IS*6110* in MTBC strains was generated in this study

Available literature (1, 4, 10-12, 18, 20, 23-25, 28-30, 33, 39-41) was also reviewed and when a location of IS*6110* in a Beijing strain was not included in the amplified regions of the primer collection new pair of primers were designs to amplify the area. In addition, as one of the objectives was to detect whether the preferred site of insertion of IS*6110* in LCS were present or absent in the selected HCS, the DK regions (12) were analyzed and when necessary, primers were designed and included in the oligonucleotide collection. Moreover, the primers used to amplify preferred locations of IS*6110* in *M. bovis* with human host (25) were included to check whether these locations are also detected in the selected *M. tuberculosis* clinical isolates as the host in both cases is the same. Finally, as one of the interesting IS*6110* was the located upstream *phoP* gene in MBZ strain and it was quiet studied in Chapter 2, the primers BCG2A and BCG2B which amplify the promoter region of this gene (33) were included with the aim of checking the frequency with which IS*6110* is inserted in this region. Figure 42 shows the distribution of the amplified regions with this collection and a detailed table has been included in Annex I (Table 5 of this Annex).

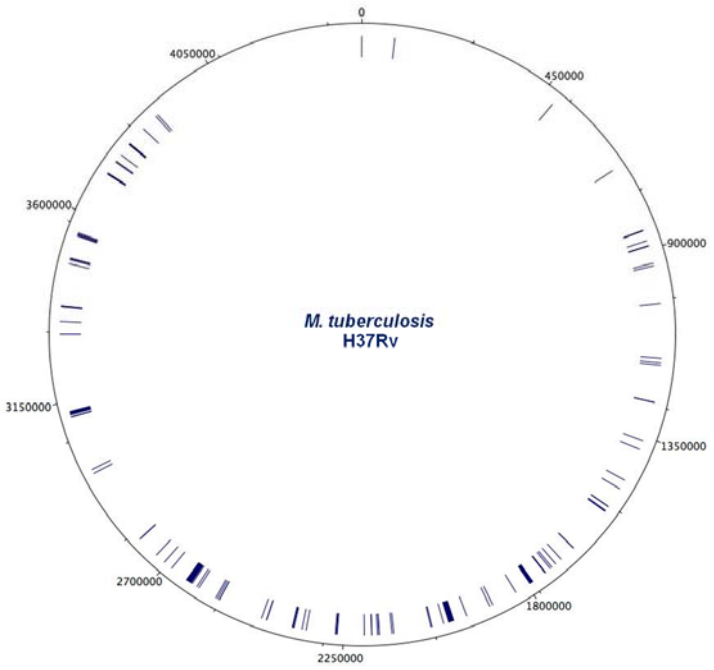


Figure 42: Distribution of the genomic regions which can be amplified with the primer collection generated in this work (Table 5 of Annex I) plotted in *M. tuberculosis* H37Rv genome.

Mapping IS*6110* locations in 61 clinical isolates of *M. tuberculosis*

With the aim of obtaining a global view of the distribution of IS*6110* throughout the genome of *M. tuberculosis* and compare them between Beijing and non-Beijing strains, a scrutiny of points of insertion of IS*6110* was performed with 61 clinical isolates of *M. tuberculosis*. These strains were selected for its high copy number of IS*6110* (HCS). Seventeen were Beijing strains and, as it has been explained above, eight of them were previously selected as representatives of this genotype. HMS1301, which presents identical RFLP as GC1237, was used as internal control and H37Rv and GC1237 strains were used as external controls.

The study was carried out by PCR using all the primer collection (Figure 42 and Table 5 of Annex I) and when a match was found, the fragment was sequenced with primers directed outwards (Table 5 of Annex I). By analysis of the sequences, it was obtained the point of insertion, the DR, the orientation and the distance to neighbouring genes of each copy of IS*6110* (Table 7). After this analysis, all the locations were plotted in the H37Rv genomic map and represented with DNA plotter in two different circles, the red one corresponds to points of insertion of IS*6110* in Beijing strains and the blue one to non-Beijing strains. The locations of IS*6110* in GC1237 were included in a separated circle (green) as reference points of insertion (Figure 43).

A total of 160 different points of insertion of IS*6110* were obtained (Table 7) with the primer collection (Table 5 of Annex I). It is stood out that several genes or intergenic regions were observed interrupted by this element in different points and in both orientations. Considering this fact, a total of 81 different regions were disrupted or altered by IS*6110* (Table 7). By this method, the common locations of IS*6110* of these strains were located; the fact that the number of copies of IS*6110* located in the group of non-Beijing strains is much lower (1 to 4) than the group of Beijing strains (at least 10) may indicate that Beijing genotype evolved separately from a common precursor at an early stage. Thus, the parental Beijing strains should carry low copy number of IS*6110* (common with other lineages) and the descendants, more evolved, would carry high copy number.

Copies of IS*6110* were detected randomly around the genome of these strains with some exceptions. One of them is the region around *OriC,* which lacks this element, corroborating the importance of keeping this region intact. In both groups of strains (Beijing and non-Beijing), this element was observed inserted in the general hot-spots such as *ipl* loci or *plc* regions confirming again that IS*6110* has preferential sites. All the strains presented one copy of IS*6110* in the DR region and two copies were obtained in one of the Beijing strains (CAM22) as explained above in this Chapter. The 90.9% of the non-Beijing strains presented the same point of insertion in this region; however, it is noteworthy that in the remaining non-Beijing strains this copy is located in the DR region but in another point, different even among these four strains.

The insertion of IS*6110* in the possible IS*6110* Beijing-hot-spot RD152 was also observed in non-Beijing strains (*plcD* region) indicating that this region is not exclusive of Beijing genotype (Figure 43 and Table 7). However, comparing these IS*6110* points of insertion, it was observed that only Beijing strains presented the deletion of the region of difference RD152. IS-locations within Rv1371, Rv2016, *ctpD*, *idsB* genes and between the two IS*1532*, were only observed in Beijing strains (Figure 43 and 44 and Table 7) indicating again that these regions could be specific hot-spots of Beijing genotype (IS*6110* Beijing-hot-spots). These locations could be the key of some advantages of the Beijing genotype and its study should be taken into account in further research.

According to Fomukong *et al*. (12) and after the analysis done in this study, among the different Beijing strains analysed until now, no IS*6110* has been detected in the DK1 locus (*mmpS1* gene). However, near to the 15% of the group non-Beijing strains presented a copy of IS*6110* in this gene (Table 7). Specifying more, 6 of them presented an IS*6110* at the exact DK1 point of insertion (12). This point is the preferred site of IS*6110* in LCS and the authors defend the idea that its prevalence decreases in

HCS suggesting a separate lineage for HCS and LCS (12). The fact that the 15% of the studied HCS strains present this copy could indicate that the increment in number of copies of IS*6110* is the result of the transposition of another IS*6110* of their genomes.

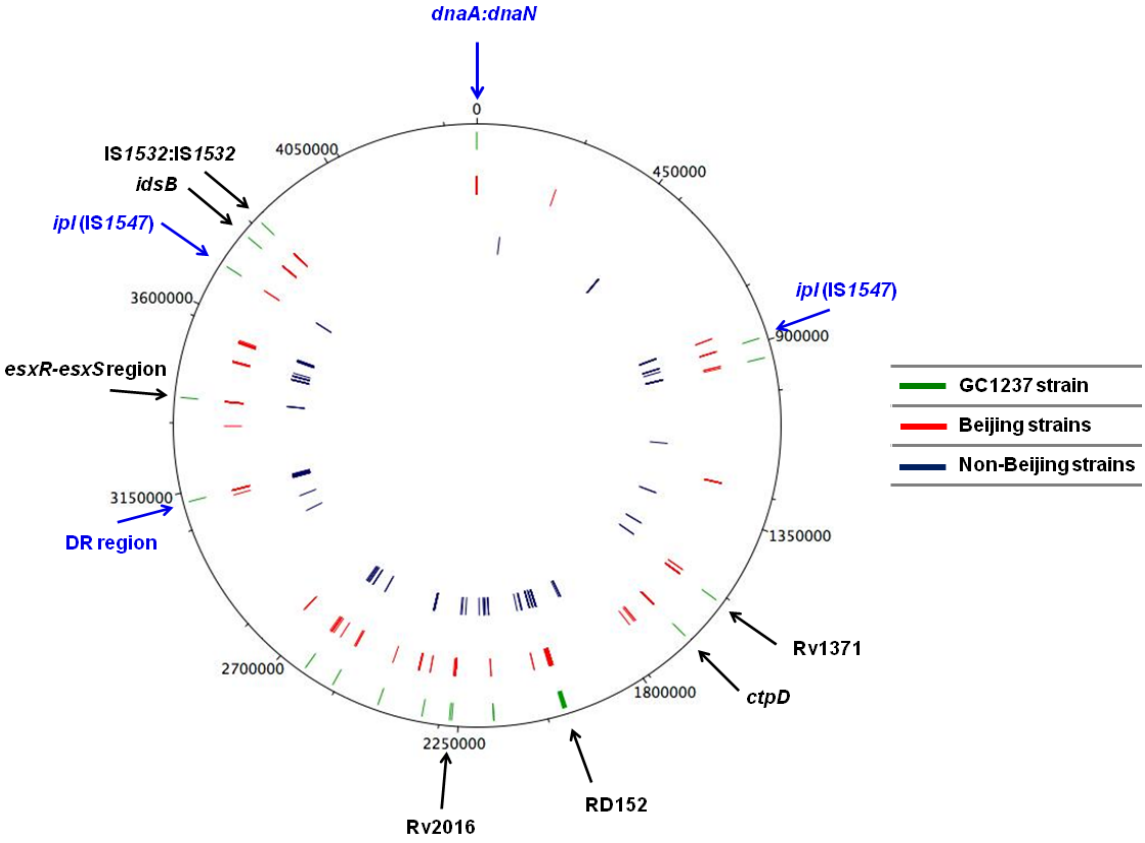


Figure 43: Distribution of IS*6110* thorough the *M. tuberculosis* genome of the 61 clinical isolates. The obtained locations of IS*6110* of each strain were plotted in *M. tuberculosis* H37Rv genome and represented in three concentric circles. The blue one corresponds to the IS-locations in non-Beijing strains. The red one represents the IS-locations in Beijing strains and the green one corresponds to the locations of IS*6110* in GC1237 strain. The general hot-spots of *M. tuberculosis* are indicated by blue arrows and with black arrows it has indicated the possible specific hot-spots of Beijing genotype.

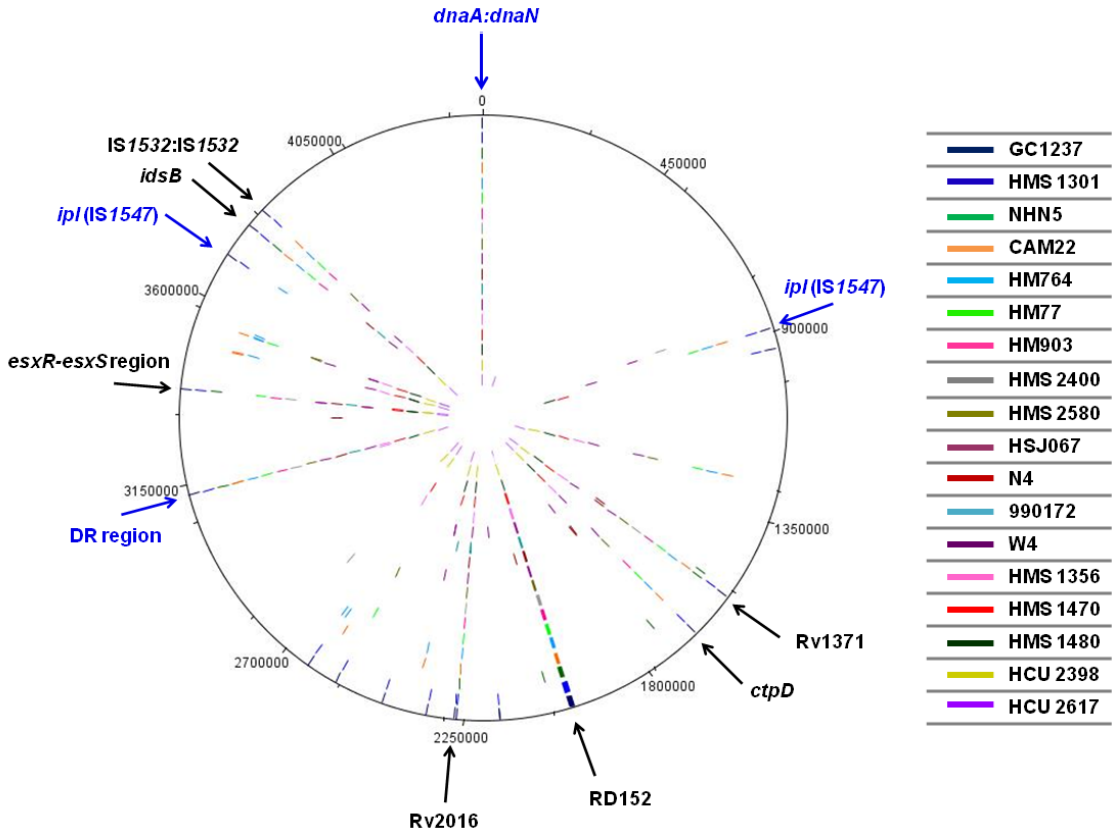


Figure 44: IS*6110* distribution of the 17 Beijing strains. The IS*6110* obtained locations in the 17 Beijing strains were plotted in *M. tuberculosis* H37Rv genome and represented each one in a concentric circle. The IS*6110* locations of GC1237 were included as control. The general hot-spots are indicated by blue arrows and with black arrows the IS*6110*-Beijing-hot-spots.

Two of the sixty-one strains presented a copy of IS*6110* in the promoter region of *phoP* gene (Table 7). One copy was located 132 bp upstream *phoP* gene (HSJ067 Beijing strain) but in opposite orientation to *phoP*, and the other one was located 196 bp upstream this gene (HMS2405 non-Beijing strain) and in the same orientation as *phoP* gene. Although this point is different of the present in MBZ, this last strain could be a possible candidate for studying the effect of IS*6110* in *phoP* gene in a sensible *M. tuberculosis* strain.

Table 7: Points of insertion, DR, orientation and distance to neighboring genes of IS*6110* located copies of the 61 studied strains

GENE	Point of Insertion	DR	Orientation 5'→3'	% Total	% BS	% NBS	% In region	Distance to neighbouring gene
dnaA: dnaN	1594	ATT	+	26,2	100	0	26,2	460 bp upstream *dnaN*
Rv0403c (*mmpS1*), (DK1 region)	483251	ACT	+	1,6	0	2,2	14,7	726 bp upstream *fadD30*
	483298	CCT	-	9,8	0	13,6		67 bp upstream *mmpL1*
	483310	TGA	+	1,6	0	2,2		667 bp upstream *fadD30*
	483537	GTG	+	1,6	0	2,2		440 bp upstream *fadD30*
Rv0755c (*PPE12*): Rv0755A	850087	TAA	+	3,2	0	4,5	4,9	-
	850182	CTT	-	1,6	0	2,2		142 bp upstream *PPE12*
Rv0755A: t*hrV*	850539	TGTC	+	4,9	0	6,8	8,2	-
	850592	-	+	1,6	0	2,2		552 bp upstream *PPE12*
	850601	CCC	-	1,6	0	2,2		561 bp upstream *PPE12*
Rv0756c: Rv0757 (*phoP*)	851412	CGAG	+	1,6	0	2,2	3,2	196 bp upsttream *phoP*
	851479	GTA	-	1,6	5,8	0		13 bp upstream Rv0756c
Rv0794c: Rv0797 (IS*1547*)	888768	TTG	+	1,6	0	2,2	16,3	1620 bp upstream IS*1547*
	888789	CCG	+	4,9	0	6,8		1599 bp upstream IS*1547*
	888846	CAC	+	1,6	0	2,2		1542 bp upstream IS*1547*
	889072	-	-	8,2	29,4	0		436 bp upstream Rv0794c
Rv0835:Rv0836c	932076	GTC	-	3,2	11,7	0	29,5	-
	932203	GTT	-	26,2	0	36,3		-
Rv0840c (*pip*)	937116	ACG	+	1,6	5,8	0	1,6	477 bp upstream Rv0841
Rv1046c	1168966	GAG	+	1,6	0	2,2	1,6	457 bp upstream Rv1047 (IS*1081*)
Rv1135c (*PPE16*)	1262963	AGC	-	18	64,7	0	18	-
Rv1213 (*glgC*)	1356403	GAC	-	1,6	0	2,2	3,2	742 bp upstream Rv1212 (*glgA*)
	1357082	ATC	-	1,6	0	2,2		1421 bp upstream Rv1212 (*glgA*)
Rv1319c	1481531	GGC	+	3,2	0	4,5	3,2	-
Rv1358:Rv1359	1530161	GCG	+	1,6	5,8	0	1,6	12 bp upstream Rv1359
Rv1359	1530203	CTC	+	1,6	5,8	0	3,2	1145 bp upstream Rv1360
	1530907	GGG	-	1,6	0	2,2		-
Rv1360	1531507	-	-	1,6	0	2,2	1,6	-
Rv1371	1543972	GAGG	-	26,2	94,1	0	26,2	-
Rv1469 (*ctpD*)	1657016	CGT	-	21,3	76,4	0	21,3	295 bp upstream Rv1468 (*PE_PGRS29*)
Rv1527c (*pks5*)	1724406	CCGA	-	3,2	11,7	0	3,2	2346 bp upstream Rv1526c

% Total: % of the indicated location in the 61 studied strains.
% BS: % of the indicated location in Beijing strains (17 strains).
% NBS: % of the indicated location in Non-Beijing strains (44 strains).
% In the Region: % of the indicated location in a gene or in an intergenic region in the 61 strains.
DR: Direct Repeat of each copy of IS*6110*, the absence of DR is indicated by minus.
Orientation 5'→3': the orientation of IS*6110* is indicated by + or – direct or complementary chain respectively.
Candidate locations to act as a mobile promoter.

Table 7: Continuation, part 2

GENE	Point of Insertion	DR	Orientation 5'→3'	% Total	% BS	% NBS	% In region	Distance to neighbouring gene
Rv1528c (*papA4*)	1729409	CTT	-	1,6	5,8	0	1,6	1000 bp upstream Rv1527c (*pks5*)
Rv1542c (*glbN*): Rv1543	1745059	ACC	+	1,6	5,8	0	1,6	5 bp upstream Rv1543
Rv1664 (*pks9*)	1889068	CGG	-	3,2	0	4,5	4,9	-
	1889623-1889816	-	+	1,6	0	2,2		1410 bp upstream Rv1665 (*pks11*)
Rv1669	1895860	-	+	1,6	0	2,2	1,6	260 bp upstream Rv1670
Rv1754c	1986625	GTT	-	8,2	0	11,3	22,9	1850 bp upstream Rv1753c (*PPE24*)
	1986626	TGTTC	+	11,4	0	15,9		-
	1986726	ATC	-	1,6	0	2,2		1951 bp upstream Rv1753c (*PPE24*)
	1986820	TAC	-	1,6	0	5,8		
Rv1754-Rv1765c (RD152)	1986638-1998625	-	-	27,8	100	0	27,8	-
Rv1755c (*plcD*)	1987008	CGG	-	3,2	0	4,5	13,1	338 bp upstream Rv1754c
	1987018	CAG	+	1,6	0	2,2		2024 bp upstream Rv1758 (*cut1*)
	1987129-1987556	-	-	1,6	0	2,2		459 bp upstream Rv1754c
	1987211	GAAG	+	1,6	0	2,2		1831 bp upstream Rv1758 (*cut1*)
	1987458-1987544	-	-	3,2	0	4,5		788 bp upstream Rv1754c
	1987601	GAC	+	1,6	0	2,2		1441 bp upstream Rv1758 (*cut1*)
Rv1765c	1998278	-	-	1,6	0	2,2	1,6	-
Rv1765c:Rv1765A	1998623	-	-	3,2	11,7	0	11,8	108 bp upstream Rv1765c
	1998698	-	-	1,6	0	2,2		183 bp upstream Rv1765c
	1998792	-	-	3,2	0	4,5		277 bp upstream Rv1765c
	1998800	-	-	1,6	0	2,2		285 bp upstream Rv1765c
	1998849	-	-	1,6	0	2,2		334bp upsream Rv1765c
Rv1777 (*cyp144*)	2010924	CTC	-	8,2	0	11,3	9,8	369 bp upstream Rv1776c
	2011104	-	-	1,6	0	2,2		549 bp upstream Rv1776c
Rv1798 (*eccA5*)	2039012	CGGC	-	1,6	0	2,2	1,6	-
Rv1800 (*PPE28*): Rv1801(*PPE29*)	2041428	TTT	+	1,6	0	2,2	6,5	573 bp upstream *PPE29*
	2041741	GCG	-	3,2	11,7	0		-
	2041820	CGCC	-	1,6	0	2,2		-
Rv1814 (*erg3*)	2057398	-	-	1,6	0	2,2	1,6	-
Rv1917c (*PPE34*)	2163392	TTA	+	1,6	5,8	0	8,2	-
	2163404	CCAG	+	1,6	0	2,2		-
	2163461	TAA	-	1,6	0	2,2		-
	2163473	CTG	+	1,6	0	2,2		-
	2163649	TCC	-	1,6	5,8	0		-

% Total: % of the indicated location in the 61 studied strains.
% BS: % of the indicated location in Beijing strains (17 strains).
% NBS: % of the indicated location in Non-Beijing strains (44 strains).
% In the Region: % of the indicated location in a gene or in an intergenic region in the 61 strains.
DR: Direct Repeat of each copy of IS*6110*, the absence of DR is indicated by minus.
Orientation 5'→3': the orientation of IS*6110* is indicated by + or – direct or complementary chain respectively.
Candidate locations to act as a mobile promoter.

Table 7: Continuation, part 3

GENE	Point of Insertion	DR	Orientation 5'→3'	% Total	% BS	% NBS	% In region	Distance to neighbouring gene
Rv1922	2174919	-	+	1,6	0	2,2	1,6	254 bp upstream Rv1923 (*lipD*)
Rv1928c	2180845	AAG	-	3,2	0	4,5	3,2	-
Rv1947	2198330	-	+	1,6	0	2,2	1,6	-
Mb2048c (MT2081)	2248279	TGC	+	1,6	0	2,2	1,6	-
Rv2015c	2262185	ACG	-	1,6	0	2,2	1,6	-
Rv2016	2263619	TCC	+	1,6	5,8	0	27,8	370 bp upstream Rv2017
	2263627	AGG	-	24,5	88,2	0		555 bp upstream Rv2015
	2263778	CCT	-	1,6	5,8	0		706 bp upstream Rv2015
Rv2019	2266015	CGCC	+	1,6	0	2,2	3,2	-
	2266167	GAC	-	1,6	5,8	0		-
Mb2047c: Mb2048c	2268726-2268693	GAA	+	1,6	5,8	0	1,6	-
Rv2077A:Rv2078	2334686	AGG	+	1,6	5,8	0	3,2	393 bp upstream Rv2078
	2334688	AGG	-	1,6	5,8	0		92 bp upstream Rv2077A
Rv2103c	2364232	-	-	1,6	0	2,2	1,6	-
Rv2104c: Rv2107 (*PE22*)	2364548	-	+	1,6	0	2,2	27,8	-
	2364912	-	+	1,6	0	2,2		-
	2365243	-	+	1,6	5,8	0		-
	2365414	-	+	3,2	0	3,2		-
	2366894	CAC	+	6,5	23,5	0		-
	2366896	CTA	-	6,5	0	9,1		2376 bp upstream Rv2103c
	2366897	CAC	-	4,9	17,6	0		2377 bp upstream Rv2103c
	2367158-2367279	-	-	1,6	0,00	2,2		2638 bp upstream Rv2103c
Rv2107 (*PE22*): Rv2108 (*PPE36*)	2367672	CGA	-	1,6	5,8	0	3,2	-
	2367679	CGAG	+	1,6	5,8	0		32 bp upstream *PPE36*
Rv2180c	2442348	AGC	-	1,6	1,6	0	1,6	31 bp upstream Rv2179c
Rv2282c	2555228	AAG	+	1,6	5,8	0	3,2	713 bp upstream Rv2283
	2555718	AAT	-	1,6	5,8	0		-
Rv2286c	2559506	GAT	-	1,6	5,8	0	4,9	-
	2559556	AAT	+	1,6	0	2,2		147 bp upstream Rv2287 (*yjcE*)
	2559570	ATC	-	1,6	0	2,2		-
Rv2332 (*mez*)	2605438	AAT	-	1,6	5,8	0	1,6	-
Rv2336	2610861	GCC	+	4,9	0	6,8	4,9	-
Rv2349c (*plcC*)	2627275	AAC	-	1,6	0	2,2	4,9	295 bp upstream Rv2348c
	2627494	TCA	+	1,6	0	2,2		-
	2627829	CAG	-	1,6	5,8	0		849 bp upstream Rv2348c

% Total: % of the indicated location in the 61 studied strains.
% BS: % of the indicated location in Beijing strains (17 strains).
% NBS: % of the indicated location in Non-Beijing strains (44 strains).
% In the Region: % of the indicated location in a gene or in an intergenic region in the 61 strains.
DR: Direct Repeat of each copy of IS*6110*, the absence of DR is indicated by minus.
Orientation 5'→3': the orientation of IS*6110* is indicated by + or – direct or complementary chain respectively.
Candidate locations to act as a mobile promoter.

Table 7: Continuation, part 4

GENE	Point of Insertion	DR	Orientation 5'→3'	% Total	% BS	% NBS	% In region	Distance to neighbouring gene
Rv2352c (*PPE38*)	2633979	CGC	-	1,6	0	2,2	8,2	1904 bp upstream Rv2351c (*plcA*)
	2634022-2634048	-	+	6,5	23,5	0		-
Rv2352c (*PPE38*): Rv2353c (*PPE39*)	2634148	TCC	-	1,6	0	2,2	1,6	50 bp upstream Rv2352c (*PPE38*)
Rv2356c (*PPE40*)	2639543	CCC	-	1,6	0	2,2	1,6	3951 bp upstream Rv2353c (*PPE39*)
Rv2356c (*PPE40*): Rv2357c (*glyS*)	2639612	-	-	1,6	5,8	0	1,6	77 bp upstream Rv2356c (*PPE40*)
Rv2435c	2732535	TAGG	+	1,6	5,8	0	4,9	695 bp upstream Rv2436 (*rbsK*)
	2732535-2732670	-	-	1,6	5,8	0		1975 bp upstream Rv2434c
	2732604	CGA	+	1,6	5,8	0		626 bp upstream Rv2436 (*rbsK*)
Rv2664:Rv2665	2982432	TGGC	-	1,6	0	2,2	1,6	-
Rv2807	3114403	TCA	-	1,6	0	2,2	4,9	-
	3114442	CAC	-	1,6	5,8	0		-
	3114795	GAT	+	1,6	0	2,2		251 bp upstream Rv2808
Rv2807:Rv2808	3114873	AGAT	+	1,6	0	2,2	1,6	173 bp upstream Rv2808
Rv2808	3115131	CGA	-	1,6	0	2,2	3,2	-
	3115231	CGA	-	1,6	0	2,2		-
Rv2809	3115489	GTC	-	1,6	0	2,2	1,6	-
Rv2813-Rv2820c, DR region and RD207	3119420	TCC	-	1,6	0	2,2	100	-
	3119936	-	-	1,6	0	2,2		-
	3120515	GAC	-	1,6	0	2,2		-
	3120523	GGG	-	65,5	0	90,9		-
	3125228	GGTG	-	1,6	0	2,2		-
	3125708	CGC	+	1,6	0	2,2		-
	3125851	CTT	-	1,6	0	2,2		-
	3127927	-	+	27,8	100	0		-
Rv2957	3310163	GTT	-	1,6	5,8	0	1,6	-
***PE27A*: Rv3019c (*esxR*)**	3378553	GGC	-	21,3	76,4	0	21,3	310 bp upstream Rv3018c (*PPE46*)
Rv3019c (*esxR*)	3378898	CTGC	-	4,9	0	6,8	4,9	655 bp upstream Rv3018c (*PPE46*)
Rv3019c (*esxR*): Rv3020c(*esxS*)	3379027	GCAG	+	9,8	35,2	0	9,8	-
Rv3113	3480373	CAG	+	24,5	0	34	24,5	386 bp upstream Rv3114
Rv3125c (*PPE49*)	3491592	CTC	+	3,2	11,7	0	3,2	-
Rv3128c	3493192	CCT	-	1,6	0	2,2	19,6	-
	3493843	AAC	+	3,2	0	4,5		817 bp upstream Rv3129
	3493844-3493908	CGA	-	13,1	47,0	0		-
	3494114	ACC	-	1,6	0	2,2		-

% Total: % of the indicated location in the 61 studied strains.
% BS: % of the indicated location in Beijing strains (17 strains).
% NBS: % of the indicated location in Non-Beijing strains (44 strains).
% In the Region: % of the indicated location in a gene or in an intergenic region in the 61 strains.
DR: Direct Repeat of each copy of IS*6110*, the absence of DR is indicated by minus.
Orientation 5'→3': the orientation of IS*6110* is indicated by + or – direct or complementary chain respectively.
Candidate locations to act as a mobile promoter.

Table 7: continuation, part 5

GENE	Point of Insertion	DR	Orientation 5'→3'	% Total	% BS	% NBS	% In region	Distance to neighbouring gene
Rv3128c:Rv3129	3494372	AAG	-	1,6	0	2,2	3,2	-
	3494395	TAC	+	1,6	0	2,2		265 bp upstream Rv3129
Rv3138 (*pflA*)	3504821	TAC	+	1,6	0	2,2	1,6	542 bp upstream Rv3139 (*fadE24*)
Rv3178:Rv3179	3547342	ATG	+	3,2	11,7	0	3,2	276 bp upstream Rv3179
Rv3179:Rv3180c	3549051	GTCG	-	1,6	0	2,2	21,3	-
	3549176	AAC	+	1,6	0	2,2		-
	3549199	CGG	-	18	64,7	0		-
Rv3183	3550960	ACC	-	1,6	5,8	0	6,5	-
	3550962	-	-	1,6	0	2,2		-
	3550996	CCTG	-	1,6	0	2,2		-
	3551030	GTCG	-	1,6	0	2,2		-
Rv3183:Rv3188	3551139	ATG	-	1,6	0	2,2	13,1	-
	3551220	GAC	+	1,6	0	2,2		-
	3551229	CTG	+	1,6	0	2,2		-
	3551230	GAA	-	1,6	0	2,2		-
	3552683	GCC	+	1,6	5,8	0		-
	3552701	CGG	+	1,6	0	2,2		-
	3554294	-	-	1,6	0	2,2		-
	3559402	-	+	1,6	0	2,2		-
Rv3189	3555250	CGCG	+	1,6	0	2,2	1,6	-
Rv3190c	3555513	CGC	-	1,6	5,8	0	3,2	-
	3555753	TGT	+	1,6	0	2,2		-
Rv3323c (*moaX*)	3709336	CTT	-	1,6	0	2,2	3,2	284 bp upstream Rv3322c
	3709622	GCA	-	1,6	0	2,2		570 pb upstream Rv3322c
Rv3324A: Rv3327 (IS*1547*)	3711737	GGC	-	3,2	11,7	0	3,2	-
Rv3382c (*lytB1*)	3796689	TGG	+	1,6	5,8	0	1,6	-
Rv3383c (*idsB*)	3797823	ATC	+	13,1	47	0	13,1	-
Rv3427c:Rv3428c	3844681	CGGG	-	21,3	76,4	0	22,9	-
	3844737	CCCG	-	1,6	0	2,2		-

% Total: % of the indicated location in the 61 studied strains.
% BS: % of the indicated location in Beijing strains (17 strains).
% NBS: % of the indicated location in Non-Beijing strains (44 strains).
% In the Region: % of the indicated location in a gene or in an intergenic region in the 61 strains.
DR: Direct Repeat of each copy of IS*6110*, the absence of DR is indicated by minus.
Orientation 5'→3': the orientation of IS*6110* is indicated by + or – direct or complementary chain respectively.
Candidate locations to act as a mobile promoter.

The 20% of the located copies of IS*6110* in the studied strains could act as mobile promoter

As already mentioned in Chapter 1 and 2, different studies have indicated that when IS*6110* is inserted in the same orientation as, and close enough to, a downstream gene could potentially function as a promoter (28, 33). The orientation of the located copies of IS*6110* in the 61 strains and the distance to the close genes were analysed in order to test the promoter function of this element. 32 locations of IS*6110* were located close enough (less than 400 bp) and with the same orientation as the neighbouring gene (Table 7). Among these 32 candidate locations to act as a mobile promoter, 4 of them were observed in two different strains and 23 in just one single strain. This result is indicating the enormous variability and genomic plasticity that IS*6110* could generate among strains. The remaining 5 locations were observed in a higher number of strains; being 3 of these locations observed in non-Beijing strains and 2 in Beijing strains (Table 7). It is noteworthy that one of the two frequent locations among Beijing strains corresponds to the copy located in *ctpD* gene and upstream Rv1468c gene and, as already explained in chapter 2, it was demonstrated that this copy is acting as a promoter inside monocytes (28). The fact that this location is quiet frequent in Beijing genotype and even is specific of this family could be one of the special features of Beijing family in terms of virulence and transmission.

Analysis of the Direct Repeats of each IS*6110* obtained copy in the 61 studied strains

The DR generated by the mechanism of the transposition of IS*6110* was determined with the sequence analysis of the flanking regions of each IS*6110*obtained copy of in the 61 strains. The DR of 3-4 nucleotides was detected 128 of the 160 located copies of IS*6110* and the other 32 copies were not flanked by DRs. Genomic regions loss and genomic reorganizations were observed in these cases. It is stood out that in one copy of IS*6110*, located in, it was detected a Direct Repeat of 5 nucleotides.

Although in some of the studied HCS the number of IS*6110* located copies was less than 5, it was possible to observe some of them without DR corroborating the idea that the probability of rearrangement process between copies rises when the number of those increases producing more variability among strains.

Comparison among IS*6110* locations and the distribution of the ORFs across the eleven functional categories of *M. tuberculosis*

The 81 regions affected or altered for IS*6110* were grouped in the eleven functional categories of *M. tuberculosis*. The 68% of the regions corresponded to interrupt ORFs and the remainder were intergenic regions. Comparing the current distribution of ORFs across the eleven functional categories and the 81 regions affected for IS*6110* it is possible to observe similar proportions (Figure 45).

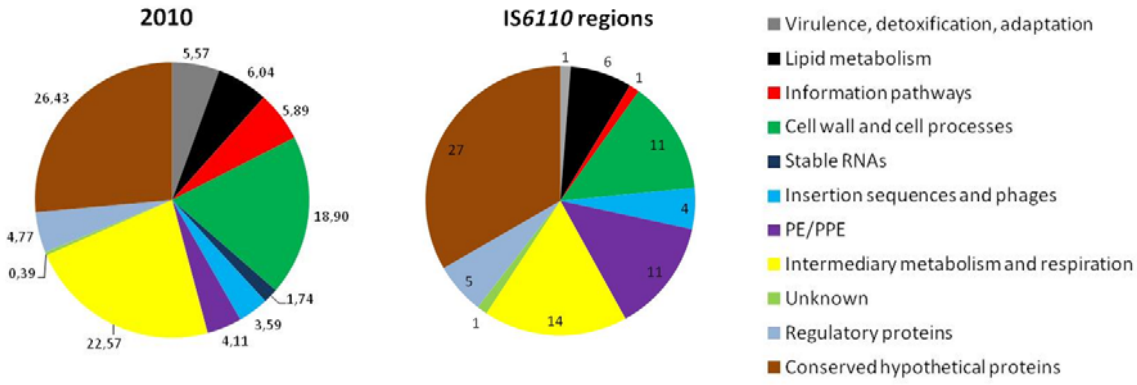


Figure 45: Distribution of the genes and the IS*6110* of *M. tuberculosis* across the eleven functional categories. The 81 obtained regions altered for the presence of IS6110 were classified in the 11 functional categories of *M. tuberculosis.*

An accurate dendrogram of the 61 strains was obtained based on IS*6110* points of insertion

The informatic analysis of spoligotyping results is carried out by Bionumerics program and is based in presence/absence or numerical analysis 1/0 of a specific sequence of *M. tuberculosis*. Based on this idea, all the obtained IS-locations were arranged in columns of an excel sheet and the 61 strains, and the two controls, GC1237 and H37Rv, were arranged in rows. When a strain presented a point of insertion was given the number 1 and if not, was assigned with the number 0. After that, the document was introduced in SpolDB4 database as a new informatic event, included to the RFLP and the sopligotyping of each strain and analysed by Bionumerics program. The obtained dendrogram (Figure 47) shows that Beijing strains were grouped better than by RFLP (Figure 46). Furthermore, and despite having very few located copies of IS*6110* in non-Beijing strains, they were also grouped in families. However, in the case of non-Beijing strains the grouping was no perfect.

Comparing with spoligotyping and RFLP techniques, with this dendrogram it is possible to obtain information of each clone or strain, information not provided by the spoligotyping and provided by RFLP, and the families are also grouped. As Figure 45 shows, although RFLP groups genotypes, if the RFLP of a Beijing strains is quite different from the others, this is not in the group. This method offers advantages over the other two as it is quickly and more economic. Moreover, as is based on PCR it is possible to complete in one day.

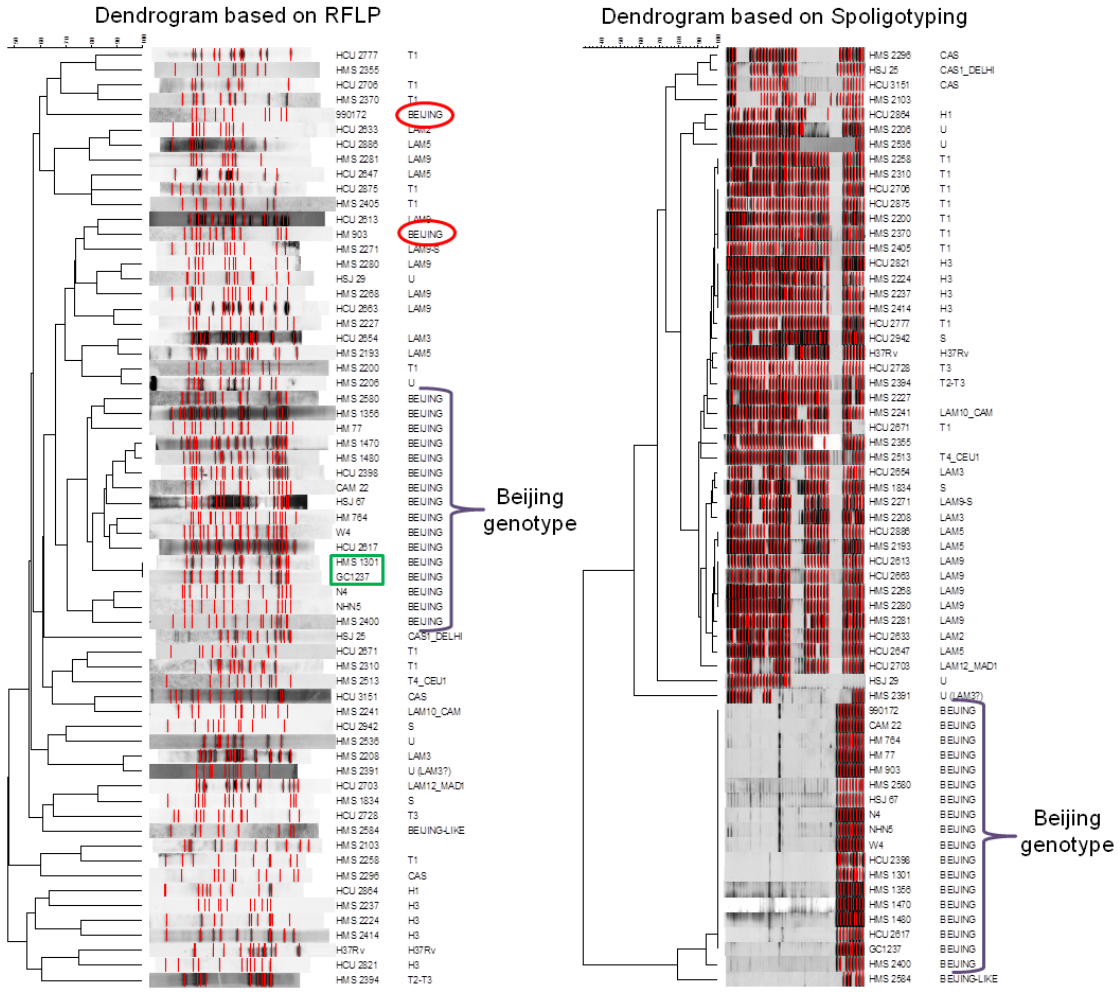


Figure 46: Comparison of two Dendrogrames based on RFLP and dendrogram spolygotiping of the 61 strains used in this work. The RFLP/Spoligotyping of GC1237 and H37Rv strains were included in the dendrogram. The Beijing genotype is showed in both cases.

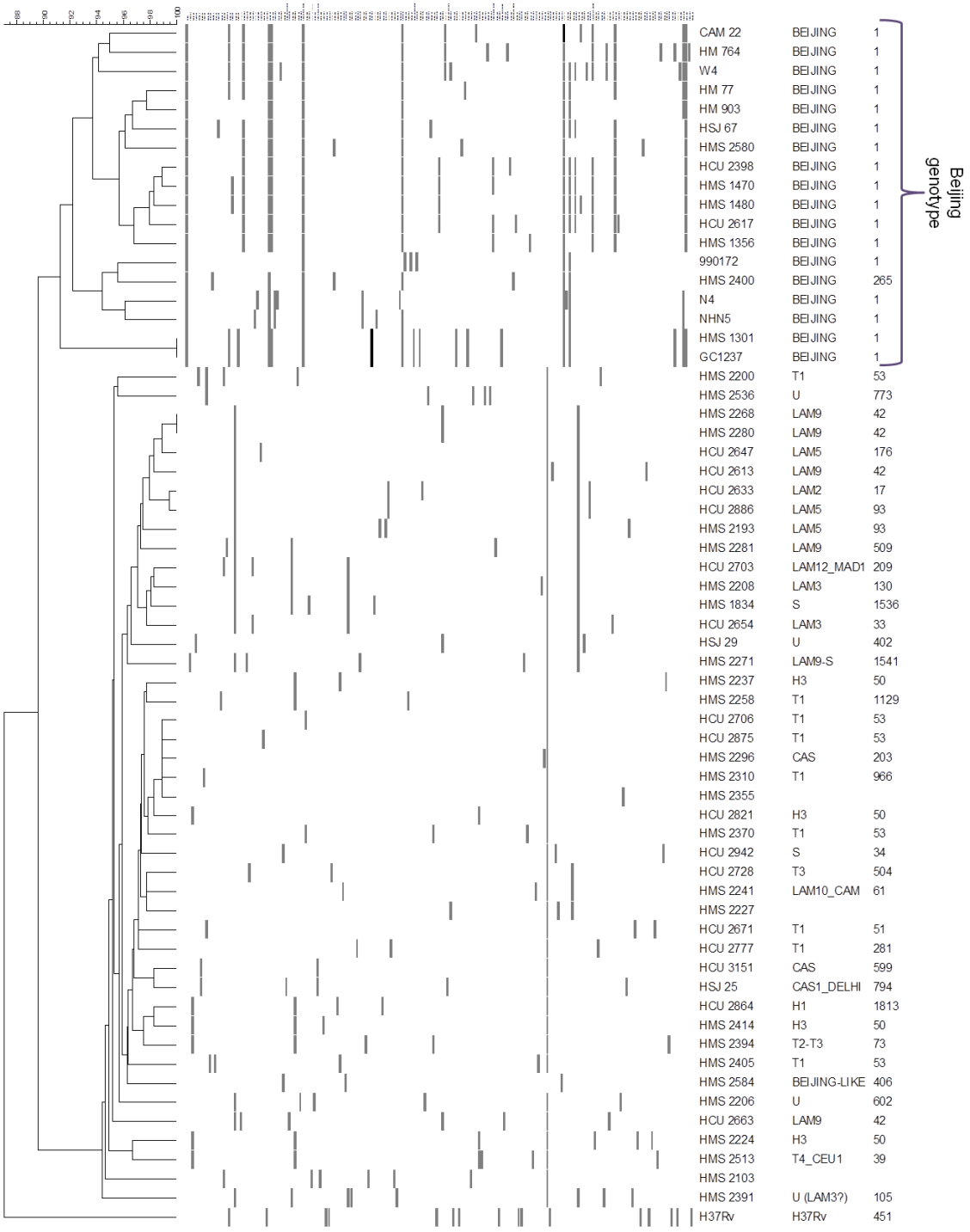


Figure 47: Dendrogram based on points of insertion of IS*6110* of the 61 strains used in this work. The points of insertion of GC1237 and H37Rv strains were included.

Conclusions

- A primer collection, which amplifies locations of IS*6110* in MTBC strains, was generated in this study.
- 160 points of insertion of IS*6110* were obtained. These points of insertion were grouped in 81 regions due to in some cases it was observed the interruption of a gene by IS*6110* in different points.
- Random distribution of IS*6110* along the genome was observed both, in the reference *M. tuberculosis* strains of the database and in the 61 studied strains.
- The common locations of IS*6110* of Beijing strains were obtained.
- The common locations of the group of used Beijing strains also common with the group of non-Beijing strains were obtained.
- The hot-spots of IS*6110*: DR region, both *ipl*, and *plcABC* and *plcD* regions were observed in the two groups of HCS (Beijing and non-Beijing) being these general hot-spots of *M. tuberculosis*.
- IS-locations within Rv1371, Rv2016, *ctpD*, *idsB* genes and between the two IS*1532*, were only observed in Beijing strains indicating that these regions could be specific hot-spots of Beijing genotype (IS*6110* Beijing-hot-spots).
- The 80% of the located copies have DR indicating that its presence was due to a transposition event. The remaining copies lack on DR and genomic reorganization was observed.
- A dendrogram based on the points of insertion of IS*6110* classifies more accurate the *M. tuberculosis* families and information of each clone is also obtained.

References

1. **Alonso, H., J. I. Aguilo, S. Samper, J. A. Caminero, M. I. Campos-Herrero, B. Gicquel, R. Brosch, C. Martin, and I. Otal.** Deciphering the role of IS6110 in a highly transmissible Mycobacterium tuberculosis Beijing strain, GC1237. Tuberculosis (Edinb).
2. **Asiimwe, B. B., M. L. Joloba, S. Ghebremichael, T. Koivula, D. P. Kateete, F. A. Katabazi, A. Pennhag, R. Petersson, and G. Kallenius.** 2009. DNA restriction fragment length polymorphism analysis of Mycobacterium tuberculosis isolates from HIV-seropositive and HIV-seronegative patients in Kampala, Uganda. BMC Infect Dis **9:**12.
3. **Barlow, R. E., D. M. Gascoyne-Binzi, S. H. Gillespie, A. Dickens, S. Qamer, and P. M. Hawkey.** 2001. Comparison of variable number tandem repeat and IS6110-restriction fragment length polymorphism analyses for discrimination of high- and low-copy-number IS6110 Mycobacterium tuberculosis isolates. J Clin Microbiol **39:**2453-7.
4. **Beggs, M. L., K. D. Eisenach, and M. D. Cave.** 2000. Mapping of IS6110 insertion sites in two epidemic strains of Mycobacterium tuberculosis. J Clin Microbiol **38:**2923-8.
5. **Bifani, P. J., B. Mathema, Z. Liu, S. L. Moghazeh, B. Shopsin, B. Tempalski, J. Driscol, R. Frothingham, J. M. Musser, P. Alcabes, and B. N. Kreiswirth.** 1999. Identification of a W variant outbreak of Mycobacterium tuberculosis via population-based molecular epidemiology. Jama **282:**2321-7.
6. **Bifani, P. J., B. B. Plikaytis, V. Kapur, K. Stockbauer, X. Pan, M. L. Lutfey, S. L. Moghazeh, W. Eisner, T. M. Daniel, M. H. Kaplan, J. T. Crawford, J. M. Musser, and B. N. Kreiswirth.** 1996. Origin and interstate spread of a New York City multidrug-resistant Mycobacterium tuberculosis clone family. Jama **275:**452-7.
7. **Caminero, J. A., M. J. Pena, M. I. Campos-Herrero, J. C. Rodriguez, I. Garcia, P. Cabrera, C. Lafoz, S. Samper, H. Takiff, O. Afonso, J. M. Pavon, M. J. Torres, D. van Soolingen, D. A. Enarson, and C. Martin.** 2001. Epidemiological evidence of the spread of a Mycobacterium tuberculosis strain of the Beijing genotype on Gran Canaria Island. Am J Respir Crit Care Med **164:**1165-70.
8. **Casart, Y., L. Turcios, I. Florez, R. Jaspe, E. Guerrero, J. de Waard, D. Aguilar, R. Hernandez-Pando, and L. Salazar.** 2008. IS6110 in oriC affects the morphology and growth of Mycobacterium tuberculosis and attenuates virulence in mice. Tuberculosis (Edinb) **88:**545-52.
9. **Chauhan, D. S., V. D. Sharma, D. Parashar, A. Chauhan, D. Singh, H. B. Singh, R. Das, B. M. Aggarwal, B. Malhotra, A. Jain, M. Sharma, V. K. Kataria, J. K. Aggarwal, M. Hanif, A. Shahani, and V. M. Katoch.** 2007. Molecular typing of Mycobacterium tuberculosis isolates from different parts of India based on IS6110 element polymorphism using RFLP analysis. Indian J Med Res **125:**577-81.
10. **Dale, J. W., H. Al-Ghusein, S. Al-Hashmi, P. Butcher, A. L. Dickens, F. Drobniewski, K. J. Forbes, S. H. Gillespie, D. Lamprecht, T. D. McHugh, R. Pitman, N. Rastogi, A. T. Smith, C. Sola, and H. Yesilkaya.** 2003. Evolutionary relationships among strains of Mycobacterium tuberculosis with few copies of IS6110. J Bacteriol **185:**2555-62.
11. **Fang, Z., and K. J. Forbes.** 1997. A Mycobacterium tuberculosis IS6110 preferential locus (ipl) for insertion into the genome. J Clin Microbiol **35:**479-81.
12. **Fomukong, N., M. Beggs, H. el Hajj, G. Templeton, K. Eisenach, and M. D. Cave.** 1997. Differences in the prevalence of IS6110 insertion sites in Mycobacterium tuberculosis strains: low and high copy number of IS6110. Tuber Lung Dis **78:**109-16.
13. **Fomukong, N. G., J. W. Dale, T. W. Osborn, and J. M. Grange.** 1992. Use of gene probes based on the insertion sequence IS986 to differentiate between BCG vaccine strains. J Appl Bacteriol **72:**126-33.
14. **Fomukong, N. G., T. H. Tang, S. al-Maamary, W. A. Ibrahim, S. Ramayah, M. Yates, Z. F. Zainuddin, and J. W. Dale.** 1994. Insertion sequence typing of Mycobacterium tuberculosis: characterization of a widespread subtype with a single copy of IS6110. Tuber Lung Dis **75:**435-40.
15. **Hanekom, M., G. D. van der Spuy, E. Streicher, S. L. Ndabambi, C. R. McEvoy, M. Kidd, N. Beyers, T. C. Victor, P. D. van Helden, and R. M. Warren.** 2007. A recently evolved sublineage of the Mycobacterium tuberculosis Beijing strain family is associated with an increased ability to spread and cause disease. J Clin Microbiol **45:**1483-90.
16. **Hermans, P. W., D. van Soolingen, E. M. Bik, P. E. de Haas, J. W. Dale, and J. D. van Embden.** 1991. Insertion element IS987 from Mycobacterium bovis BCG is located in a hot-spot

integration region for insertion elements in Mycobacterium tuberculosis complex strains. Infect Immun **59:**2695-705.

17. **Ioerger, T. R., Y. Feng, K. Ganesula, X. Chen, K. M. Dobos, S. Fortune, W. R. Jacobs, Jr., V. Mizrahi, T. Parish, E. Rubin, C. Sassetti, and J. C. Sacchettini.** Variation among genome sequences of H37Rv strains of Mycobacterium tuberculosis from multiple laboratories. J Bacteriol **192:**3645-53.
18. **Kong, Y., M. D. Cave, D. Yang, L. Zhang, C. F. Marrs, B. Foxman, J. H. Bates, F. Wilson, L. N. Mukasa, and Z. H. Yang.** 2005. Distribution of insertion- and deletion-associated genetic polymorphisms among four Mycobacterium tuberculosis phospholipase C genes and associations with extrathoracic tuberculosis: a population-based study. J Clin Microbiol **43:**6048-53.
19. **Kremer, K., J. R. Glynn, T. Lillebaek, S. Niemann, N. E. Kurepina, B. N. Kreiswirth, P. J. Bifani, and D. van Soolingen.** 2004. Definition of the Beijing/W lineage of Mycobacterium tuberculosis on the basis of genetic markers. J Clin Microbiol **42:**4040-9.
20. **Kurepina, N. E., S. Sreevatsan, B. B. Plikaytis, P. J. Bifani, N. D. Connell, R. J. Donnelly, D. van Sooligen, J. M. Musser, and B. N. Kreiswirth.** 1998. Characterization of the phylogenetic distribution and chromosomal insertion sites of five IS6110 elements in Mycobacterium tuberculosis: non-random integration in the dnaA-dnaN region. Tuber Lung Dis **79:**31-42.
21. **Marais, B. J., A. C. Hesseling, H. S. Schaaf, R. P. Gie, P. D. van Helden, and R. M. Warren.** 2009. Mycobacterium tuberculosis transmission is not related to household genotype in a setting of high endemicity. J Clin Microbiol **47:**1338-43.
22. **McEvoy, C. R., A. A. Falmer, N. C. Gey van Pittius, T. C. Victor, P. D. van Helden, and R. M. Warren.** 2007. The role of IS6110 in the evolution of Mycobacterium tuberculosis. Tuberculosis (Edinb) **87:**393-404.
23. **McEvoy, C. R., P. D. van Helden, R. M. Warren, and N. C. Gey van Pittius.** 2009. Evidence for a rapid rate of molecular evolution at the hypervariable and immunogenic Mycobacterium tuberculosis PPE38 gene region. BMC Evol Biol **9:**237.
24. **McEvoy, C. R., R. M. Warren, P. D. van Helden, and N. C. Gey van Pittius.** 2009. Multiple, independent, identical IS6110 insertions in Mycobacterium tuberculosis PPE genes. Tuberculosis (Edinb) **89:**439-42.
25. **Otal, I., A. B. Gomez, K. Kremer, P. de Haas, M. J. Garcia, C. Martin, and D. van Soolingen.** 2008. Mapping of IS6110 insertion sites in Mycobacterium bovis isolates in relation to adaptation from the animal to human host. Vet Microbiol **129:**333-41.
26. **Plikaytis, B. B., J. L. Marden, J. T. Crawford, C. L. Woodley, W. R. Butler, and T. M. Shinnick.** 1994. Multiplex PCR assay specific for the multidrug-resistant strain W of Mycobacterium tuberculosis. J Clin Microbiol **32:**1542-6.
27. **Purwar, S., S. Chaudhari, V. M. Katoch, A. Sampath, P. Sharma, P. Upadhyay, and D. S. Chauhan.** Determination of drug susceptibility patterns and genotypes of Mycobacterium tuberculosis isolates from Kanpur district, North India. Infect Genet Evol **11:**469-75.
28. **Safi, H., P. F. Barnes, D. L. Lakey, H. Shams, B. Samten, R. Vankayalapati, and S. T. Howard.** 2004. IS6110 functions as a mobile, monocyte-activated promoter in Mycobacterium tuberculosis. Mol Microbiol **52:**999-1012.
29. **Sampson, S., R. Warren, M. Richardson, G. van der Spuy, and P. van Helden.** 2001. IS6110 insertions in Mycobacterium tuberculosis: predominantly into coding regions. J Clin Microbiol **39:**3423-4.
30. **Sampson, S. L., R. M. Warren, M. Richardson, G. D. van der Spuy, and P. D. van Helden.** 1999. Disruption of coding regions by IS6110 insertion in Mycobacterium tuberculosis. Tuber Lung Dis **79:**349-59.
31. **Sankar, S., M. Ramamurthy, B. Nandagopal, and G. Sridharan.** An appraisal of PCR-based technology in the detection of Mycobacterium tuberculosis. Mol Diagn Ther **15:**1-11.
32. **Shanmugam, S., N. Selvakumar, and S. Narayanan.** Drug resistance among different genotypes of Mycobacterium tuberculosis isolated from patients from Tiruvallur, South India. Infect Genet Evol **11:**980-6.
33. **Soto, C. Y., M. C. Menendez, E. Perez, S. Samper, A. B. Gomez, M. J. Garcia, and C. Martin.** 2004. IS6110 mediates increased transcription of the phoP virulence gene in a multidrug-resistant clinical isolate responsible for tuberculosis outbreaks. J Clin Microbiol **42:**212-9.
34. **Suffys, P. N., M. E. Ivens de Araujo, M. L. Rossetti, A. Zahab, E. W. Barroso, A. M. Barreto, E. Campos, D. van Soolingen, K. Kremer, H. Heersma, and W. M. Degrave.** 2000. Usefulness of IS6110-restriction fragment length polymorphism typing of Brazilian strains of Mycobacterium

tuberculosis and comparison with an international fingerprint database. Res Microbiol **151:**343-51.

35. **Thierry, D., A. Brisson-Noel, V. Vincent-Levy-Frebault, S. Nguyen, J. L. Guesdon, and B. Gicquel.** 1990. Characterization of a Mycobacterium tuberculosis insertion sequence, IS6110, and its application in diagnosis. J Clin Microbiol **28:**2668-73.
36. **Tsolaki, A. G., S. Gagneux, A. S. Pym, Y. O. Goguet de la Salmoniere, B. N. Kreiswirth, D. Van Soolingen, and P. M. Small.** 2005. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of Mycobacterium tuberculosis. J Clin Microbiol **43:**3185-91.
37. **Turcios, L., Y. Casart, I. Florez, J. de Waard, and L. Salazar.** 2009. Characterization of IS6110 insertions in the dnaA-dnaN intergenic region of Mycobacterium tuberculosis clinical isolates. Clin Microbiol Infect **15:**200-3.
38. **Valway, S. E., M. P. Sanchez, T. F. Shinnick, I. Orme, T. Agerton, D. Hoy, J. S. Jones, H. Westmoreland, and I. M. Onorato.** 1998. An outbreak involving extensive transmission of a virulent strain of Mycobacterium tuberculosis. N Engl J Med **338:**633-9.
39. **Vera-Cabrera, L., M. A. Hernandez-Vera, O. Welsh, W. M. Johnson, and J. Castro-Garza.** 2001. Phospholipase region of Mycobacterium tuberculosis is a preferential locus for IS6110 transposition. J Clin Microbiol **39:**3499-504.
40. **Viana-Niero, C., C. A. Rodriguez, F. Bigi, M. S. Zanini, J. S. Ferreira-Neto, A. Cataldi, and S. C. Leao.** 2006. Identification of an IS6110 insertion site in plcD, the unique phospholipase C gene of Mycobacterium bovis. J Med Microbiol **55:**451-7.
41. **Warren, R. M., S. L. Sampson, M. Richardson, G. D. Van Der Spuy, C. J. Lombard, T. C. Victor, and P. D. van Helden.** 2000. Mapping of IS6110 flanking regions in clinical isolates of Mycobacterium tuberculosis demonstrates genome plasticity. Mol Microbiol **37:**1405-16.
42. **Yang, Z., P. F. Barnes, F. Chaves, K. D. Eisenach, S. E. Weis, J. H. Bates, and M. D. Cave.** 1998. Diversity of DNA fingerprints of Mycobacterium tuberculosis isolates in the United States. J Clin Microbiol **36:**1003-7.

GENERAL CONCLUSIONS

Conclusiones Generales

- La cepa ancestral de *M. tuberculosis* GC1237 perteneciente a la familia Beijing tiene 19 copias de la secuencia de inserción IS*6110*. La localización de todas las copias de este elemento fue posible gracias a la combinación de varias técnicas como son LMPCR y PCR con el uso de la librería de BACs y del DNA genómico de la cepa de estudio GC1237.

- Una de las copias de IS*6110* de la cepa GC1237 se localizó a 31 nucleótidos del inicio del gen esencial Rv2179c. Mediante varios métodos se confirmó la actividad promotora de esta copia de IS*6110.* Además se observó que esta actividad se ve incrementada en el interior del macrófago y sólo está relacionada con la distancia de la IS al gen.

- La copia de GC1237 delante del gen Rv2179c resultó ser única de esta cepa lo que permitió el diseño de un test rápido para su detección basada en un multiplex PCR.

- Cuando IS*6110* actúa como secuencia promotora lo hace de forma independiente al fondo genético.

- La copia de IS entre los genes *dnaA-dnaN* (inserción A1), característica de todas las cepas Beijing, resultó no tener efecto en la expresión del gen *dnaN* tanto en cultivo líquido como en el interior del macrófago.

- La copia de IS*6110* localizada a 75 nucleótidos del gen *phoP* de MBZ actúa como una secuencia promotora de ese gen tanto en cultivo como en el interior del macrófago.

- No se observaron diferencias en la replicación tanto en el interior del macrófago como en ratones de las cepas H37Rv ΔphoP pSO5k y H37Rv ΔphoP pSO7k.

- Las cepas Beijing y las cepas no-Beijing comparten regiones donde IS6110 tiende a insertarse pero no comparten puntos de inserción. Concretamente las localizaciones de IS en los genes *ctpD, idsB*, Rv1371, Rv2016, y entre las dos IS*1532,* son únicos de las cepas Beijing y podrían ser la clave de algunas características especiales de esta familia.

General Conclusions

- *M. tuberculosis* Beijing GC1237 is an ancestral strain which contains 19 copies of the insertion sequence IS*6110*. The location of all its copies was possible with the complementation of different techniques, as LMPCR and PCR, and with the use of BACs library and genomic DNA of GC1237 strain.

- One of the copies of I*S6110* of GC1237 strain was located 31 bp upstream the essential gene Rv2179c. The promoter activity of this copy was demonstrated by several methods. In addition, this activity enhances inside macrophages and is only related to disposition of the sequence with respect to this gene.

- The copy of IS*6110* upstream Rv2179c gene is unique of GC1237 strain, and permitted the design a promising test for its rapid detection based on multiplex PCR.

- When I*S6110* acts as a promoter, acts regardless of the genetic background.

- The well-known insertion A1, one of the genetic markers of all Beijing strains, causes minimal effect in *dnaN* gene expression both in broth and inside macrophages.

- The copy of IS*6110* located 75 bp upstream *phoP* gene of MBZ is acting as a promoter of this gene both in broth and inside macrophages.

- No significant differences were observed in replication experiments both in macrophages and in mice between the strains H37Rv ΔphoP pSO5k y H37Rv ΔphoP pSO7k.

- Beijing strains and non-Beijing strains share insertion regions of IS*6110* but not points of insertion. Concretely, IS-locations within *ctpD, idsB*, Rv1371, Rv2016, genes and between the two IS*1532*, were only observed in Beijing strains and they could be the key of some advantages of the Beijing genotype.

Annex I

Table 1: Primers used in the study of genomic deletions of Regions of Differences (RD) (Chapter 1)

Name	Sequence 5'-3'	Tm ºC	Reference
RD105d	GGAGTCGTTGAGGGTGTTCATCAGCTCAGTC	77.3	Tsolaki *et al.* (8)
RD105r	CGCCAAGGCCGCATAGTCACGGTCG	81.6	Tsolaki *et al.* (8)
RD105int	GCAACACCCGCTTGTCTTTG	62	Hanekom *et al*.(2)
RD108d	CGCCGCGTCTGCCGCTCTT	77.1	Tsolaki *et al.* (8)
RD108r	GACCAGGACACCGCGCAGATACG	75.4	Tsolaki *et al.* (8)
RD110ad	GAACTGCCGCTGATCGAAGAATGG	73.3	Tsolaki *et al.* (8)
RD110ar	GGTGTCGATGCCGTGCTATGTGC	74.5	Tsolaki *et al.* (8)
RD127d	GACTCCGATCTGGAGGAGCTG	67.5	Tsolaki *et al.* (8)
RD127r	CAGGCCGGCCAGAAAGTCCAG	74	Tsolaki *et al.* (8)
RD129d	GAAAGCCGCTCGTGGTCAGGTTG	74.5	Tsolaki *et al.* (8)
RD129r	GCTGCGGCCGGCTCCATC	75.9	Tsolaki *et al.* (8)
RD139BWd	GATCCTGATCATCCCGTTCGCGCTGTGCTT	82	Tsolaki *et al.* (8)
RD139BWr	AGGACCTCGACCTGACCGACCTGGTCAT	78.5	Tsolaki *et al.* (8)
RD142d	TCCGCGACGACGAACAACGAC	74.7	Tsolaki *et al.* (8)
RD142r	TCACTTCCATTTCCAGCGGCAACT	73.1	Tsolaki *et al.* (8)
RD142int	GCTCGAGCATGATCAGCAAAG	62	Hanekom *et al*.(2)
RD149d	CATGTCACCCTGGCCCGACGGGTC	81.16	Tsolaki *et al.* (8)
RD149r	CCAGGCGATCTTCGACCACGGCACAC	81.33	Tsolaki *et al.* (8)
RD150d	TGTGGCGTGGCTCGGCAAATAG	74.7	Tsolaki *et al.* (8)
RD150r	CGGGACGGCAAACGGGTGAT	75.1	Tsolaki *et al.* (8)
RD150int	CACCGGCACTTACCATCTCG	62	Hanekom *et al*.(2)
RD152d	CCGGGTTGAGCAATGCGGATATCAGTGGAC	80.5	Tsolaki *et al.* (8)
RD152r	TGGGATAGTTCAGGTGGCCATCGTGGGCAT	81.9	Tsolaki *et al.* (8)
RD165d	GCCTTCGCTGCCCGTGTTCTG	75.1	Tsolaki *et al.* (8)
RD165r	GCTCGCGGGGTGCTGGTGTTT	76.2	Tsolaki *et al.* (8)
RD166d	CCAGGGTTGGCAGCCGTGCGTCTTGAA	77.7	Tsolaki *et al.* (8)
RD166r	GTACTCGCTGTCGCCAAGCGCATAGAGAAGAT	77.8	Tsolaki *et al.* (8)
RD182ad	CGGTGGGGCTCTGCGTGGTC	76.5	Tsolaki *et al.* (8)
RD182ar	AGGCGTTCGGGGATGCGTCTT	74.6	Tsolaki *et al.* (8)
RD207d	GACGAGTTCGCGCTCAAAATGT	69.4	Tsolaki *et al.* (8)
RD207r	CCCCGGCGAGGAACAGAA	70.8	Tsolaki *et al.* (8)
RD181d	CGCAACGGCCGCGGTGAACT	79.4	Tsolaki *et al.* (8)
RD181r	CGGGCGGCTGCGGGAACCTT	79.6	Tsolaki *et al.* (8)

Table 2: Primers used in the study of *pks15/1* polymorphism and in LMPCR technique (Chapter 1)

Name	Sequence 5'-3'	Tm ºC	Used for	Reference
Pks1I	GCAGGCGATGCGTCATGGGG	77.2	*pks15/1* polymorphism	Constant *et al*.(1)
Pks1J	TCTTGCCCACCGACCCTGGC	75.8		Constant *et al*.(1)
SalGd	TAGCTTATTCCTCAAGGCACGAGC	68.1	LMPCR	Prod'hom *et al.* (5)
SalPt	TCGAGCTCGTGC	49.2		Prod'hom *et al.* (5)
Xmapt	CCGGGCTCGTGC	59.3		Otal *et al.* (4)
ISA1	CCTGACATGACCCCATCCTTTCC	71.4		Mendiola *et al*.(3)
ISA3	GAGGCTGCCTACTACGCTCAACG	69.9		Mendiola *et al*.(3)

Table 3: Primers used for multiplex PCR designed to detect GC1237 strain (Chapter 1)

Name	Sequence 5'-3'	Tm ºC	Reference
dna II	CCACCCACGACACCGCAT	70.8	This work
dna III	GGGCGGTTCAATTGGCTGT	69	This work
Rv2179c fw	CAATACGTGATCGCCGGGAC	69.6	This work
Rv2179c rv	CGAATTCCGTGGATACTGCGTAG	68.2	This work
RD105 fw	AGTCGTTGAGGGTGTTCATCAGC	68.5	This work
RD105 rv	CCAAGGCCGCATAGTCACG	68.8	This work

Table 4: Primers used in the study of gene expression qRT-PCR and RT-PCR (Chapter 2)

Name	Sequence 5'-3'	Tm ºC	Reference
RT-rrnAP1 fw	CCTATGGATATCTATGGATGACCGA	65.5	This work
RT-rrnAP1 rv	GGCGACCCTGCCAGTCTAA	63.2	This work
RTRv0794c d	CGGATCAAAGCGATCGATGT	67.2	This work
RTRv0794c r	CGCCCTGCCGGTGTATC	67.8	This work
RTRv0841 d	CTGGTCGATGACCCCTGTAGTT	66.2	This work
RTRv0841 r	CCAAGAGATGAGCACCATGAATC	66.7	This work
RTRv2078 d	CATTTCCTGTCGCCCAGACT	66.4	This work
RTRv2078 r	TTTCTCGCCCACCGTGAT	66.7	This work
RT Rv2179c fw	CTACGACCATGTAGCGCTGTGT	66.2	This work
RT Rv2179c rv	CAGTTCCCGGGTGAAACG	66.6	This work
RT AroG fw	TGCCAGCCATGAAGCTTTG	67.5	This work
RT AroG rv	CCGTCGGACAACCTCAACAT	67.2	This work
RT Rv2280 fw H	TCCTGGAGTACATTGACAACACTTC	65.1	This work
RT Rv2280 rv H	GGAATACCTAGCTCCAGGTTCTGA	65.4	This work
RT ctpD F	GTGGCGCTGCTGCTGTTTCT	70.1	This work
RT ctpD R	CGCGGCAATCATCAGCAGAT	70.6	This work
RTdnaA F	CCAGACACCACAACCGACAA	67.2	This work
RTdnaA R	TGGCCAACTGTGCTGGTTATC	67.1	This work
RTdnaN d	CGGTGAGACGGTGGTTTTG	66.6	This work
RTdnaN r	CGACGCCGACCACTTCAG	68.3	This work
RT1371 F	GTATACCCAGAGCGAGGTG	59.7	This work
RT1371 R	TCGCTACGAATTGCAGTGCCC	71.3	This work
RTPPE34 F	ATCACCAGTTCCAAGGGTGTG	66.1	This work
RTPPE34 R	CGCCCATCTGTCCGAAGC	69.4	This work
RTRv2016 F	ACCACGCAGGTGACGAG	64	This work
RTRv2016 R	ACCGAAGTCGATGGACG	62.5	This work
RTRv2077A F	GGGTCGAATGAACTACAGGTTGTA	64.9	This work
RTRv2077A R	AATGGTTGGCCGGATTCTG	67	This work
RTRv2077cF	GCACCGAGCACCTTATCGAT	66.3	This work
RTRv2077cR	GCATCTGCAGAAACACGTCTTC	66	This work
RTRv2286cF	CCTGGTAGCGGGCAAGAAG	66.9	This work
RTRv2286cR	CCGATCCGCATCAACGA	67.3	This work
AroGend1	GTCGTGTTCGCAGGTCAGTC	66.1	This work
Rv2180cfw	ACGCACTGATCCTGCATCTG	66.6	This work
Rv2180crv	GGCCACTGCGGCAAAG	67	This work
Rv2179c fw	CAATACGTGATCGCCGGGAC	69.6	This work
Rv2179c rv	CGAATTCCGTGGATACTGCGTAG	68.2	This work

Table 5: Primers used to amplify IS*6110* insertion sequence of MTBC (Chapter 1 and 3)

Name	Sequence 5'-3'	Tm ºC	Reference
IS61	GACCGCGGATCTCTGCGACCATCC	78.8	This work
IS62	ACCATCGCCGCCTCTACC	67.4	This work
Gab1	CTGACCGAGCTGGGTGTGC	69.6	Samper *et al*. (6)
Gab2	TCTGATCTGAGACCTCAGCC	62.2	Samper *et al*. (6)
ISAcom	ATGTCAGGTGGTTCATCGAGGAGG	71.2	This work
NTFfw	CGTGAGGCACCGAGGGTGTTTC	73.9	This work
NTFrv	CGGTAGTAACTGGGCGCGTTCG	73	This work
dnaAI	TCCGAGATGGCCGAGCGCCG	80.5	This work
dnaAII	CCACCCACGACACCGCAT	71	This work
Rv0402cext	GGTGCGCCTCATCACCCAG	71.6	This work
Rv0403cext	GTTCGGCGTTGCCAAACGC	73	This work
Rv0594d	ATGTCCGCTACACGCCGAACCC	74.7	This work
Rv0594r	GGCTCAGCCGGTTGGTGCC	74.5	This work
Rv0756cint	CCTCGTGGGTATTGCGA	63.5	This work
Rv0782ext	ACGAGAACGTCACGGCCC	68.8	This work
BCG2A	GCCGTCCATCCCGGGCATC	76.2	Soto *et al.*(7)
BCG2B	CCATGTTCAAACCGGTGTC	63.9	Soto *et al.*(7)
Rv0783cext	AAACCAACTTCGCGGCCC	69.4	This work
0795F	CCTATCGCCCACCAGACGC	70	This work
0795R	ACACCGTGCGCCTCTACCTGC	72.7	This work
Rv0795d	AGGCTTTGCTGGGCACACAGG	72.6	This work
Rv0795r	GTCCTTCTCACCTAGTGCCG	63.7	This work
Rv0835d	TTGCTCCACTGCTGCCAAGTCGG	76.2	This work
Rv0836cd	AACAATTGGGACCACTTCGAGG	68	This work
pipF	ACGAATGGCTGAAGATGTGAA	64.5	This work
pipR	AATCGGAGGTCAAGTGGA	60.6	This work
Rv0920cr	GCAGCATCCATGCCCTTGAG	70.3	This work
Rv0921r	AATCCGCCAGATTCACGCCG	73.3	This work
Rv1032c out	AGCACAACGGTAACGACC	61.3	This work
Rv1033c int	GGGGTACACGCGCAGTCAA	69.5	This work
Rv1046d	CGTAACTGCTGATATTATCGAGG	66.5	This work
Rv1046r	ACGACGAATGGAATCGCCC	69.4	This work
Rv1051c out	CGATGTCGCGAACCACC	66.8	This work
Rv1052out	AACTTGGCTGAGGCCCTTG	66.5	This work
PPE12out	CCGCCGCGTGCAACTC	70.9	This work
PPE16d	ACAACATGATCGGCATCGGC	69.3	This work
PPE16r	GAAACCGAAGTTGCCGCTGCCC	76	This work
PPE16end1	GCGTATTCAGCCAGCGGAGG	71.1	This work
PPE16end2	CGAAGGATCTCGCGCGG	70.7	This work
glgCd	CATGAGAGAAGTGCCGCACG	69.4	This work
glgCr	AACCACCGCGATGCTGGC	72	This work
Rv1234d	CTCATGCAGGTTGAACGGGTC	68.9	This work
Rv1235r	TACCTCTGATTGTGCTGGTCACG	68.2	This work
Rv1301d	GAGCAGCGTTCGCGTGG	69.8	This work
Rv1301r	CTGGCCGCGTCCACACC	71.9	This work
Rv1319cd	GAGACCGTGACACTTCGCGGG	73.3	This work
Rv1319cr	TACCCGTCGGCGCCGTTGG	77.2	This work
Rv1359d	TTTCGTCTCACCGCGTACCGTG	73.1	This work

Table 5: Continuation

Name	Sequence 5'-3'	Tm ºC	Reference
Rv1359r	GCAGGACGATCGGGGACAACG	75	This work
Rv1359ext	CGTCGCCGCGTTCGCCC	78	This work
Rv1360ext	GACGAATTCACGCATCCGACG	71.9	This work
Rv1368ext	GACAATCCAGGGACGTGAC	63.2	This work
Rv1371ext	TTGGAAATCCACTCACTCAGG	64	This work
Rv1371d	ATCGTTGCAGTACTGGCCGTGGG	74.5	This work
Rv1371r	TATACGGGTGGTGCAGCAGGGTG	73.5	This work
Rv1468cext	GCTAATCATCGAACCGAGGCG	69.7	This work
Rv1469cext	CACGGAGACTACCGACCACAG	67	This work
CtpDF	ACCCCGCCAGAATGGTTAATCC	70.4	This work
CtpDR	TTGTTCCGCAACGCTTGTGC	71.5	This work
CtpDbIS	GAGGCTCCTTTCGACCG	68	This work
Rv1506d	ATCTATCGTCTCCACGCTTCC	64.3	This work
Rv1506r	TAGCCGAGAGCCTTAGCGGC	69.7	This work
FadD25d	GGTCGAATCCTCTCTTCCTGG	66.1	This work
FadD25r	CCAATAGCCGATGGCGACG	70.8	This work
Rv1528cd	ATGTAGGTGATGGCCGC	66	This work
Rv1528cr	GCCAATTGGTGCCCG	64	This work
Rv1541cd	GAGGCCTACCAGACCCG	63.4	This work
Rv1541cr	CGACCATACCGGGCGCCAC	75	This work
Rv1541cf	GCACCAGTCTGAGCCCTG	65	This work
Rv1541ci	ATCCCAAGCTAGCGAAGAGG	65.4	This work
Rv1543d	CAGATCCTTGGTCGCCGCAC	72.4	This work
Rv1543r	GTTGAGCTGCATCGTTCGC	67.2	This work
cyar	TTGGCGGTCAGACCCCTGC	73.1	This work
LeuVd	GGCGATTTCGGATGCGGCG	76.8	This work
Rv1664d	GCCAGCTTTTTGCCATCCTCG	71	This work
Rv1664r	GCATAATTTGAGCCTGCCAGGAC	69.4	This work
Rv1669d	GGTGTTTGAACATTTCGTGGG	66.2	This work
Rv1669r	GGTAGTACCAGGCAGCGTCC	65.8	This work
Rv1728cext	GGTTTCTGAGCCGTAGTCCG	65.9	This work
Rv1729cext	GGGTTTGACTCGGCAGCACC	71.3	This work
Rv1754cd	AGGAACTCAAATCCGGTCAGG	66.2	This work
Rv1754cr	GAACCATGAGTCCAATAGCGGC	68.5	This work
1765	CGGTTCCAACTTGAGCGCGGTC	75.5	Otal *et al.* (4)
1765-ISA1	GTTGGTGCATCATTCCGACC	67.5	Otal *et al.* (4)
plcDd	GTCAGCTGGAAGGTGTATCGC	65.8	This work
plcDr	GGATTTGAGTTCCTCGAGCC	64.8	This work
Rv1758d	ATTGGCGAGAAGTCTATGGGC	66.2	This work
Rv17589d	TTGGTGCCGGCGGCAT	72.6	This work
Rv1761cext	GCGTTGATGTTCGGGATCAC	67.7	This work
Rv1762cint	TCGCTCGATCCAGTAGCC	64.2	This work
Rv1777d	GAGACGTTCGCCGAAAGGC	69.3	This work
Rv1777r	CTGCCTCAATTACCGAGG	60.8	This work
Rv1798d	GTTTGGCACAAGTATCGGT	60	This work
Rv1798r	TCTCATGTTTCGATATATTTGGC	61.5	This work
PPE28d	GGAGGGCATCCAGCAAGCG	72.5	This work
PPE28r	CTTAATCCCTAATCGCGTGGC	66.4	This work

Table 5: Continuation

Name	Sequence 5'-3'	Tm ºC	Reference
Rv1879d	AAACGACCGTAGACCCGAC	63.9	This work
Rv1879r	ATAGATGCGTGCGGCAGTGC	70.7	This work
1883-ISA1	ACGATAATGCCAATTGTTCACCG	68.3	Otal *et al.* (4)
1883-ISA3	GTTGCCAATCGGGTTCTCGCC	73.6	Otal *et al.* (4)
PPE34d	TATTTCCCGCATCCCCCTTACG	71	This work
PPE34r	ATGCCAGACTCCAGGGTGCC	71	This work
PPE34a	AACGTCGCTGGTATCTTCACGG	69	This work
PPE34t	TGCCCGACGTTGTAGATTCCC	70	This work
Rv1917cd	TCGATTCCTAAAGCGGCT	62.4	This work
Rv1917cr	CGAATATGAGGGCCGAG	61.8	This work
Rv1928cd	ATTCCATCAGGTATCGGCCC	66.8	This work
Rv1928cr	CCGTGTGTCGTTCAGTGACG	68	This work
Rv1929d	GATTCAGCTGGCGATCGAG	66.1	This work
Rv1929r	TCACATCGCAGCAGACGGGC	74.4	This work
1947-Rv1948c	GTGCTGGAGCAAGACCTCGC	69.7	Otal *et al.* (4)
1947-ISA3	ACTGCGGATATGGAACTGGG	66	Otal *et al.* (4)
Rv2015rev	AATCCTTGATGCCCCAAACC	66.8	This work
Rv2016d	ATGCATAGCCGGTGTCTGTCC	68.3	This work
Rv2016r	AATGCGAACCTCTTCCAGCC	67.5	This work
Rv2017rev	AATTCGATCTGGGTGAGTCC	62.8	This work
Rv2019d	AGCCTGATCGGAATCTCCTCGCC	74.2	This work
Rv2019r	TCGCCGGCGGTTAACG	69.5	This work
Rv2078d	ATCGAGACCACGCAGAGGGC	71.1	This work
Rv2078r	AGTCCTCATGATGCGCACG	67.6	This work
Rv2088d	TGACCTCGATTTTCGGGC	66.2	This work
Rv2088r	TCGTCGATTGTCAACCCG	66	This work
Rv2102fw	CGTGGCCGCAGCTATGC	69.3	This work
Rv2103fw	TGATCCCAGCGCACCCCT	71.6	This work
Rv2106d	GTCCACATTTTGTACACGAAATACC	64	This work
Rv2106r	TCCTCCTCGAAACTCAAAGAA	62.7	This work
PE22d	ACATGGCGGTAAGCAATGC	66	This work
PE22r	ATCGCGGTCTTATGGGC	63.6	This work
Rv2166cext	CCGATCACCACACAGTCC	62.9	This work
Rv2169cext	TCTCTACGCCGAAGATCCCAAG	68.5	This work
Rv2179cd	AATACGTGATCGCCGGGAC	67	This work
Rv2179cr	TTCCGTGGATACTGCG	59.2	This work
Rv2277c	AAGTCCCTCCCAAGGCG	65.27	This work
Rv2280r	TGTCATTTCACTCATCGGAGGC	68.3	This work
Rv2282cd	TTCTGGCCATCGCTGAGGC	72	This work
Rv2282cr	GCAGGTCTCGTATCGCACCC	74	This work
Rv2286cd	ATCAGCTGGAGGACTCAGG	62	This work
Rv2286cr	AATCGCCGCATTGACGCC	71.6	This work
mezd	GCACCAAGTCTCAACCGTG	64.5	This work
mezr	ATCGTTGAACACGCAGTAATCCG	68.8	This work
Rv2336d	TCGCCGAATTCACCTCGGGC	75.5	This work
Rv2336r	CCGCCCATTACCGCTTCCAG	72.5	This work
Rv2348cd	ATGGTTCAGACGTAACGGTTG	63.7	This work
Rv2348cr	GAGAATGCGGAATCACG	60.4	This work

Table 5: Continuation

Name	Sequence 5'-3'	Tm ºC	Reference
plcCd	TTCTCGTCGCTTCTTGGC	64.4	This work
plcCr	ATTGCAGCTGCGATGTGTGG	69.7	This work
plcBd	AATTCGTCACGGTCCCC	63.5	This work
plcBr	CAGGAATGCTCGGCGTGAC	69.4	This work
plcAd	TGCAGGAAGGCAGGGC	68	This work
plcAr	TTTCAGCTGCACAGCCCG	69	This work
PPE38d	AAGTGCGGATTTTCGGTGTGG	70	This work
PPE38r	ATCAAACGAGGACGCCGAGG	70.5	This work
PPE39int1	ATGCCGGGGCGGTTCAGAAA	74.6	This work
PPE39int2	GAAGAAACCCGAAAGCCC	63.5	This work
PPE39R	AAGCCTGTGTTGCCGGCGCCG	80.4	This work
PPE40-3'	ACTCCGGGTTCTTCAACACAGGC	71.2	This work
PPE40F	CGGATGTTTTTTGGTGCG	65	This work
PPE40R	GGCGTTGGAGAAGCCCG	69.3	This work
Rv2345d	ATGCGTCTCGTTCGCCTGC	71	This work
Rv2345r	AACTTCCCGGCGCGTTCG	72.6	This work
Rv2356cf	CCGGCAGCTGGCTTGACC	71.8	This work
Rv2356crev	CAACTGCACAAACGCACTGC	68	This work
PPE40d	AAGGGTGGGATGGATAAGCG	67	This work
PPE40r	ATAGGTGGCCTCAGCAGCGG	70.6	This work
GlyS-3'	AATGCCGTCACCGTGCGGGAG	77.2	This work
Rv2390cd	AAAACTACGCAACAAGCCTGG	65.1	This work
Rv2390cr	ATTGCGCGATGGCGTCCC	73.9	This work
2407-ISA1	ATCCATCGCATTGAGCAGGC	66	Otal *et al.* (4)
2407	AGCTATGACCAACCAGCCC	64	Otal *et al.* (4)
Rv2435cr	ATGATCGCTCCGACAATCGG	69.4	This work
Rv2435cd	AAACGCGATGATCGTGGGC	70	This work
Rv2478cext	GCCGATATGGGCAGGGGG	71.9	This work
Rv2481cext	GCGTAGACGCCATGAGCC	67.1	This work
Rv2664d	GTTCTACGACTACCTGAGGTG	57.9	This work
Rv2664r	CCAACTGGTGTTGCTGTCAG	64.5	This work
Rv2807int	TGACGATGACTGATCTGGTG	62.5	This work
Rv2809int	GCACCGCCTGGTCAGCC	70.9	This work
Rv2809d	TTAGAGAATCGTAATCCCGC	60.6	This work
Rv2809r	AACGGACGACTGGAAGCCC	68.7	This work
Rv2813d	CAAGCCGAGCGCGGACGC	77	This work
DR-ISA1	GGGACGGAAACCTTGAATAACG	67	Otal *et al.* (4)
DR-ISA3	CCTGTATTTCGCTGGTTTCCGTC	69	Otal *et al.* (4)
Rv2818cd	TTTGTCACCGACCAAGACGAC	67.3	This work
Rv2818cr	TTGCGCACCTGTTAGCCCAG	70.5	This work
Rv2957d	TTTCCGCTGGGGTGG	64	This work
Rv2957r	AAGATCACGCCGCCCAAG	69	This work
Rv2986cd	CCCCCTGAGGCGTAGAACCG	71.9	This work
Rv2986cr	TCGAGCAAGTAGGCACACCC	67.6	This work
esxQrev	GTGCTGACATTTCCACCAC	61.7	This work
PPE46d	ATGTGGTCGGCGAGGAGC	69	This work
PPE46r	TTCCTTAGGCGCCTTCCC	66.3	This work
PPE46rev	CCTTTCAAATATTCCGGGCA	65.5	This work

Table 5: Continuation

Name	Sequence 5'-3'	Tm ºC	Reference
esxRd	TTTCAGGGTGCGCATGCC	70.7	This work
esxRr	TGTTGGACTCATGGGTGCC	67.2	This work
PPE47rev	CCTCTGTAAATGCTTGGGC	62.2	This work
Rv3113d	CAATTCATCGCGCCGCTGT	71	This work
Rv3113r	AAAATAGGTGTTGCCACCCG	65.6	This work
Rv3124d	GTGCATGAAACCCGGAG	62.9	This work
Rv3124r	CTCCAATGAGAAGGCGT	58.9	This work
3126-ISA1	GGTCGATGAGCTGAGCGC	67.5	Otal *et al.* (4)
3126-ISA3	CCAACGCGGCCTCAAACG	72.2	Otal *et al.* (4)
Rv3128cd	GACATGGTGACCGTCGTGAGGC	73.3	This work
Rv3128cr	GAGAAAGGGCGTAATGCCTC	64.8	This work
Rv3179d	AAGAACCGCCTGCTCGAATCTGC	73.2	This work
Rv3179r	TTGTGGCTGTAGTCGACCTTCCAGG	72.7	This work
Rv3180cd	TAGTCGCCGTTTGGGACCG	70.3	This work
Rv3180cr	AACGATGGGCGGCTCTAGCG	72.4	This work
Rv3186d	AACCAGTCGCAATCTCTCCCA	68	This work
Rv3186r	AATAGAGGCTGATCTGGACG	60.8	This work
Rv3190cd	ACTACCAGTCCGCACTTCG	63	This work
Rv3190cr	ACGAACACATGCCCGAC	63.5	This work
moaxd	ATCGGGTCATTACCGGCGGC	73.8	This work
moaxr	CCAGTCGACGCGGTTGGGG	75.4	This work
Rv3324Ad	TTCAGTGTCCATAACAAGCGG	64.9	This work
Rv3324Ar	TAAACCGTGAGCGCTGTCACC	69.4	This work
3326F	TGAGCGCTGTCACCGTAACG	69.6	This work
3326R	ACTTGTGCGCATCGGTTCC	68.5	This work
Rv3328cd	AAGTTCACCGGCTCTCCGCTG	71.8	This work
Rv3327-SigJ	CGCACAAGTACAGCCACACC	66.8	This work
SigJd	TTTCCGAATTCGAGGCAC	63.8	This work
dsx2ext	GATAGGTTTGGTGCCCGGTG	72	This work
lytBext	TACTTGATTGACGGGCCTG	68	This work
lytB1d	GAACCTGCCCGAGGGTACCCAG	73.5	This work
lytB1r	AAGGCCGAATCGCACCGTCTC	73.5	This work
IdsBd	TTTGAGGATTGTTTATTGGAGGG	64.6	This work
IdsBr	AACGCAAGACCCAACTCATGGC	71.2	This work
Rv3427cd	TATCGCACCATCAAGGGC	64.7	This work
Rv3427cr	AATTCCAGATGCCCCAGG	64.8	This work
Rv3346cr	CCCCAACGGGTCAATGGTGG	73.5	This work
Rv3347cd	CCAGCACGCTGCCGCTTGGG	79	This work
Rv3347cr	AAGCGGCAGCGTGCTGGTG	74	This work
Rv3347cint	ATCCTGATCCATACCGGTCCC	67.9	This work
PPE56d	GAATGGCATCAGTGCTCCC	65.8	This work
PPE56r	GAATTGGGAGATCGTGGGC	66.7	This work
rmlB2-ISA1	AACACATGCAGCCACCATGAGAG	70.4	Otal *et al.* (4)
rmlB2-ISA3	CTGCGCGGCGTTGAGGTG	74	Otal *et al.* (4)
Mb2047cd	TGTTCTCAAGTGTCCCCAA	61.8	This work
Mb2047cr	CCTAAAAGGATAGCGTTGAA	58.4	This work
Mb3109d	TGATGGGGGTCTCTCAGG	63.9	This work

Table 6: Primers used in the construction of integrative plasmids and replicative plasmids (Chapter 2)

Name	Sequence 5'-3'	Tm ºC	Reference
Rv2179c-Eco	TCGAGAACCCGAATTCACCATAGC	70.6	This work
Rv2179c-Kpn	GTGTCGTAGAGGTACCGCACCG	69.7	This work
ISA4-Eco	CTCGAATTCTCGACTGGTTCAACCATCG	74.9	Previous work
dnaN-KpnI	GAGCACCGGTACCGCGGG	76	This work
dnaN-EcoRI	ACCGCCGAATTCTTCGACACTACC	76	This work
BCG2B-Eco	CTCGAATTCCCATGTTCAAACCGGTGTC	76	Previous work
PhoBO3.1-Kpn	CTCGGTACCAACTTGAGGCTCACCGACAG	77	Previous work
Hygfw	CTGCGGAACGACCAGGAATTC	69.9	This work
Hygrv	GACCTCGGAATGGGGGG	67.9	This work

References

1. **Constant, P., E. Perez, W. Malaga, M. A. Laneelle, O. Saurel, M. Daffe, and C. Guilhot.** 2002. Role of the pks15/1 gene in the biosynthesis of phenolglycolipids in the Mycobacterium tuberculosis complex. Evidence that all strains synthesize glycosylated p-hydroxybenzoic methyl esters and that strains devoid of phenolglycolipids harbor a frameshift mutation in the pks15/1 gene. J Biol Chem **277:**38148-58.
2. **Hanekom, M., G. D. van der Spuy, E. Streicher, S. L. Ndabambi, C. R. McEvoy, M. Kidd, N. Beyers, T. C. Victor, P. D. van Helden, and R. M. Warren.** 2007. A recently evolved sublineage of the Mycobacterium tuberculosis Beijing strain family is associated with an increased ability to spread and cause disease. J Clin Microbiol **45:**1483-90.
3. **Mendiola, M. V., C. Martin, I. Otal, and B. Gicquel.** 1992. Analysis of the regions responsible for IS6110 RFLP in a single Mycobacterium tuberculosis strain. Res Microbiol **143:**767-72.
4. **Otal, I., A. B. Gomez, K. Kremer, P. de Haas, M. J. Garcia, C. Martin, and D. van Soolingen.** 2008. Mapping of IS6110 insertion sites in Mycobacterium bovis isolates in relation to adaptation from the animal to human host. Vet Microbiol **129:**333-41.
5. **Prod'hom, G., B. Lagier, V. Pelicic, A. J. Hance, B. Gicquel, and C. Guilhot.** 1998. A reliable amplification technique for the characterization of genomic DNA sequences flanking insertion sequences. FEMS Microbiol Lett **158:**75-81.
6. **Samper, S., I. Otal, M. C. Rubio, M. A. Vitoria, R. Gomez-Lus, and C. Martin.** 1993. [Application of RFLP to typing strains of Mycobacterium tuberculosis]. Enferm Infecc Microbiol Clin **11:**547-51.
7. **Soto, C. Y., M. C. Menendez, E. Perez, S. Samper, A. B. Gomez, M. J. Garcia, and C. Martin.** 2004. IS6110 mediates increased transcription of the phoP virulence gene in a multidrug-resistant clinical isolate responsible for tuberculosis outbreaks. J Clin Microbiol **42:**212-9.
8. **Tsolaki, A. G., S. Gagneux, A. S. Pym, Y. O. Goguet de la Salmoniere, B. N. Kreiswirth, D. Van Soolingen, and P. M. Small.** 2005. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of Mycobacterium tuberculosis. J Clin Microbiol **43:**3185-91.

Anex II

Table 1. Obtained Points of insertion of IS*6110* of the 61 studied strains.

Name of the strain	1594	483251	483298	483310	483537	850087	850182	850539	850592
[a] **H37Rv**	0 [b]	0	0	0	0	0	0	0	0
HMS 1301	1 [c]	0	0	0	0	0	0	0	0
HMS 1356	1	0	0	0	0	0	0	0	0
HMS 1470	1	0	0	0	0	0	0	0	0
HMS 1480	1	0	0	0	0	0	0	0	0
HMS 1834	0	0	0	0	0	0	0	0	0
HMS 2103	0	0	0	0	0	0	0	0	0
HMS 2193	0	0	0	0	0	0	0	0	0
HMS 2200	0	0	0	0	1	0	0	1	0
HMS 2206	0	0	0	0	0	0	0	0	0
HMS 2208	0	0	0	0	0	0	0	0	0
HMS 2224	0	0	1	0	0	0	0	0	0
HMS 2227	0	0	0	0	0	0	0	0	0
HMS 2237	0	0	0	0	0	0	0	0	0
HMS 2241	0	0	0	0	0	0	0	0	0
HMS 2258	0	0	0	0	0	0	0	0	0
HMS 2268	0	0	0	0	0	0	0	0	0
HMS 2271	0	1	0	0	0	0	0	0	0
HMS 2280	0	0	0	0	0	0	0	0	0
HMS 2281	0	0	0	0	0	0	0	0	0
HMS 2296	0	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	1	0	0
HMS 2355	0	0	0	0	0	0	0	0	0
HMS 2370	0	0	0	0	0	0	0	0	0
HMS 2391	0	0	0	0	0	0	0	0	0
HMS 2394	0	0	1	0	0	0	0	0	0
HMS 2400	1	0	0	0	0	0	0	0	0
HMS 2405	0	0	0	0	0	0	0	0	1
HMS 2414	0	0	1	0	0	0	0	0	0
HMS 2513	0	0	1	0	0	0	0	0	0
HMS 2536	0	0	0	0	0	0	0	1	0
HMS 2580	1	0	0	0	0	0	0	0	0
HMS 2584	0	0	0	0	0	0	0	0	0
HCU 2398	1	0	0	0	0	0	0	0	0
HCU 2613	0	0	0	0	0	0	0	0	0
HCU 2617	1	0	0	0	0	0	0	0	0
HCU 2633	0	0	0	0	0	0	0	0	0
HCU 2647	0	0	0	0	0	0	0	0	0
HCU 2654	0	0	0	0	0	0	0	0	0
HCU 2663	0	0	0	0	0	0	0	0	0
HCU 2671	0	0	0	0	0	0	0	1	0
HCU 2703	0	0	0	0	0	0	0	0	0
HCU 2706	0	0	0	0	0	0	0	0	0
HCU 2728	0	0	0	0	0	0	0	0	0
HCU 2777	0	0	0	0	0	0	0	0	0
HCU 2821	0	0	1	0	0	0	0	0	0
HCU 2864	0	0	1	0	0	0	0	0	0
HCU 2886	0	0	0	0	0	0	0	0	0
HCU 2875	0	0	0	0	0	0	0	0	0
HCU 2942	0	0	0	0	0	0	0	0	0
HCU 3151	0	0	0	0	0	1	0	0	0
HSJ 025	0	0	0	0	0	1	0	0	0
HSJ 029	0	0	0	1	0	0	0	0	0
HSJ 067	1	0	0	0	0	0	0	0	0
NHN5	1	0	0	0	0	0	0	0	0
HM77	1	0	0	0	0	0	0	0	0
HM764	1	0	0	0	0	0	0	0	0
HM903	1	0	0	0	0	0	0	0	0
N4	1	0	0	0	0	0	0	0	0
W4	1	0	0	0	0	0	0	0	0
990172	1	0	0	0	0	0	0	0	0
CAM22	1	0	0	0	0	0	0	0	0

[a] The points of insertion of IS6110 of H37Rv strain were included
[b] The absence of a point of insertion of IS*6110* in a strain was indicated by the number 0.
[c] The presence of a point of insertion of IS*6110* in a strain was indicated by the number 1.

Name of the strain	850601	851412	851479	888768	888789	888846	889072	932076	932203
[a] **H37Rv**	0	0	0	0	0	0	1	0	0
HMS 1301	0	0	0	0	0	0	1	0	0
HMS 1356	0	0	0	0	0	0	0	0	0
HMS 1470	0	0	0	0	0	0	0	1	0
HMS 1480	0	0	0	0	0	0	0	1	0
HMS 1834	0	0	0	0	0	0	0	0	1
HMS 2103	0	0	0	0	1	0	0	0	0
HMS 2193	0	0	0	0	0	0	0	0	1
HMS 2200	0	0	0	0	1	0	0	0	0
HMS 2206	0	0	0	0	0	0	0	0	1
HMS 2208	0	0	0	0	0	0	0	0	1
HMS 2224	0	0	0	0	0	0	0	0	0
HMS 2227	0	0	0	0	0	0	0	0	0
HMS 2237	0	0	0	0	0	0	0	0	0
HMS 2241	0	0	0	0	0	0	0	0	0
HMS 2258	0	0	0	1	0	0	0	0	0
HMS 2268	0	0	0	0	0	0	0	0	1
HMS 2271	0	0	0	0	0	0	0	0	1
HMS 2280	0	0	0	0	0	0	0	0	1
HMS 2281	0	0	0	0	0	1	0	0	1
HMS 2296	0	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	0	0	0
HMS 2355	0	0	0	0	0	0	0	0	0
HMS 2370	0	0	0	0	0	0	0	0	0
HMS 2391	0	0	0	0	0	0	0	0	1
HMS 2394	0	0	0	0	0	0	0	0	0
HMS 2400	1	0	0	0	0	0	0	0	0
HMS 2405	0	1	0	0	0	0	0	0	0
HMS 2414	0	0	0	0	0	0	0	0	0
HMS 2513	0	0	0	0	0	0	0	0	0
HMS 2536	0	0	0	0	0	0	0	0	0
HMS 2580	0	0	0	0	0	0	0	0	0
HMS 2584	0	0	0	0	0	0	0	0	0
HCU 2398	0	0	0	0	0	0	0	0	0
HCU 2613	0	0	0	0	0	0	0	0	1
HCU 2617	0	0	0	0	0	0	0	0	0
HCU 2633	0	0	0	0	0	0	0	0	1
HCU 2647	0	0	0	0	0	0	0	0	1
HCU 2654	0	0	0	0	0	0	0	0	1
HCU 2663	0	0	0	0	0	0	0	0	1
HCU 2671	0	0	0	0	0	0	0	0	0
HCU 2703	0	0	0	0	1	0	0	0	1
HCU 2706	0	0	0	0	0	0	0	0	0
HCU 2728	0	0	0	0	0	0	0	0	0
HCU 2777	0	0	0	0	0	0	0	0	0
HCU 2821	0	0	0	0	0	0	0	0	0
HCU 2864	0	0	0	0	0	0	0	0	0
HCU 2886	0	0	0	0	0	0	0	0	1
HCU 2875	0	0	0	0	0	0	0	0	0
HCU 2942	0	0	0	0	0	0	0	0	0
HCU 3151	0	0	0	0	0	0	0	0	0
HSJ 025	0	0	0	0	0	0	0	0	0
HSJ 029	0	0	0	0	0	0	0	0	0
HSJ 067	0	0	1	0	0	0	0	0	0
NHN5	0	0	0	0	0	0	0	0	0
HM77	0	0	0	0	0	0	1	0	0
HM764	0	0	0	0	0	0	1	0	0
HM903	0	0	0	0	0	0	0	0	0
N4	0	0	0	0	0	0	0	0	0
W4	0	0	0	0	0	0	1	0	0
990172	0	0	0	0	0	0	0	0	0
CAM22	0	0	0	0	0	0	1	0	0

Name of the strain	937116	1168966	1262963	1356403	1357082	1481531	1530161	1530203	1530907
[a] **H37Rv**	0	0	0	0	0	0	0	0	0
HMS 1301	1	0	0	0	0	0	0	0	0
HMS 1356	0	0	1	0	0	0	0	0	0
HMS 1470	0	0	1	0	0	0	0	0	0
HMS 1480	0	0	1	0	0	0	0	0	0
HMS 1834	0	0	0	0	0	0	0	0	0
HMS 2103	0	0	0	0	0	0	0	0	0
HMS 2193	0	0	0	0	0	0	0	0	0
HMS 2200	0	0	0	0	0	0	0	0	0
HMS 2206	0	0	0	0	0	0	0	0	0
HMS 2208	0	0	0	0	0	0	0	0	0
HMS 2224	0	0	0	0	0	0	0	0	0
HMS 2227	0	0	0	0	0	0	0	0	0
HMS 2237	0	0	0	0	0	0	0	0	0
HMS 2241	0	0	0	0	0	0	0	0	0
HMS 2258	0	0	0	0	0	0	0	0	0
HMS 2268	0	0	0	0	0	0	0	0	0
HMS 2271	0	0	0	1	0	0	0	0	0
HMS 2280	0	0	0	0	0	0	0	0	0
HMS 2281	0	0	0	0	0	0	0	0	0
HMS 2296	0	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	0	0	0
HMS 2355	0	0	0	0	0	0	0	0	0
HMS 2370	0	0	0	0	0	0	0	0	0
HMS 2391	0	0	0	0	0	0	0	0	0
HMS 2394	0	0	0	0	0	0	0	0	0
HMS 2400	0	0	0	0	0	0	0	0	0
HMS 2405	0	0	0	0	0	0	0	0	0
HMS 2414	0	0	0	0	0	0	0	0	0
HMS 2513	0	0	0	0	0	0	0	0	0
HMS 2536	0	0	0	0	0	0	0	0	0
HMS 2580	0	0	1	0	0	0	0	0	0
HMS 2584	0	0	0	0	0	0	0	0	0
HCU 2398	0	0	1	0	0	0	0	0	0
HCU 2613	0	0	0	0	0	0	0	0	0
HCU 2617	0	0	1	0	0	0	0	0	0
HCU 2633	0	0	0	0	0	0	0	0	0
HCU 2647	0	0	0	0	0	0	0	0	1
HCU 2654	0	0	0	0	0	1	0	0	0
HCU 2663	0	1	0	0	0	0	0	0	0
HCU 2671	0	0	0	0	0	0	0	0	0
HCU 2703	0	0	0	0	0	1	0	0	0
HCU 2706	0	0	0	0	0	0	0	0	0
HCU 2728	0	0	0	0	1	0	0	0	0
HCU 2777	0	0	0	0	0	0	0	0	0
HCU 2821	0	0	0	0	0	0	0	0	0
HCU 2864	0	0	0	0	0	0	0	0	0
HCU 2886	0	0	0	0	0	0	0	0	0
HCU 2875	0	0	0	0	0	0	0	0	0
HCU 2942	0	0	0	0	0	0	0	0	0
HCU 3151	0	0	0	0	0	0	0	0	0
HSJ 025	0	0	0	0	0	0	0	0	0
HSJ 029	0	0	0	0	0	0	0	0	0
HSJ 067	0	0	1	0	0	0	0	0	0
NHN5	0	0	0	0	0	0	1	0	0
HM77	0	0	1	0	0	0	0	0	0
HM764	0	0	1	0	0	0	0	0	0
HM903	0	0	0	0	0	0	0	0	0
N4	0	0	0	0	0	0	0	1	0
W4	0	0	1	0	0	0	0	0	0
990172	0	0	0	0	0	0	0	0	0
CAM22	0	0	1	0	0	0	0	0	0

Name of the strain	1531507	1541994	1543972	1657016	1724406	1729409	1745059	1889068	1889623 - 1889816
[a] **H37Rv**	0	1	0	0	0	0	0	0	0
HMS 1301	0	0	1	1	0	0	0	0	0
HMS 1356	0	0	1	1	0	0	0	0	0
HMS 1470	0	0	1	1	0	0	0	0	0
HMS 1480	0	0	1	1	0	0	0	0	0
HMS 1834	0	0	0	0	0	0	0	0	0
HMS 2103	0	0	0	0	0	0	0	0	0
HMS 2193	0	0	0	0	0	0	0	0	0
HMS 2200	0	0	0	0	0	0	0	0	0
HMS 2206	0	0	0	0	0	0	0	0	0
HMS 2208	0	0	0	0	0	0	0	0	0
HMS 2224	0	0	0	0	0	0	0	0	0
HMS 2227	0	0	0	0	0	0	0	0	0
HMS 2237	0	0	0	0	0	0	0	0	0
HMS 2241	0	0	0	0	0	0	0	0	0
HMS 2258	0	0	0	0	0	0	0	0	0
HMS 2268	0	0	0	0	0	0	0	0	0
HMS 2271	0	0	0	0	0	0	0	0	0
HMS 2280	0	0	0	0	0	0	0	0	0
HMS 2281	0	0	0	0	0	0	0	0	0
HMS 2296	0	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	0	0	0
HMS 2355	0	0	0	0	0	0	0	0	0
HMS 2370	0	0	0	0	0	0	0	0	0
HMS 2391	0	0	0	0	0	0	0	0	0
HMS 2394	0	0	0	0	0	0	0	0	0
HMS 2400	0	0	1	0	0	0	0	0	0
HMS 2405	0	0	0	0	0	0	0	0	0
HMS 2414	0	0	0	0	0	0	0	0	0
HMS 2513	0	0	0	0	0	0	0	0	0
HMS 2536	0	0	0	0	0	0	0	0	0
HMS 2580	0	0	1	1	0	0	0	0	0
HMS 2584	0	0	0	0	0	0	0	1	0
HCU 2398	0	0	1	1	0	0	0	0	0
HCU 2613	0	0	0	0	0	0	0	0	0
HCU 2617	0	0	1	1	0	0	0	0	0
HCU 2633	0	0	0	0	0	0	0	0	0
HCU 2647	0	0	0	0	0	0	0	0	0
HCU 2654	0	0	0	0	0	0	0	0	0
HCU 2663	0	0	0	0	0	0	0	0	0
HCU 2671	0	0	0	0	0	0	0	0	0
HCU 2703	0	0	0	0	0	0	0	0	0
HCU 2706	0	0	0	0	0	0	0	0	0
HCU 2728	0	0	0	0	0	0	0	0	0
HCU 2777	0	0	0	0	0	0	0	0	0
HCU 2821	0	0	0	0	0	0	0	0	0
HCU 2864	0	0	0	0	0	0	0	0	0
HCU 2886	0	0	0	0	0	0	0	0	0
HCU 2875	1	0	0	0	0	0	0	0	0
HCU 2942	0	0	0	0	0	0	0	1	0
HCU 3151	0	0	0	0	0	0	0	0	0
HSJ 025	0	0	0	0	0	0	0	0	1
HSJ 029	0	0	0	0	0	0	0	0	0
HSJ 067	0	0	1	1	0	0	0	0	0
NHN5	0	0	1	0	1	0	0	0	0
HM77	0	0	1	1	0	0	0	0	0
HM764	0	0	1	1	0	0	0	0	0
HM903	0	0	1	1	0	0	0	0	0
N4	0	0	1	0	1	1	0	0	0
W4	0	0	1	1	0	0	1	0	0
990172	0	0	0	0	0	0	0	0	0
CAM22	0	0	1	1	0	0	0	0	0

Name of the strain	1895860	1986625	1986626	1986726	1986820	1986638-1998625	1987008	1987018
[a] **H37Rv**	0	0	0	0	0	0	0	0
HMS 1301	0	0	0	0	0	1	0	0
HMS 1356	0	0	0	0	0	1	0	0
HMS 1470	0	0	0	0	0	1	0	0
HMS 1480	0	0	0	0	0	1	0	0
HMS 1834	0	1	0	0	0	0	0	1
HMS 2103	0	0	0	0	0	0	0	0
HMS 2193	0	0	0	0	0	0	0	0
HMS 2200	0	0	0	1	0	0	0	0
HMS 2206	0	0	0	0	1	0	0	0
HMS 2208	0	1	0	0	0	0	0	0
HMS 2224	0	0	1	0	0	0	0	0
HMS 2227	0	0	0	0	0	0	0	0
HMS 2237	0	0	1	0	0	0	0	0
HMS 2241	0	0	0	0	0	0	0	0
HMS 2258	0	0	1	0	0	0	0	0
HMS 2268	0	0	0	0	0	0	0	0
HMS 2271	0	0	0	0	0	0	0	0
HMS 2280	0	0	0	0	0	0	0	0
HMS 2281	0	1	0	0	0	0	0	0
HMS 2296	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	0	0
HMS 2355	0	0	0	0	0	0	0	0
HMS 2370	0	0	0	0	0	0	1	0
HMS 2391	0	1	0	0	0	0	0	0
HMS 2394	0	0	1	0	0	0	0	0
HMS 2400	0	0	0	0	0	1	0	0
HMS 2405	0	0	0	0	0	0	0	0
HMS 2414	0	0	1	0	0	0	0	0
HMS 2513	0	0	1	0	0	0	0	0
HMS 2536	0	0	0	0	0	0	0	0
HMS 2580	0	0	0	0	0	1	0	0
HMS 2584	0	0	0	0	0	0	0	0
HCU 2398	0	0	0	0	0	1	0	0
HCU 2613	0	0	0	0	0	0	0	0
HCU 2617	0	0	0	0	0	1	0	0
HCU 2633	0	0	0	0	0	0	0	0
HCU 2647	0	0	0	0	0	0	0	0
HCU 2654	0	0	0	0	0	0	0	0
HCU 2663	1	0	0	0	0	0	0	0
HCU 2671	0	0	0	0	0	0	0	0
HCU 2703	0	1	0	0	0	0	0	0
HCU 2706	0	0	0	0	0	0	1	0
HCU 2728	0	0	0	0	0	0	0	0
HCU 2777	0	0	0	0	0	0	0	0
HCU 2821	0	0	0	0	0	0	0	0
HCU 2864	0	0	1	0	0	0	0	0
HCU 2886	0	0	0	0	0	0	0	0
HCU 2875	0	0	0	0	0	0	0	0
HCU 2942	0	0	0	0	0	0	0	0
HCU 3151	0	0	0	0	0	0	0	0
HSJ 025	0	0	0	0	0	0	0	0
HSJ 029	0	0	0	0	0	0	0	0
HSJ 067	0	0	0	0	0	1	0	0
NHN5	0	0	0	0	0	1	0	0
HM77	0	0	0	0	0	1	0	0
HM764	0	0	0	0	0	1	0	0
HM903	0	0	0	0	0	1	0	0
N4	0	0	0	0	0	1	0	0
W4	0	0	0	0	0	1	0	0
990172	0	0	0	0	0	1	0	0
CAM22	0	0	0	0	0	1	0	0

Name of the strain	1987129-1987556	1987211	1987458-1987544	1987556	1987601	1987745	1996152	1998278
[a] **H37Rv**	0	0	0	0	0	1	1	0
HMS 1301	0	0	0	0	0	0	0	0
HMS 1356	0	0	0	0	0	0	0	0
HMS 1470	0	0	0	0	0	0	0	0
HMS 1480	0	0	0	0	0	0	0	0
HMS 1834	0	0	0	0	0	0	0	0
HMS 2103	1	0	0	1	0	0	0	0
HMS 2193	0	0	0	0	0	0	0	0
HMS 2200	0	0	0	0	0	0	0	0
HMS 2206	0	1	0	0	0	0	0	0
HMS 2208	0	0	0	0	0	0	0	0
HMS 2224	0	0	0	0	0	0	0	0
HMS 2227	0	0	0	0	0	0	0	0
HMS 2237	0	0	0	0	0	0	0	0
HMS 2241	0	0	0	0	0	0	0	0
HMS 2258	0	0	0	0	0	0	0	0
HMS 2268	0	0	0	0	0	0	0	0
HMS 2271	0	0	0	0	0	0	0	0
HMS 2280	0	0	0	0	0	0	0	0
HMS 2281	0	0	0	0	0	0	0	0
HMS 2296	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	0	0
HMS 2355	0	0	0	0	0	0	0	0
HMS 2370	0	0	0	0	0	0	0	0
HMS 2391	0	0	0	0	0	0	0	0
HMS 2394	0	0	0	0	0	0	0	0
HMS 2400	0	0	0	0	0	0	0	0
HMS 2405	0	0	0	0	0	0	0	0
HMS 2414	0	0	0	0	1	0	0	0
HMS 2513	0	0	0	0	0	0	0	0
HMS 2536	0	0	0	0	0	0	0	0
HMS 2580	0	0	0	0	0	0	0	0
HMS 2584	0	0	0	0	0	0	0	0
HCU 2398	0	0	0	0	0	0	0	0
HCU 2613	0	0	0	0	0	0	0	0
HCU 2617	0	0	0	0	0	0	0	0
HCU 2633	0	0	0	0	0	0	0	0
HCU 2647	0	0	0	0	0	0	0	0
HCU 2654	0	0	0	0	0	0	0	0
HCU 2663	0	0	0	0	0	0	0	0
HCU 2671	0	0	0	0	0	0	0	0
HCU 2703	0	0	0	0	0	0	0	0
HCU 2706	0	0	0	0	0	0	0	0
HCU 2728	0	0	0	0	0	0	0	1
HCU 2777	0	0	0	0	0	0	0	0
HCU 2821	0	0	0	0	0	0	0	0
HCU 2864	0	0	0	0	0	0	0	0
HCU 2886	0	0	0	0	0	0	0	0
HCU 2875	0	0	0	0	0	0	0	0
HCU 2942	0	0	0	0	0	0	0	0
HCU 3151	0	0	1	0	0	0	0	0
HSJ 025	0	0	1	0	0	0	0	0
HSJ 029	0	0	0	0	0	0	0	0
HSJ 067	0	0	0	0	0	0	0	0
NHN5	0	0	0	0	0	0	0	0
HM77	0	0	0	0	0	0	0	0
HM764	0	0	0	0	0	0	0	0
HM903	0	0	0	0	0	0	0	0
N4	0	0	0	0	0	0	0	0
W4	0	0	0	0	0	0	0	0
990172	0	0	0	0	0	0	0	0
CAM22	0	0	0	0	0	0	0	0

Name of the strain	1998623	1998698	1998792	1998800	1998849	2010924	2011104	2011114	2039012
[a] **H37Rv**	0	0	0	0	0	0	0	0	0
HMS 1301	0	0	0	0	0	0	0	0	0
HMS 1356	0	0	0	0	0	0	0	0	0
HMS 1470	0	0	0	0	0	0	0	0	0
HMS 1480	0	0	0	0	0	0	0	0	0
HMS 1834	0	0	0	0	0	1	0	0	0
HMS 2103	0	0	0	0	0	0	0	0	0
HMS 2193	0	0	0	0	0	0	0	0	0
HMS 2200	0	0	0	0	0	0	0	0	0
HMS 2206	0	0	0	0	0	0	0	0	0
HMS 2208	0	0	0	0	0	1	0	0	0
HMS 2224	0	0	0	0	0	0	0	0	0
HMS 2227	0	0	0	0	0	0	0	0	0
HMS 2237	0	0	1	0	0	0	0	0	0
HMS 2241	0	0	0	1	0	0	0	0	0
HMS 2258	0	0	0	0	0	0	0	0	0
HMS 2268	0	0	0	0	0	0	0	0	0
HMS 2271	0	0	0	0	0	0	0	0	0
HMS 2280	0	0	0	0	0	0	0	0	0
HMS 2281	0	0	0	0	0	0	0	0	0
HMS 2296	0	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	0	0	0
HMS 2355	0	0	0	0	0	0	0	0	0
HMS 2370	0	0	0	0	0	0	0	0	0
HMS 2391	0	0	0	0	0	1	1	0	0
HMS 2394	0	0	0	0	0	0	0	0	0
HMS 2400	1	0	0	0	0	0	0	0	0
HMS 2405	0	0	1	0	0	0	0	0	0
HMS 2414	0	0	0	0	0	0	0	0	0
HMS 2513	0	0	0	0	0	0	0	0	0
HMS 2536	0	0	0	0	0	0	0	0	0
HMS 2580	1	0	0	0	0	0	0	0	0
HMS 2584	0	0	0	0	1	0	0	0	0
HCU 2398	0	0	0	0	0	0	0	0	0
HCU 2613	0	0	0	0	0	0	0	0	0
HCU 2617	0	0	0	0	0	0	0	0	0
HCU 2633	0	0	0	0	0	0	0	0	0
HCU 2647	0	0	0	0	0	0	0	0	0
HCU 2654	0	0	0	0	0	1	0	1	0
HCU 2663	0	0	0	0	0	0	0	0	0
HCU 2671	0	0	0	0	0	0	0	0	0
HCU 2703	0	0	0	0	0	1	0	0	0
HCU 2706	0	0	0	0	0	0	0	0	0
HCU 2728	0	0	0	0	0	0	0	0	0
HCU 2777	0	0	0	0	0	0	0	0	1
HCU 2821	0	0	0	0	0	0	0	0	0
HCU 2864	0	1	0	0	0	0	0	0	0
HCU 2886	0	0	0	0	0	0	0	0	0
HCU 2875	0	0	0	0	0	0	0	0	0
HCU 2942	0	0	0	0	0	0	0	0	0
HCU 3151	0	0	0	0	0	0	0	0	0
HSJ 025	0	0	0	0	0	0	0	0	0
HSJ 029	0	0	0	0	0	0	0	0	0
HSJ 067	0	0	0	0	0	0	0	0	0
NHN5	0	0	0	0	0	0	0	0	0
HM77	0	0	0	0	0	0	0	0	0
HM764	0	0	0	0	0	0	0	0	0
HM903	0	0	0	0	0	0	0	0	0
N4	0	0	0	0	0	0	0	0	0
W4	0	0	0	0	0	0	0	0	0
990172	0	0	0	0	0	0	0	0	0
CAM22	0	0	0	0	0	0	0	0	0

Name of the strain	2041428	2041741	2041820	2057398	2163392	2163404	2163461	2163473
[a] **H37Rv**	0	0	0	0	0	0	0	0
HMS 1301	0	0	0	0	2	0	0	0
HMS 1356	0	0	0	0	0	0	0	0
HMS 1470	0	0	0	0	0	0	0	0
HMS 1480	0	0	0	0	0	0	0	0
HMS 1834	0	0	0	0	0	1	0	0
HMS 2103	0	0	0	1	0	0	0	0
HMS 2193	0	0	0	0	0	0	1	0
HMS 2200	0	0	0	0	0	0	0	0
HMS 2206	0	0	0	0	0	0	0	0
HMS 2208	0	0	0	0	0	0	0	0
HMS 2224	0	0	0	0	0	0	0	0
HMS 2227	0	0	0	0	0	0	0	0
HMS 2237	0	0	0	0	0	0	0	0
HMS 2241	0	0	0	0	0	0	0	0
HMS 2258	0	0	0	0	0	0	0	0
HMS 2268	0	0	0	0	0	0	0	0
HMS 2271	1	0	0	0	0	0	0	0
HMS 2280	0	0	0	0	0	0	0	0
HMS 2281	0	0	0	0	0	0	0	0
HMS 2296	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	0	0
HMS 2355	0	0	0	0	0	0	0	0
HMS 2370	0	0	0	0	0	0	0	0
HMS 2391	0	0	0	0	0	0	0	0
HMS 2394	0	0	1	0	0	0	0	0
HMS 2400	0	0	0	0	0	0	0	0
HMS 2405	0	0	0	0	0	0	0	0
HMS 2414	0	0	0	0	0	0	0	0
HMS 2513	0	0	0	0	0	0	0	0
HMS 2536	0	0	0	0	0	0	0	0
HMS 2580	0	0	0	0	0	0	0	0
HMS 2584	0	0	0	0	0	0	0	0
HCU 2398	0	0	0	0	0	0	0	0
HCU 2613	0	0	0	0	0	0	0	0
HCU 2617	0	0	0	0	0	0	0	0
HCU 2633	0	0	0	0	0	0	0	0
HCU 2647	0	0	0	0	0	0	0	0
HCU 2654	0	0	0	0	0	0	0	0
HCU 2663	0	0	0	0	0	0	0	0
HCU 2671	0	0	0	0	0	0	0	0
HCU 2703	0	0	0	0	0	0	0	0
HCU 2706	0	0	0	0	0	0	0	0
HCU 2728	0	0	0	0	0	0	0	0
HCU 2777	0	0	0	0	0	0	0	0
HCU 2821	0	0	0	0	0	0	0	0
HCU 2864	0	0	0	0	0	0	0	1
HCU 2886	0	0	0	0	0	0	0	0
HCU 2875	0	0	0	0	0	0	0	0
HCU 2942	0	0	0	0	0	0	0	0
HCU 3151	0	0	0	0	0	0	0	0
HSJ 025	0	0	0	0	0	0	0	0
HSJ 029	0	0	0	0	0	0	0	0
HSJ 067	0	0	0	0	0	0	0	0
NHN5	0	1	0	0	0	0	0	0
HM77	0	0	0	0	0	0	0	0
HM764	0	0	0	0	0	0	0	0
HM903	0	0	0	0	0	0	0	0
N4	0	1	0	0	0	0	0	0
W4	0	0	0	0	0	0	0	0
990172	0	0	0	0	0	0	0	0
CAM22	0	0	0	0	0	0	0	0

Name of the strain	2163649	2174919	2180845	2198330	2248279	2262185	2263619	2263627
[a] **H37Rv**	0	0	0	0	0	0	0	0
HMS 1301	0	0	0	0	0	0	0	1
HMS 1356	0	0	0	0	0	0	0	1
HMS 1470	0	0	0	0	0	0	0	1
HMS 1480	0	0	0	0	0	0	0	1
HMS 1834	0	0	0	0	0	0	0	0
HMS 2103	0	0	0	0	1	0	0	0
HMS 2193	0	1	0	0	0	0	0	0
HMS 2200	0	0	0	0	0	0	0	0
HMS 2206	0	0	0	0	0	0	0	0
HMS 2208	0	0	0	0	0	0	0	0
HMS 2224	0	0	0	0	0	0	0	0
HMS 2227	0	0	0	0	0	0	0	0
HMS 2237	0	0	0	0	0	0	0	0
HMS 2241	0	0	0	0	0	0	0	0
HMS 2258	0	0	0	0	0	0	0	0
HMS 2268	0	0	0	0	0	0	0	0
HMS 2271	0	0	0	0	0	0	0	0
HMS 2280	0	0	0	0	0	0	0	0
HMS 2281	0	0	0	0	0	0	0	0
HMS 2296	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	0	0
HMS 2355	0	0	0	0	0	0	0	0
HMS 2370	0	0	0	0	0	0	0	0
HMS 2391	0	0	0	0	0	1	0	0
HMS 2394	0	0	0	0	0	0	0	0
HMS 2400	0	0	0	0	0	0	0	1
HMS 2405	0	0	0	0	0	0	0	0
HMS 2414	0	0	0	0	0	0	0	0
HMS 2513	0	0	0	0	0	0	0	0
HMS 2536	0	0	0	0	0	0	0	0
HMS 2580	0	0	0	0	0	0	0	1
HMS 2584	0	0	0	0	0	0	0	0
HCU 2398	0	0	0	0	0	0	0	1
HCU 2613	0	0	0	0	0	0	0	0
HCU 2617	0	0	0	0	0	0	0	1
HCU 2633	0	0	1	0	0	0	0	0
HCU 2647	0	0	0	0	0	0	0	0
HCU 2654	0	0	0	0	0	0	0	0
HCU 2663	0	0	0	0	0	0	0	0
HCU 2671	0	0	0	0	0	0	0	0
HCU 2703	0	0	0	0	0	0	0	0
HCU 2706	0	0	0	0	0	0	0	0
HCU 2728	0	0	0	0	0	0	0	0
HCU 2777	0	0	0	1	0	0	0	0
HCU 2821	0	0	0	0	0	0	0	0
HCU 2864	0	0	0	0	0	0	0	0
HCU 2886	0	0	1	0	0	0	0	0
HCU 2875	0	0	0	0	0	0	0	0
HCU 2942	0	0	0	0	0	0	0	0
HCU 3151	0	0	0	0	0	0	0	0
HSJ 025	0	0	0	0	0	0	0	0
HSJ 029	0	0	0	0	0	0	0	0
HSJ 067	0	0	0	0	0	0	0	1
NHN5	1	0	0	0	0	0	0	1
HM77	0	0	0	0	0	0	0	1
HM764	0	0	0	0	0	0	0	1
HM903	0	0	0	0	0	0	0	1
N4	0	0	0	0	0	0	1	0
W4	0	0	0	0	0	0	0	1
990172	0	0	0	0	0	0	0	0
CAM22	0	0	0	0	0	0	0	1

Name of the strain	2263778	2266015	2266167	2268726-2268693	2334666	2334686	2364232	2364548
[a] **H37Rv**	0	0	0	0	0	0	0	0
HMS 1301	0	0	0	1	0	1	0	0
HMS 1356	0	0	0	0	0	0	0	0
HMS 1470	0	0	0	0	0	0	0	0
HMS 1480	0	0	0	0	0	0	0	0
HMS 1834	0	0	0	0	0	0	0	0
HMS 2103	0	0	0	0	0	0	0	0
HMS 2193	0	0	0	0	0	0	0	0
HMS 2200	0	0	0	0	0	0	0	0
HMS 2206	0	0	0	0	0	0	0	1
HMS 2208	0	0	0	0	0	0	0	0
HMS 2224	0	0	0	0	0	0	0	0
HMS 2227	0	0	0	0	0	0	0	0
HMS 2237	0	0	0	0	0	0	0	0
HMS 2241	0	0	0	0	0	0	0	0
HMS 2258	0	1	0	0	0	0	0	0
HMS 2268	0	0	0	0	0	0	0	0
HMS 2271	0	0	0	0	0	0	0	0
HMS 2280	0	0	0	0	0	0	0	0
HMS 2281	0	0	0	0	0	0	0	0
HMS 2296	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	0	0
HMS 2355	0	0	0	0	0	0	0	0
HMS 2370	0	0	0	0	0	0	0	0
HMS 2391	0	0	0	0	0	0	0	0
HMS 2394	0	0	0	0	0	0	0	0
HMS 2400	0	0	0	0	0	0	0	0
HMS 2405	0	0	0	0	0	0	0	0
HMS 2414	0	0	0	0	0	0	0	0
HMS 2513	0	0	0	0	0	0	0	0
HMS 2536	0	0	0	0	0	0	0	0
HMS 2580	0	0	0	0	0	0	0	0
HMS 2584	0	0	0	0	0	0	0	0
HCU 2398	0	0	0	0	0	0	0	0
HCU 2613	0	0	0	0	0	0	0	0
HCU 2617	0	0	0	0	0	0	0	0
HCU 2633	0	0	0	0	0	0	1	0
HCU 2647	0	0	0	0	0	0	0	0
HCU 2654	0	0	0	0	0	0	0	0
HCU 2663	0	0	0	0	0	0	0	0
HCU 2671	0	0	0	0	0	0	0	0
HCU 2703	0	0	0	0	0	0	0	0
HCU 2706	0	0	0	0	0	0	0	0
HCU 2728	0	0	0	0	0	0	0	0
HCU 2777	0	0	0	0	0	0	0	0
HCU 2821	0	0	0	0	0	0	0	0
HCU 2864	0	0	0	0	0	0	0	0
HCU 2886	0	0	0	0	0	0	0	0
HCU 2875	0	0	0	0	0	0	0	0
HCU 2942	0	0	0	0	0	0	0	0
HCU 3151	0	0	0	0	0	0	0	0
HSJ 025	0	0	0	0	0	0	0	0
HSJ 029	0	0	0	0	0	0	0	0
HSJ 067	0	0	0	0	0	0	0	0
NHN5	0	0	0	0	0	0	0	0
HM77	0	0	0	0	0	0	0	0
HM764	0	0	0	0	0	0	0	0
HM903	0	0	0	0	0	0	0	0
N4	0	0	0	0	0	0	0	0
W4	0	0	0	0	0	0	0	0
990172	1	0	1	0	1	0	0	0
CAM22	0	0	0	0	0	0	0	0

Name of the strain	2364912	2365243	2365414	2365465	2366894	2366896	2366897	2367158-2367279
[a] **H37Rv**	0	0	0	1	0	0	0	0
HMS 1301	0	0	0	0	0	0	0	0
HMS 1356	0	0	0	0	0	0	0	0
HMS 1470	0	0	0	0	1	0	0	0
HMS 1480	0	0	0	0	1	0	0	0
HMS 1834	0	0	0	0	0	0	0	0
HMS 2103	0	0	0	0	0	0	0	0
HMS 2193	0	0	0	0	0	0	0	0
HMS 2200	0	0	0	0	0	0	0	0
HMS 2206	0	0	0	0	0	0	0	0
HMS 2208	0	0	0	0	0	0	0	0
HMS 2224	0	0	0	0	0	0	0	0
HMS 2227	0	0	0	0	0	0	0	0
HMS 2237	0	0	0	0	0	0	0	0
HMS 2241	0	0	0	0	0	0	0	0
HMS 2258	0	0	0	0	0	0	0	0
HMS 2268	0	0	0	0	0	1	0	0
HMS 2271	0	0	0	0	0	0	0	0
HMS 2280	0	0	0	0	0	1	0	0
HMS 2281	0	0	0	0	0	0	0	0
HMS 2296	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	0	0
HMS 2355	0	0	0	0	0	0	0	0
HMS 2370	0	0	1	0	0	0	0	0
HMS 2391	0	0	0	0	0	0	0	0
HMS 2394	0	0	1	0	0	0	0	0
HMS 2400	0	0	0	0	0	0	0	0
HMS 2405	0	0	0	0	0	0	0	0
HMS 2414	0	0	0	0	0	0	0	0
HMS 2513	0	0	0	0	0	0	0	0
HMS 2536	1	0	0	0	0	0	0	0
HMS 2580	0	0	0	0	0	0	0	0
HMS 2584	0	0	0	0	0	0	0	0
HCU 2398	0	0	0	0	1	0	0	0
HCU 2613	0	0	0	0	0	0	0	0
HCU 2617	0	0	0	0	1	0	0	0
HCU 2633	0	0	0	0	0	0	0	0
HCU 2647	0	0	0	0	0	0	0	0
HCU 2654	0	0	0	0	0	0	0	0
HCU 2663	0	0	0	0	0	1	0	0
HCU 2671	0	0	0	0	0	0	0	0
HCU 2703	0	0	0	0	0	0	0	0
HCU 2706	0	0	0	0	0	0	0	0
HCU 2728	0	0	0	0	0	0	0	0
HCU 2777	0	0	0	0	0	0	0	0
HCU 2821	0	0	0	0	0	0	0	0
HCU 2864	0	0	0	0	0	0	0	0
HCU 2886	0	0	0	0	0	0	0	0
HCU 2875	0	0	0	0	0	0	0	0
HCU 2942	0	0	0	0	0	0	0	0
HCU 3151	0	0	0	0	0	0	0	0
HSJ 025	0	0	0	0	0	0	0	1
HSJ 029	0	0	0	0	0	1	0	0
HSJ 067	0	1	0	0	0	0	0	0
NHN5	0	0	0	0	0	0	0	0
HM77	0	0	0	0	0	0	0	0
HM764	0	0	0	0	0	0	1	0
HM903	0	0	0	0	0	0	0	0
N4	0	0	0	0	0	0	0	0
W4	0	0	0	0	0	0	1	0
990172	0	0	0	0	0	0	0	0
CAM22	0	0	0	0	0	0	1	0

Name of the strain	2367679	2430159	2442348	2550065	2555228	2555718	2559506	2559556
[a] **H37Rv**	0	1	0	1	0	0	0	0
HMS 1301	0	0	1	0	0	0	1	0
HMS 1356	0	0	0	0	0	0	0	0
HMS 1470	0	0	0	0	0	0	0	0
HMS 1480	0	0	0	0	0	0	0	0
HMS 1834	0	0	0	0	0	0	0	0
HMS 2103	0	0	0	0	0	0	0	1
HMS 2193	0	0	0	0	0	0	0	0
HMS 2200	0	0	0	0	0	0	0	0
HMS 2206	0	0	0	0	0	0	0	0
HMS 2208	0	0	0	0	0	0	0	0
HMS 2224	0	0	0	0	0	0	0	0
HMS 2227	1	0	0	0	0	0	0	0
HMS 2237	0	0	0	0	0	0	0	0
HMS 2241	0	0	0	0	0	0	0	0
HMS 2258	0	0	0	0	0	0	0	0
HMS 2268	0	0	0	0	0	0	0	0
HMS 2271	0	0	0	0	0	0	0	0
HMS 2280	0	0	0	0	0	0	0	0
HMS 2281	0	0	0	0	0	0	0	0
HMS 2296	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	0	0
HMS 2355	0	0	0	0	0	0	0	0
HMS 2370	0	0	0	0	0	0	0	0
HMS 2391	0	0	0	0	0	0	0	0
HMS 2394	0	0	0	0	0	0	0	0
HMS 2400	0	0	0	0	0	0	0	0
HMS 2405	0	0	0	0	0	0	0	0
HMS 2414	0	0	0	0	0	0	0	0
HMS 2513	0	0	0	0	0	0	0	0
HMS 2536	0	0	0	0	0	0	0	0
HMS 2580	0	0	0	0	1	0	0	0
HMS 2584	0	0	0	0	0	0	0	0
HCU 2398	0	0	0	0	0	0	0	0
HCU 2613	0	0	0	0	0	0	0	0
HCU 2617	0	0	0	0	0	0	0	0
HCU 2633	0	0	0	0	0	0	0	0
HCU 2647	0	0	0	0	0	0	0	0
HCU 2654	0	0	0	0	0	0	0	0
HCU 2663	0	0	0	0	0	0	0	0
HCU 2671	0	0	0	0	0	0	0	0
HCU 2703	0	0	0	0	0	0	0	0
HCU 2706	0	0	0	0	0	0	0	0
HCU 2728	0	0	0	0	0	0	0	0
HCU 2777	0	0	0	0	0	0	0	0
HCU 2821	0	0	0	0	0	0	0	0
HCU 2864	0	0	0	0	0	0	0	0
HCU 2886	0	0	0	0	0	0	0	0
HCU 2875	0	0	0	0	0	0	0	0
HCU 2942	0	0	0	0	0	0	0	0
HCU 3151	0	0	0	0	0	0	0	0
HSJ 025	0	0	0	0	0	0	0	0
HSJ 029	0	0	0	0	0	0	0	0
HSJ 067	0	0	0	0	0	0	0	0
NHN5	0	0	0	0	0	0	0	0
HM77	0	0	0	0	0	1	0	0
HM764	0	0	0	0	0	0	0	0
HM903	0	0	0	0	0	0	0	0
N4	0	0	0	0	0	0	0	0
W4	1	0	0	0	0	0	0	0
990172	0	0	0	0	0	0	0	0
CAM22	0	0	0	0	0	0	0	0

Name of the strain	2559570	2605438	2610861	2627275	2627494	2627829	2633979	2634022-2634048
[a] **H37Rv**	0	0	0	0	0	0	0	0
HMS 1301	0	0	0	0	0	0	0	0
HMS 1356	0	0	0	0	0	0	0	1
HMS 1470	0	0	0	0	0	0	0	1
HMS 1480	0	0	0	0	0	0	0	0
HMS 1834	0	0	0	0	0	0	0	0
HMS 2103	0	0	0	0	0	0	0	0
HMS 2193	0	0	0	0	0	0	0	0
HMS 2200	0	0	0	0	0	0	0	0
HMS 2206	0	0	0	0	0	0	0	0
HMS 2208	0	0	0	0	0	0	0	0
HMS 2224	0	0	1	0	0	0	0	0
HMS 2227	0	0	0	0	0	0	0	0
HMS 2237	0	0	0	0	0	0	0	0
HMS 2241	0	0	0	0	0	0	0	0
HMS 2258	0	0	0	0	0	0	0	0
HMS 2268	0	0	0	0	0	0	0	0
HMS 2271	0	0	0	0	0	0	0	0
HMS 2280	0	0	0	0	0	0	0	0
HMS 2281	0	0	0	0	0	0	0	0
HMS 2296	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	0	0
HMS 2355	0	0	0	0	0	0	0	0
HMS 2370	0	0	0	0	0	0	0	0
HMS 2391	0	0	0	0	0	0	0	0
HMS 2394	0	0	0	0	0	0	0	0
HMS 2400	0	0	0	0	0	0	0	0
HMS 2405	0	0	0	0	0	0	0	0
HMS 2414	0	0	0	0	0	0	0	0
HMS 2513	0	0	1	1	0	0	0	0
HMS 2536	1	0	0	0	1	0	1	0
HMS 2580	0	0	0	0	0	0	0	0
HMS 2584	0	0	0	0	0	0	0	0
HCU 2398	0	0	0	0	0	0	0	1
HCU 2613	0	0	0	0	0	0	0	0
HCU 2617	0	0	0	0	0	0	0	1
HCU 2633	0	0	0	0	0	0	0	0
HCU 2647	0	0	0	0	0	0	0	0
HCU 2654	0	0	0	0	0	0	0	0
HCU 2663	0	0	0	0	0	0	0	0
HCU 2671	0	0	0	0	0	0	0	0
HCU 2703	0	0	0	0	0	0	0	0
HCU 2706	0	0	0	0	0	0	0	0
HCU 2728	0	0	0	0	0	0	0	0
HCU 2777	0	0	0	0	0	0	0	0
HCU 2821	0	0	1	0	0	0	0	0
HCU 2864	0	0	0	0	0	0	0	0
HCU 2886	0	0	0	0	0	0	0	0
HCU 2875	0	0	0	0	0	0	0	0
HCU 2942	0	0	0	0	0	0	0	0
HCU 3151	0	0	0	0	0	0	0	0
HSJ 025	0	0	0	0	0	0	0	0
HSJ 029	0	0	0	0	0	0	0	0
HSJ 067	0	0	0	0	0	0	0	0
NHN5	0	0	0	0	0	0	0	0
HM77	0	0	0	0	0	0	0	0
HM764	0	0	0	0	0	1	0	0
HM903	0	0	0	0	0	0	0	0
N4	0	0	0	0	0	0	0	0
W4	0	0	0	0	0	0	0	0
990172	0	0	0	0	0	0	0	0
CAM22	0	1	0	0	0	0	0	0

Name of the strain	2634148	2635628	2635592-2637688	2639543	2639612	2732535	2732535-2732670	2732604
[a] **H37Rv**	0	1	0	0	0	0	0	0
HMS 1301	0	0	1	0	0	0	0	0
HMS 1356	0	0	0	0	0	0	0	0
HMS 1470	0	0	0	0	0	0	0	0
HMS 1480	0	0	0	0	0	0	0	0
HMS 1834	0	0	0	0	0	0	0	0
HMS 2103	0	0	0	0	0	0	0	0
HMS 2193	0	0	0	0	0	0	0	0
HMS 2200	0	0	0	0	0	0	0	0
HMS 2206	0	0	0	0	0	0	0	0
HMS 2208	0	0	0	0	0	0	0	0
HMS 2224	0	0	0	0	0	0	0	0
HMS 2227	0	0	0	0	0	0	0	0
HMS 2237	0	0	0	0	0	0	0	0
HMS 2241	0	0	0	0	0	0	0	0
HMS 2258	0	0	0	0	0	0	0	0
HMS 2268	0	0	0	0	0	0	0	0
HMS 2271	0	0	0	0	0	0	0	0
HMS 2280	0	0	0	0	0	0	0	0
HMS 2281	1	0	0	0	0	0	0	0
HMS 2296	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	0	0
HMS 2355	0	0	0	0	0	0	0	0
HMS 2370	0	0	0	0	0	0	0	0
HMS 2391	0	0	0	0	0	0	0	0
HMS 2394	0	0	0	0	0	0	0	0
HMS 2400	0	0	0	0	0	0	1	0
HMS 2405	0	0	0	0	0	0	0	0
HMS 2414	0	0	0	0	0	0	0	0
HMS 2513	0	0	0	0	0	0	0	0
HMS 2536	0	0	0	0	0	0	0	0
HMS 2580	0	0	0	0	0	0	0	0
HMS 2584	0	0	0	0	0	0	0	0
HCU 2398	0	0	0	0	0	1	0	0
HCU 2613	0	0	0	0	0	0	0	0
HCU 2617	0	0	0	0	0	0	0	1
HCU 2633	0	0	0	0	0	0	0	0
HCU 2647	0	0	0	0	0	0	0	0
HCU 2654	0	0	0	0	0	0	0	0
HCU 2663	0	0	0	1	0	0	0	0
HCU 2671	0	0	0	0	0	0	0	0
HCU 2703	0	0	0	0	0	0	0	0
HCU 2706	0	0	0	0	0	0	0	0
HCU 2728	0	0	0	0	0	0	0	0
HCU 2777	0	0	0	0	0	0	0	0
HCU 2821	0	0	0	0	0	0	0	0
HCU 2864	0	0	0	0	0	0	0	0
HCU 2886	0	0	0	0	0	0	0	0
HCU 2875	0	0	0	0	0	0	0	0
HCU 2942	0	0	0	0	0	0	0	0
HCU 3151	0	0	0	0	0	0	0	0
HSJ 025	0	0	0	0	0	0	0	0
HSJ 029	0	0	0	0	0	0	0	0
HSJ 067	0	0	0	0	0	0	0	0
NHN5	0	0	0	0	0	0	0	0
HM77	0	0	0	0	0	0	0	0
HM764	0	0	0	0	1	0	0	0
HM903	0	0	0	0	0	0	0	0
N4	0	0	0	0	0	0	0	0
W4	0	0	0	0	0	0	0	0
990172	0	0	0	0	0	0	0	0
CAM22	0	0	0	0	0	0	0	0

Name of the strain	2784657	2972160	2982432	3114403	3114442	3114795	3114873	3115131
[a] **H37Rv**	1	1	0	0	0	0	0	0
HMS 1301	0	0	0	0	0	0	0	0
HMS 1356	0	0	0	0	1	0	0	0
HMS 1470	0	0	0	0	0	0	0	0
HMS 1480	0	0	0	0	0	0	0	0
HMS 1834	0	0	0	0	0	0	0	0
HMS 2103	0	0	0	0	0	0	0	0
HMS 2193	0	0	0	0	0	0	0	0
HMS 2200	0	0	0	0	0	0	0	0
HMS 2206	0	0	0	0	0	0	0	0
HMS 2208	0	0	0	0	0	0	0	0
HMS 2224	0	0	0	0	0	0	0	0
HMS 2227	0	0	0	0	0	0	0	0
HMS 2237	0	0	0	0	0	0	0	0
HMS 2241	0	0	0	0	0	0	1	0
HMS 2258	0	0	0	0	0	0	0	0
HMS 2268	0	0	0	0	0	0	0	0
HMS 2271	0	0	1	0	0	0	0	0
HMS 2280	0	0	0	0	0	0	0	0
HMS 2281	0	0	0	0	0	0	0	0
HMS 2296	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	0	0
HMS 2355	0	0	0	0	0	0	0	0
HMS 2370	0	0	0	1	0	0	0	0
HMS 2391	0	0	0	0	0	0	0	0
HMS 2394	0	0	0	0	0	0	0	0
HMS 2400	0	0	0	0	0	0	0	0
HMS 2405	0	0	0	0	0	0	0	1
HMS 2414	0	0	0	0	0	0	0	0
HMS 2513	0	0	0	0	0	1	0	0
HMS 2536	0	0	0	0	0	0	0	0
HMS 2580	0	0	0	0	0	0	0	0
HMS 2584	0	0	0	0	0	0	0	0
HCU 2398	0	0	0	0	0	0	0	0
HCU 2613	0	0	0	0	0	0	0	0
HCU 2617	0	0	0	0	0	0	0	0
HCU 2633	0	0	0	0	0	0	0	0
HCU 2647	0	0	0	0	0	0	0	0
HCU 2654	0	0	0	0	0	0	0	0
HCU 2663	0	0	0	0	0	0	0	0
HCU 2671	0	0	0	0	0	0	0	0
HCU 2703	0	0	0	0	0	0	0	0
HCU 2706	0	0	0	0	0	0	0	0
HCU 2728	0	0	0	0	0	0	0	0
HCU 2777	0	0	0	0	0	0	0	0
HCU 2821	0	0	0	0	0	0	0	0
HCU 2864	0	0	0	0	0	0	0	0
HCU 2886	0	0	0	0	0	0	0	0
HCU 2875	0	0	0	0	0	0	0	0
HCU 2942	0	0	0	0	0	0	0	0
HCU 3151	0	0	0	0	0	0	0	0
HSJ 025	0	0	0	0	0	0	0	0
HSJ 029	0	0	0	0	0	0	0	0
HSJ 067	0	0	0	0	0	0	0	0
NHN5	0	0	0	0	0	0	0	0
HM77	0	0	0	0	0	0	0	0
HM764	0	0	0	0	0	0	0	0
HM903	0	0	0	0	0	0	0	0
N4	0	0	0	0	0	0	0	0
W4	0	0	0	0	0	0	0	0
990172	0	0	0	0	0	0	0	0
CAM22	0	0	0	0	0	0	0	0

Name of the strain	3115231	3115489	3120523	3120566	3125228	3125708	3125851	3127417
[a] **H37Rv**	0	0	0	1	0	0	0	0
HMS 1301	0	0	0	0	0	0	0	0
HMS 1356	0	0	0	0	0	0	0	0
HMS 1470	0	0	0	0	0	0	0	0
HMS 1480	0	0	0	0	0	0	0	0
HMS 1834	0	0	1	0	0	0	0	0
HMS 2103	0	0	1	0	0	0	0	0
HMS 2193	0	0	1	0	0	0	0	0
HMS 2200	0	0	1	0	0	0	0	0
HMS 2206	0	0	1	0	0	0	0	0
HMS 2208	1	0	1	0	0	0	0	0
HMS 2224	0	0	1	0	0	0	0	0
HMS 2227	0	0	1	0	0	0	1	0
HMS 2237	0	0	1	0	0	0	0	0
HMS 2241	0	0	1	0	0	0	0	0
HMS 2258	0	0	1	0	0	0	0	0
HMS 2268	0	0	1	0	0	0	0	0
HMS 2271	0	0	1	0	0	0	0	0
HMS 2280	0	0	1	0	0	0	0	0
HMS 2281	0	0	1	0	0	0	0	0
HMS 2296	0	1	1	0	0	0	0	0
HMS 2310	0	0	1	0	0	0	0	0
HMS 2355	0	0	1	0	0	0	0	0
HMS 2370	0	0	1	0	0	0	0	0
HMS 2391	0	0	1	0	0	0	0	0
HMS 2394	0	0	1	0	0	0	0	0
HMS 2400	0	0	0	0	0	0	0	0
HMS 2405	0	0	1	0	0	0	0	0
HMS 2414	0	0	1	0	0	0	0	0
HMS 2513	0	0	1	0	0	0	0	0
HMS 2536	0	0	1	0	0	0	0	0
HMS 2580	0	0	0	0	0	0	0	0
HMS 2584	0	0	0	0	0	0	0	1
HCU 2398	0	0	0	0	0	0	0	0
HCU 2613	0	0	1	0	1	0	0	0
HCU 2617	0	0	0	0	0	0	0	0
HCU 2633	0	0	1	0	0	0	0	0
HCU 2647	0	0	1	0	0	0	0	0
HCU 2654	0	0	1	0	0	0	0	0
HCU 2663	0	0	1	0	0	0	0	0
HCU 2671	0	0	1	0	0	0	0	0
HCU 2703	0	0	1	0	0	0	0	0
HCU 2706	0	0	1	0	0	0	0	0
HCU 2728	0	0	1	0	0	0	0	0
HCU 2777	0	0	1	0	0	0	0	0
HCU 2821	0	0	1	0	0	0	0	0
HCU 2864	0	0	1	0	0	0	0	0
HCU 2886	0	0	1	0	0	0	0	0
HCU 2875	0	0	1	0	0	0	0	0
HCU 2942	0	0	1	0	0	1	0	0
HCU 3151	0	0	1	0	0	0	0	0
HSJ 025	0	0	1	0	0	0	0	0
HSJ 029	0	0	1	0	0	0	0	0
HSJ 067	0	0	0	0	0	0	0	0
NHN5	0	0	0	0	0	0	0	0
HM77	0	0	0	0	0	0	0	0
HM764	0	0	0	0	0	0	0	0
HM903	0	0	0	0	0	0	0	0
N4	0	0	0	0	0	0	0	0
W4	0	0	0	0	0	0	0	0
990172	0	0	0	0	0	0	0	0
CAM22	0	0	0	0	0	0	0	0

Name of the strain	3127927	3310163	3378553	3378898	3379027	3480373	3491592	3493192
[a] **H37Rv**	0	0	0	0	0	0	0	0
HMS 1301	1	0	1	0	0	0	0	0
HMS 1356	1	0	0	0	0	0	0	0
HMS 1470	1	0	1	0	1	0	0	0
HMS 1480	1	0	1	0	1	0	1	0
HMS 1834	0	0	0	0	0	1	0	0
HMS 2103	0	0	0	0	0	0	0	0
HMS 2193	0	0	0	0	0	1	0	0
HMS 2200	0	0	0	0	0	0	0	0
HMS 2206	0	0	0	0	0	0	0	0
HMS 2208	0	0	0	0	0	1	0	0
HMS 2224	0	0	0	0	0	0	0	0
HMS 2227	0	0	0	1	0	0	0	0
HMS 2237	0	0	0	0	0	0	0	0
HMS 2241	0	0	0	1	0	0	0	0
HMS 2258	0	0	0	0	0	0	0	0
HMS 2268	0	0	0	0	0	1	0	0
HMS 2271	0	0	0	0	0	1	0	0
HMS 2280	0	0	0	0	0	1	0	0
HMS 2281	0	0	0	0	0	1	0	0
HMS 2296	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	0	0
HMS 2355	0	0	0	0	0	0	0	0
HMS 2370	0	0	0	0	0	0	0	0
HMS 2391	0	0	0	0	0	1	0	0
HMS 2394	0	0	0	0	0	0	0	0
HMS 2400	1	0	1	0	0	0	0	0
HMS 2405	0	0	0	0	0	0	0	0
HMS 2414	0	0	0	0	0	0	0	0
HMS 2513	0	0	0	0	0	0	0	0
HMS 2536	0	0	0	0	0	0	0	0
HMS 2580	1	0	0	0	0	0	0	0
HMS 2584	0	0	0	0	0	0	0	0
HCU 2398	1	0	1	0	1	0	0	0
HCU 2613	0	0	0	0	0	1	0	0
HCU 2617	1	0	1	0	1	0	0	0
HCU 2633	0	0	0	0	0	1	0	0
HCU 2647	0	0	0	0	0	1	0	0
HCU 2654	0	0	0	0	0	1	0	0
HCU 2663	0	0	0	0	0	0	0	0
HCU 2671	0	0	0	0	0	0	0	0
HCU 2703	0	0	0	0	0	1	0	0
HCU 2706	0	0	0	0	0	0	0	0
HCU 2728	0	0	0	1	0	0	0	0
HCU 2777	0	0	0	0	0	0	0	0
HCU 2821	0	0	0	0	0	0	0	0
HCU 2864	0	0	0	0	0	0	0	0
HCU 2886	0	0	0	0	0	1	0	0
HCU 2875	0	0	0	0	0	0	0	0
HCU 2942	0	0	0	0	0	0	0	0
HCU 3151	0	0	0	0	0	0	0	0
HSJ 025	0	0	0	0	0	0	0	0
HSJ 029	0	0	0	0	0	1	0	1
HSJ 067	1	0	1	0	1	0	0	0
NHN5	1	0	1	0	0	0	0	0
HM77	1	0	1	0	0	0	0	0
HM764	1	0	0	0	0	0	0	0
HM903	1	0	1	0	0	0	0	0
N4	1	1	1	0	0	0	0	0
W4	1	0	1	0	1	0	0	0
990172	1	0	1	0	0	0	0	0
CAM22	2	0	0	0	0	0	1	0

Name of the strain	3493843	3493844-3493908	3494114	3494372	3494395	3504821	3547342	3549051
[a] **H37Rv**	0	0	0	0	0	0	0	0
HMS 1301	0	0	0	0	0	0	0	0
HMS 1356	0	1	0	0	0	0	0	0
HMS 1470	0	1	0	0	0	0	0	0
HMS 1480	0	1	0	0	0	0	0	0
HMS 1834	0	0	0	0	0	0	0	0
HMS 2103	0	0	0	0	0	0	0	0
HMS 2193	0	0	0	0	0	0	0	0
HMS 2200	0	0	0	0	1	0	0	0
HMS 2206	0	0	0	0	0	0	0	0
HMS 2208	0	0	0	0	0	0	0	0
HMS 2224	0	0	1	0	0	0	0	0
HMS 2227	0	0	0	0	0	0	0	0
HMS 2237	0	0	0	0	0	0	0	0
HMS 2241	0	0	0	0	0	0	0	0
HMS 2258	0	0	0	0	0	0	0	0
HMS 2268	0	0	0	0	0	0	0	0
HMS 2271	0	0	0	0	0	0	0	0
HMS 2280	0	0	0	0	0	0	0	0
HMS 2281	0	0	0	0	0	0	0	0
HMS 2296	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	0	0
HMS 2355	0	0	0	0	0	0	0	0
HMS 2370	0	0	0	0	0	0	0	0
HMS 2391	0	0	0	0	0	1	0	0
HMS 2394	0	0	0	0	0	0	0	0
HMS 2400	0	0	0	0	0	0	0	0
HMS 2405	0	0	0	0	0	0	0	0
HMS 2414	0	0	0	0	0	0	0	0
HMS 2513	0	0	0	0	0	0	0	0
HMS 2536	0	0	0	0	0	0	0	0
HMS 2580	0	0	0	0	0	0	0	0
HMS 2584	0	0	0	0	0	0	0	0
HCU 2398	0	1	0	0	0	0	0	0
HCU 2613	0	0	0	0	0	0	0	0
HCU 2617	0	1	0	0	0	0	0	0
HCU 2633	1	0	0	0	0	0	0	0
HCU 2647	0	0	0	0	0	0	0	0
HCU 2654	0	0	0	0	0	0	0	0
HCU 2663	0	0	0	0	0	0	0	1
HCU 2671	0	0	0	0	0	0	0	0
HCU 2703	0	0	0	0	0	0	0	0
HCU 2706	0	0	0	0	0	0	0	0
HCU 2728	0	0	0	0	0	0	0	0
HCU 2777	0	0	0	1	0	0	0	0
HCU 2821	0	0	0	0	0	0	0	0
HCU 2864	0	0	0	0	0	0	0	0
HCU 2886	1	0	0	0	0	0	0	0
HCU 2875	0	0	0	0	0	0	0	0
HCU 2942	0	0	0	0	0	0	0	0
HCU 3151	0	0	0	0	0	0	0	0
HSJ 025	0	0	0	0	0	0	0	0
HSJ 029	0	0	0	0	0	0	0	0
HSJ 067	0	0	0	0	0	0	0	0
NHN5	0	0	0	0	0	0	0	0
HM77	0	0	0	0	0	0	0	0
HM764	0	1	0	0	0	0	1	0
HM903	0	0	0	0	0	0	0	0
N4	0	0	0	0	0	0	0	0
W4	0	1	0	0	0	0	1	0
990172	0	0	0	0	0	0	0	0
CAM22	0	1	0	0	0	0	0	0

Name of the strain	3549176	3549199	3550960	3550962	3550996	3551030	3551139	3551220
[a] **H37Rv**	0	0	0	0	0	0	0	0
HMS 1301	0	0	0	0	0	0	0	0
HMS 1356	0	1	0	0	0	0	0	0
HMS 1470	0	1	0	0	0	0	0	0
HMS 1480	0	1	0	0	0	0	0	0
HMS 1834	0	0	0	0	0	0	0	0
HMS 2103	0	0	0	0	0	0	0	0
HMS 2193	0	0	0	0	0	0	1	0
HMS 2200	0	0	0	0	0	0	0	0
HMS 2206	0	0	0	1	0	0	0	0
HMS 2208	0	0	0	0	0	0	0	0
HMS 2224	0	0	0	0	0	0	0	0
HMS 2227	0	0	0	0	0	0	0	0
HMS 2237	0	0	0	0	0	0	0	0
HMS 2241	0	0	0	0	0	0	0	0
HMS 2258	0	0	0	0	0	0	0	0
HMS 2268	0	0	0	0	0	0	0	0
HMS 2271	0	0	0	0	0	0	0	0
HMS 2280	0	0	0	0	0	0	0	0
HMS 2281	0	0	0	0	0	0	0	0
HMS 2296	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	0	0
HMS 2355	0	0	0	0	1	0	0	0
HMS 2370	0	0	0	0	0	0	0	0
HMS 2391	0	0	0	0	0	0	0	1
HMS 2394	0	0	0	0	0	0	0	0
HMS 2400	0	0	0	0	0	0	0	0
HMS 2405	0	0	0	0	0	0	0	0
HMS 2414	0	0	0	0	0	0	0	0
HMS 2513	0	0	0	0	0	0	0	0
HMS 2536	0	0	0	0	0	0	0	0
HMS 2580	0	1	0	0	0	0	0	0
HMS 2584	0	0	0	0	0	0	0	0
HCU 2398	0	1	0	0	0	0	0	0
HCU 2613	0	0	0	0	0	0	0	0
HCU 2617	0	1	1	0	0	0	0	0
HCU 2633	0	0	0	0	0	0	0	0
HCU 2647	0	0	0	0	0	0	0	0
HCU 2654	1	0	0	0	0	0	0	0
HCU 2663	0	0	0	0	0	0	0	0
HCU 2671	0	0	0	0	0	0	0	0
HCU 2703	0	0	0	0	0	0	0	0
HCU 2706	0	0	0	0	0	0	0	0
HCU 2728	0	0	0	0	0	0	0	0
HCU 2777	0	0	0	0	0	0	0	0
HCU 2821	0	0	0	0	0	0	0	0
HCU 2864	0	0	0	0	0	0	0	0
HCU 2886	0	0	0	0	0	0	0	0
HCU 2875	0	0	0	0	0	0	0	0
HCU 2942	0	0	0	0	0	0	0	0
HCU 3151	0	0	0	0	0	0	0	0
HSJ 025	0	0	0	0	0	1	0	0
HSJ 029	0	0	0	0	0	0	0	0
HSJ 067	0	1	0	0	0	0	0	0
NHN5	0	0	0	0	0	0	0	0
HM77	0	1	0	0	0	0	0	0
HM764	0	1	0	0	0	0	0	0
HM903	0	0	0	0	0	0	0	0
N4	0	0	0	0	0	0	0	0
W4	0	1	0	0	0	0	0	0
990172	0	0	0	0	0	0	0	0
CAM22	0	1	0	0	0	0	0	0

Name of the strain	3551229	3551230	3551281	3552683	3552701	3552764	3554294	3559402
[a] **H37Rv**	0	0	1	0	0	1	0	0
HMS 1301	0	0	0	0	0	0	0	0
HMS 1356	0	0	0	0	0	0	0	0
HMS 1470	0	0	0	0	0	0	0	0
HMS 1480	0	0	0	0	0	0	0	0
HMS 1834	0	0	0	0	0	0	0	0
HMS 2103	0	0	0	0	0	0	0	0
HMS 2193	0	0	0	0	0	0	0	0
HMS 2200	0	0	0	0	0	0	0	0
HMS 2206	0	0	0	0	0	0	0	0
HMS 2208	0	0	0	0	0	0	0	0
HMS 2224	0	1	0	0	0	0	1	0
HMS 2227	0	0	0	0	0	0	0	0
HMS 2237	0	0	0	0	0	0	0	0
HMS 2241	0	0	0	0	0	0	0	0
HMS 2258	0	0	0	0	0	0	0	0
HMS 2268	0	0	0	0	0	0	0	0
HMS 2271	0	0	0	0	0	0	0	0
HMS 2280	0	0	0	0	0	0	0	0
HMS 2281	0	0	0	0	0	0	0	0
HMS 2296	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	0	0
HMS 2355	0	0	0	0	0	0	0	0
HMS 2370	0	0	0	0	0	0	0	0
HMS 2391	0	0	0	0	0	0	0	0
HMS 2394	0	0	0	0	0	0	0	0
HMS 2400	0	0	0	0	0	0	0	0
HMS 2405	0	0	0	0	0	0	0	0
HMS 2414	0	0	0	0	0	0	0	0
HMS 2513	0	0	0	0	0	0	0	0
HMS 2536	0	0	0	0	0	0	0	0
HMS 2580	0	0	0	1	0	0	0	0
HMS 2584	0	0	0	0	0	0	0	0
HCU 2398	0	0	0	0	0	0	0	0
HCU 2613	0	0	0	0	1	0	0	0
HCU 2617	0	0	0	0	0	0	0	0
HCU 2633	0	0	0	0	0	0	0	0
HCU 2647	0	0	0	0	0	0	0	0
HCU 2654	0	0	0	0	0	0	0	0
HCU 2663	0	0	0	0	0	0	0	0
HCU 2671	1	0	0	0	0	0	0	1
HCU 2703	0	0	0	0	0	0	0	0
HCU 2706	0	0	0	0	0	0	0	0
HCU 2728	0	0	0	0	0	0	0	0
HCU 2777	0	0	0	0	0	0	0	0
HCU 2821	0	0	0	0	0	0	0	0
HCU 2864	0	0	0	0	0	0	0	0
HCU 2886	0	0	0	0	0	0	0	0
HCU 2875	0	0	0	0	0	0	0	0
HCU 2942	0	0	0	0	0	0	0	0
HCU 3151	0	0	0	0	0	0	0	0
HSJ 025	0	0	0	0	0	0	0	0
HSJ 029	0	0	0	0	0	0	0	0
HSJ 067	0	0	0	0	0	0	0	0
NHN5	0	0	0	0	0	0	0	0
HM77	0	0	0	0	0	0	0	0
HM764	0	0	0	0	0	0	0	0
HM903	0	0	0	0	0	0	0	0
N4	0	0	0	0	0	0	0	0
W4	0	0	0	0	0	0	0	0
990172	0	0	0	0	0	0	0	0
CAM22	0	0	0	0	0	0	0	0

Name of the strain	3555250	3555513	3555753	3709336	3709622	3710433	3711737	3795100
[a] **H37Rv**	0	0	0	0	0	1	0	1
HMS 1301	0	0	0	0	0	0	1	0
HMS 1356	0	0	0	0	0	0	0	0
HMS 1470	0	0	0	0	0	0	0	0
HMS 1480	0	0	0	0	0	0	0	0
HMS 1834	0	0	0	0	0	0	0	0
HMS 2103	0	0	0	0	0	0	0	0
HMS 2193	0	0	0	0	0	0	0	0
HMS 2200	0	0	0	0	0	0	0	0
HMS 2206	0	0	0	0	0	0	0	0
HMS 2208	0	0	0	0	0	0	0	0
HMS 2224	0	0	0	0	0	0	0	0
HMS 2227	0	0	0	0	0	0	0	0
HMS 2237	0	0	0	1	0	0	0	0
HMS 2241	0	0	0	0	0	0	0	0
HMS 2258	0	0	0	0	0	0	0	0
HMS 2268	0	0	0	0	0	0	0	0
HMS 2271	0	0	0	0	0	0	0	0
HMS 2280	0	0	0	0	0	0	0	0
HMS 2281	0	0	0	0	0	0	0	0
HMS 2296	0	0	0	0	0	0	0	0
HMS 2310	0	0	0	0	0	0	0	0
HMS 2355	0	0	0	0	0	0	0	0
HMS 2370	0	0	0	0	0	0	0	0
HMS 2391	0	0	0	0	0	0	0	0
HMS 2394	0	0	0	0	1	0	0	0
HMS 2400	0	0	0	0	0	0	0	0
HMS 2405	0	0	0	0	0	0	0	0
HMS 2414	0	0	0	0	0	0	0	0
HMS 2513	1	0	0	0	0	0	0	0
HMS 2536	0	0	0	0	0	0	0	0
HMS 2580	0	0	0	0	0	0	0	0
HMS 2584	0	0	0	0	0	0	0	0
HCU 2398	0	0	0	0	0	0	0	0
HCU 2613	0	0	0	0	0	0	0	0
HCU 2617	0	0	0	0	0	0	0	0
HCU 2633	0	0	0	0	0	0	0	0
HCU 2647	0	0	0	0	0	0	0	0
HCU 2654	0	0	0	0	0	0	0	0
HCU 2663	0	0	0	0	0	0	0	0
HCU 2671	0	0	0	0	0	0	0	0
HCU 2703	0	0	0	0	0	0	0	0
HCU 2706	0	0	0	0	0	0	0	0
HCU 2728	0	0	0	0	0	0	0	0
HCU 2777	0	0	0	0	0	0	0	0
HCU 2821	0	0	0	0	0	0	0	0
HCU 2864	0	0	0	0	0	0	0	0
HCU 2886	0	0	0	0	0	0	0	0
HCU 2875	0	0	0	0	0	0	0	0
HCU 2942	0	0	1	0	0	0	0	0
HCU 3151	0	0	0	0	0	0	0	0
HSJ 025	0	0	0	0	0	0	0	0
HSJ 029	0	0	0	0	0	0	0	0
HSJ 067	0	0	0	0	0	0	0	0
NHN5	0	0	0	0	0	0	0	0
HM77	0	0	0	0	0	0	0	0
HM764	0	1	0	0	0	0	1	0
HM903	0	0	0	0	0	0	0	0
N4	0	0	0	0	0	0	0	0
W4	0	0	0	0	0	0	0	0
990172	0	0	0	0	0	0	0	0
CAM22	0	0	0	0	0	0	0	0

Name of the strain	3796689	3797823	3844681	3844737	3890830
[a] **H37Rv**	0	0	0	0	1
HMS 1301	0	1	1	0	0
HMS 1356	0	0	1	0	0
HMS 1470	0	0	1	0	0
HMS 1480	0	0	1	0	0
HMS 1834	0	0	0	0	0
HMS 2103	0	0	0	0	0
HMS 2193	0	0	0	0	0
HMS 2200	0	0	0	0	0
HMS 2206	0	0	0	0	0
HMS 2208	0	0	0	0	0
HMS 2224	0	0	0	0	0
HMS 2227	0	0	0	0	0
HMS 2237	0	0	0	0	0
HMS 2241	0	0	0	0	0
HMS 2258	0	0	0	0	0
HMS 2268	0	0	0	0	0
HMS 2271	0	0	0	0	0
HMS 2280	0	0	0	0	0
HMS 2281	0	0	0	0	0
HMS 2296	0	0	0	0	0
HMS 2310	0	0	0	0	0
HMS 2355	0	0	0	0	0
HMS 2370	0	0	0	0	0
HMS 2391	0	0	0	0	0
HMS 2394	0	0	0	0	0
HMS 2400	0	0	0	0	0
HMS 2405	0	0	0	0	0
HMS 2414	0	0	0	0	0
HMS 2513	0	0	0	0	0
HMS 2536	0	0	0	0	0
HMS 2580	0	0	1	0	0
HMS 2584	0	0	0	0	0
HCU 2398	0	0	1	0	0
HCU 2613	0	0	0	0	0
HCU 2617	0	0	1	0	0
HCU 2633	0	0	0	0	0
HCU 2647	0	0	0	0	0
HCU 2654	0	0	0	0	0
HCU 2663	0	0	0	0	0
HCU 2671	0	0	0	0	0
HCU 2703	0	0	0	0	0
HCU 2706	0	0	0	0	0
HCU 2728	0	0	0	0	0
HCU 2777	0	0	0	0	0
HCU 2821	0	0	0	0	0
HCU 2864	0	0	0	0	0
HCU 2886	0	0	0	0	0
HCU 2875	0	0	0	0	0
HCU 2942	0	0	0	0	0
HCU 3151	0	0	0	0	0
HSJ 025	0	0	0	0	0
HSJ 029	0	0	0	0	0
HSJ 067	0	0	1	0	0
NHN5	0	1	0	0	0
HM77	0	1	1	0	0
HM764	0	1	1	1	0
HM903	0	1	1	0	0
N4	0	1	0	0	0
W4	1	1	1	0	0
990172	0	0	0	0	0
CAM22	0	1	1	0	0

Annex III

Tuberculosis xxx (2011) 1–10

Contents lists available at ScienceDirect

Tuberculosis

journal homepage: http://intl.elsevierhealth.com/journals/tube

MOLECULAR ASPECTS

Deciphering the role of IS*6110* in a highly transmissible *Mycobacterium tuberculosis* Beijing strain, GC1237

Henar Alonso [a,g], Juan Ignacio Aguilo [a,g], Sofía Samper [a,b,g], Jose Antonio Caminero [c], María Isolina Campos-Herrero [d], Brigitte Gicquel [e], Roland Brosch [f], Carlos Martín [a,g], Isabel Otal [a,g,*]

[a] *Grupo de Genética de Micobacterias, Departamento de Microbiología, Medicina Preventiva y Salud Pública, Universidad de Zaragoza, Spain*
[b] *Laboratorio de Investigación Molecular, Hospital Universitario Miguel Servet, IIS Aragón, Zaragoza, Spain*
[c] *Servicio de Neumología, Hospital General de Gran Canaria, Las Palmas, Canary Islands, Spain*
[d] *Servicio de Microbiología, Hospital Universitario de Gran Canaria Dr Negrín, Las Palmas, Canary Islands, Spain*
[e] *Institut Pasteur, Unit of Mycobacterial Genetics, Paris, France*
[f] *Institut Pasteur, Integrated Mycobacterial Pathogenomics, Paris, France*
[g] *CIBER Enfermedades Respiratorias, Spain*

ARTICLE INFO

Article history:
Received 29 October 2010
Received in revised form
20 December 2010
Accepted 28 December 2010

Keywords:
Beijing genotype
IS*6110*
Promoter activity
Rv2179c

SUMMARY

The capacity of infection and the ability of *Mycobacterium tuberculosis* strains belonging to the Beijing family to spread rapidly probably result from genetic advantages and unidentified mechanisms of virulence not yet thoroughly investigated. Among the mechanisms proposed to be responsible for the varying virulence phenotypes of *M. tuberculosis* strains we find IS*6110* insertions, genetic reorganizations and deletions, which have strong influences on fitness.

Beijing family is one of the lineages with the highest number of copies of IS*6110*. By studying genetic markers characteristic for this lineage, here we have characterized the clinical isolate *M. tuberculosis* GC1237 strain responsible for important epidemic outbreaks in the Gran Canary Island. We have identified and analyzed each point of insertion of IS*6110* using a bacterial artificial chromosome (BAC) library of this strain, in addition to the use of other approximations.

Nineteen copies of IS*6110* have been localized in GC1237 genome of which, four copies of IS*6110* can act as a promoter and we have focused in the characterization of one copy located 31 bp upstream of the essential gene Rv2179c and compared to the reference strain H37Rv.

1. Introduction

Strains of the Beijing genotype were first described in China and neighboring countries in 1995.[1] They are widespread in many regions of the world and frequently cause epidemic outbreaks. Different studies have indicated that one-third of global tuberculosis (TB) cases is caused by Beijing family strains assigning this lineage to one of the most successful mycobacterial families in terms of morbidity and mortality.[2] The predominance of the Beijing lineage probably results from genetic advantages, including unidentified virulence factors and the modulation of specific host responses not yet thoroughly investigated. There are some studies that relate hypervirulence of W-Beijing strains with production of phenolic glycolipid PGL,[3,4] which is a putative virulence factor that attenuates the host's innate immune response and ability to control infection.[4] The *pks15/1* locus, described to be polymorphic among members of the *Mycobacterium tuberculosis* complex,[5] is involved in the biosynthesis of the glycolipid PGL and although it is characteristic of Beijing strains to have an intact *pks15/1*, not all members of this family are producers of PGL.[5–7]

Nowadays, Beijing strains are currently attracting considerable worldwide attention because they display important pathogenic features.[8,9] These strains are often associated with multi-drug resistance as the well-known case in New York in the 1990s caused by the W strain.[10–12] The clinical isolate *M. tuberculosis* GC1237 which belongs to the Beijing family has been responsible for epidemic outbreaks in the Gran Canary Island since its first introduction in the community in 1993.[13] Nowadays, this strain continues being predominant in the area due to its rapid and successful dissemination within the community. The increased capacity of infection and the high success rate of the Beijing family to spread rapidly could be a consequence of genetic advantages and unidentified mechanisms of virulence not yet thoroughly investigated.

* Corresponding author. Departamento de Microbiología, Medicina Preventiva y Salud Pública, Facultad de Medicina, Universidad de Zaragoza, C/Domingo Miral sn, 50009 Zaragoza, Spain. Tel.: +(34) 976 761742; fax: +(34) 976 761664.
E-mail address: otali@unizar.es (I. Otal).

1472-9792/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tube.2010.12.007

IS*6110* insertions, genetic reorganizations and deletions are some of the mechanisms proposed to be responsible for differences in the virulence phenotypes of *M. tuberculosis*.

It has been demonstrated that IS*6110* may increase the expression of neighboring virulence genes by generating new promoter sequences capable of driving their expression.[14,15] IS*6110* can upregulate downstream genes through an outward-directed promoter in its 3′ end. This activity has been demonstrated for upregulation of the two-component system PhoP/PhoR.[15] Promoter activity was orientation dependent and was localized within 110-bp fragment adjacent to the right terminal inverted repeat.[14] The fact that the Beijing lineage contains a larger number of IS*6110* copies than other lineages[16] could be related with the special characteristics of this family in terms of virulence and capacity for rapid dissemination.

In this study we have classified the *M. tuberculosis* GC1237 within the Beijing family.[6,17] We also identified the locations of IS*6110* insertions in this strain and compared them to the IS*6110* insertions in two Beijing strains, 210 and W.[18] We studied the orientation and distance to neighboring genes of each copy focusing our study in one copy of IS*6110* that is acting as a promoter located 31 bp upstream the hypothetical essential gene Rv2179c.

2. Material and methods

2.1. Bacterial strains, culture media, and growth conditions

The clinical isolate *M. tuberculosis* GC1237 and the reference *M. tuberculosis* H37Rv strains were used. Mycobacterial strains were grown in Middlebrook 7H9 broth supplemented with albumin-dextrose-catalase and 0.05% Tween 80 or in Middlebrook 7H10 medium Bacto agar supplemented with oleic acid-albumin-dextrose-catalase (Difco Laboratories, Detroit, Mich.) and 0.05% Tween 80.[19] Liquid cultures were grown to logarithmic phase to be used for macrophage infection in vitro and for mycobacterial RNA extraction. *Escherichia coli* DH10B cultures were grown in Lurina-Bertani (LB) medium supplemented with chloramphenicol (12.5 μg/ml) in order to isolate the BACs. All of the strains were grown at 37 °C pBeloBAC11 plasmid was used for the construction of the BAC library of *M. tuberculosis* GC1237. pFPV27-int, derived from pFPV27,[20] was used for the construction of the GFP strains.

2.2. Cell culture and infections

MH-S murine alveolar macrophages were obtained from HPA culture collections. Cells were grown in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 4 mM L-glutamine. Infections were performed during 4 h at a multiplicity of infection (MOI) of 10 bacteria per cell. After incubating with bacteria, cells were washed three times with PBS, and cultured in complete medium during the time indicated for each experiment.

2.3. Isolation of genomic DNA

Genomic DNA of mycobacterial strains was isolated using the CTAB method as previously described by van Soolingen et al.[21]

2.4. Construction and characterization of BAC library of M. tuberculosis GC1237 strain

The construction of BACs was carried out as previously described by Brosch et al.[22] Briefly, the *M. tuberculosis* GC1237 library was constructed by ligation of partially digested *Hin*dIII fragments (50–125 Kb) into pBeloBAC11 plasmid. From almost 10,000 clones obtained 2000 were placed into 96-well plates and stored at −80 °C. Plasmid preparations of recombinant clones for sequencing reactions were prepared in 96-well plates containing an overnight culture in 250 μl of 2X yeast-tryptone medium with 12.5 μg/ml of chloramphenicol.

BAC DNA extraction was done as previously described by Birnboim et al.[23] with minor modifications. Briefly, 100 ml of BAC-transformed *E. coli* was prepared in LB medium containing 12.5 μg/ml chloramphenicol, and the cultures were grown overnight at 37 °C with vigorous and continuous agitation. Then, the bacterial cells were collected by centrifugation. The bacterial pellet was softly resuspended in a solution of 5 ml of 50 mM glucose 10 mM EDTA, 25 mM Tris pH 8, 200 mg lysozime was added to the tube and a solution of 4 ml of ice-cold NaAc pH 4.8 was added to the mixture. The tube was placed on ice and the precipitated debris was removed by centrifugation. Then, 14 ml of chloroform/isoamilic alcohol (24/1) were added to the supernatant and after centrifugation the aqueous phase was transferred to a new microfuge tube. DNA of the BACs was precipitated adding isopropanol and finally obtained by centrifugation.

End-sequencing reactions were performed with Taq DyeDeoxy Terminator cycle sequencing kit (Applied Biosystems) by using a mixture of 13 μl of BAC DNA solution, 2 μl of primer SP6-BAC1 or T7-BAC1 (supplementary material), 2.5 μl of Big Dye, and 2.5 μl of a 5X buffer (50 mM $MgCl_2$, 50 mM Tris). Thermal cycling was performed on a thermocycler (MJ Research Inc.) with an initial denaturation step of 60 s at 95 °C, followed by 26 cycles of 15 s at 96 °C, 15 s at 56 °C, and 4 min at 60 °C. DNA was precipitated with 70 μl of 70% ethanol, centrifuged, rinsed with 70% ethanol, dried and resuspended in 2 ml of formamide-EDTA buffer. The sequencing was performed on a model 373A automatic DNA sequencer (Applied Biosistems) for 12–16 h. Sequence data were transferred to Digital workstations and edited with TED software from the Staden package.[24] Edited sequences were compared using BLAST programs to the *M. tuberculosis* H37Rv sequence database.

2.5. Identification of genomic deletions and analysis of pks15/1 region

The study of the genomic deletions of the regions of difference RD105, RD142, RD150, RD181 and RD207 in *M. tuberculosis* GC1237 strain, which identify and phylogenetically sub-classify the Beijing lineage, was performed by PCR. Other genomic deletions (RD108, RD110a, RD127, RD129, RD139BW, RD149, RD152, RD165, RD166 and RD182a) also were analyzed. The primers used in these amplifications were as described elsewhere.[6] The PCR was carried out in a total volume of 50 μl, containing 0.5 μg of DNA, 5 μl of 10x PCR buffer, 200 μM dNTPs, 12.5 pmol of each primer and 1 U Taq Gold polymerase (Roche). Before the amplification, the template was initially denatured by incubation at 94 °C for 9 min then the amplification was performed for 35 cycles of 94 °C for 30 s, corresponding annealing temperature for 30 s, and 72 °C for 1–2 min depending on the amplified product. After the last cycle, the samples were incubated at 72 °C for 10 min. The RD deletions were confirmed by DNA sequencing using H37Rv as reference genome.

The pks15/1 polymorphism in this strain was determined by PCR and sequencing as previously described in Ref.[6]

2.6. Location of the copies of IS6110 insertion sequence

The study of the presence of IS*6110* insertion sequence in GC1237 BACs library was carried out by PCR as previously described in this study with the specific primers of this sequence, Gab 1 and Gab 2 (supplementary material).

We used Ligation-mediated PCR (LMPCR) to locate the copies of IS*6110* in GC1237 strain as described by Prod'hom et al.[25] This

Please cite this article in press as: Alonso H, et al., Deciphering the role of IS*6110* in a highly transmissible *Mycobacterium tuberculosis* Beijing strain, GC1237, Tuberculosis (2011), doi:10.1016/j.tube.2010.12.007

technique amplifies both ends of each copy of IS*6110*. Briefly, genomic DNA and BACs were digested with *Sal*I or with *Sma*I and ligated to a linker containing a *Sal*I restriction site or a *Sma*I restriction site. The resulting template was then digested by *Sal*I or with *Sma*I. PCR was performed using ISA1 and ISA3, specific primers for IS*6110* directed outwards[26] and the common linker primer Salgd (supplementary material). PCR products were purified using GFX PCR DNA gel band purification kit (Amersham Pharmacia Biotech) and the restriction enzyme ExoSAP-IT® (Affymetrix).

The amplified products were sequenced with the corresponding oligonucleotides and when a match was found for a flanking region in the databases, additional primers were designed to verify the point of insertion (supplementary material). PCR amplification was carried out as previously described in this study. PCR products included the complete sequence of IS*6110* and approximately 100–200 bp of both flanking sequences.

Each amplified PCR product was sequenced using CNIO service and analyzed for homology with Tuberculist (http://genolist.pasteur.fr/TubercuList), Bovilist (http://genolist.pasteur.fr/BoviList) and NCBI (http://www.ncbi.nlm.nih.gov/) database Blast analysis.

In addition, we designed primers (supplementary material) to amplify the different regions that might include IS*6110* based in genome known locations of IS*6110* in 210 and W-Beijing strains.[18]

2.7. Construction of FPV79c and FPVIS79c strains

To construct pFPV79c plasmid, primers Rv2179c-Eco and Rv2179c-Kpn were used to amplify a 670 bp fragment from H37Rv strain. To construct pFPVIS79c plasmid, primers ISA4-Eco and Rv2179c-Kpn were used to amplify a 210 bp segment from GC1237 strain. The PCR products were inserted in pFPV27-int that had been cut with *Eco*RI and *Kpn*I enzimes. To verify the constructions, pFPV79c and pFPVIS79c plasmids were sequenced with the corresponding primers (supplementary material).

The two new plasmids were electroporated into H37Rv to obtain FPV79c and FPVIS79c strains.

2.8. Analysis of GFP expression

To study the expression of the GFP protein in FPV79c and FPVIS79c strains, the fluorescence was measured by fluorometric method (Synergy™HT, BioTek) during the bacterial growth and normalized with OD_{600}. Fluorescence median intensity (FMI) of the different GFP expressing strains was measured by FACS.

To compare GFP expression under intracellular conditions, MH-S cells were infected with different GFP expressing strains; H37Rv was used as reference. Infections were carried out in 24-well plates, where cells were seeded at 10^5 cells per well. Cells were collected at 0, 24, 48 and 72 h post-infection and they were analyzed by FACS. Previously, cells were labeled with annexinV to discern between live and dead cells. Briefly, after washing collected cells once with AnnexinV binding buffer (ABB) (BD Biosciences), they were incubated for 15 min at room temperature with annexinV APC-conjugated diluted in ABB, and then fixed with paraformaldehide 4% in calcium-containing buffer. The cytometer used in this work was a FACSaria (BDBiosciences) and the analysis was performed with the software Weasel.

2.9. Isolation of RNA from mycobacteria

2.9.1. Extracellular conditions

M. tuberculosis strains were grown at 37 °C until the desired OD_{600} under aerobic conditions. The RNA from bacterial pellet was stabilized using the RNAprotect Bacteria Reagent (QIAGEN) following manufacturer's recommendations. Cells were resuspended in 1 ml acid phenol:chloroform (5:1) and 0.4 ml lysis buffer (0.5% SDS, 20 mM NaAc, 0.1 mM EDTA) and transferred to 2 ml Lysing Matrix B screw-cap tubes containing 0.1 mm silica spheres (Q-BIOgene). Cells were disrupted by three 30 s pulses in a FastPrep homogenizer (Q-BIOgene). After centrifugation, RNA from the supernatant was further extracted with 0.9 ml chloroform:isoamyl alcohol (24:1). Total RNA was precipitated with NaAc/isopropanol and washed with 70% ethanol. RNA pellet was treated with RNase-free DNase (Ambion), and the DNA free RNA was then further purified using and RNeasy kit (Qiagen). DNA contamination was ruled out by lack of amplification products after 35 cycles of PCR and the integrity of the RNA from the different strains were checked by gel electrophoresis on a 1% agarose gel. Purified RNA was kept at −80 °C until further use.

2.9.2. Intracellular conditions

Intracellular mycobacteria RNA extraction was performed with a modified protocol based on the one described by Fontan et al.[27] Briefly, 12.10^6 MH-S cells per flask were seeded in 150 cm^2 flasks. An appropriate volume of liquid mycobacteria culture was suspended in complete medium and added to each flask containing cells. After incubation, cells were washed with phosphate buffered saline (PBS) to remove extracellular bacteria, and cultured in complete medium for 0 and 48 h post-infection, when they were treated to extract intracellular mycobacteria. With this aim, cells were lysed and homogeneized during 5 min adding 10 ml per flask of GTC buffer (25 mM sodium citrate, 4 M guanidine thiocyanate, 0.5% N-lauryl sarcosine, 0.125 M 2-mercaptoethanol and 0.5% Tween 80, pH 7.0). Next, samples were collected in 15-ml centrifuge tubes and centrifuged for 1 h at 4000 rpm. Bacterial pellets were washed with GTC buffer and centrifuged again for 1 h at 4000 rpm. After that, dry bacterial pellets were treated as described above to extract mycobacterial RNA.

2.10. RT-PCR

Reverse transcription-PCR (RT-PCR) was carried out in two steps. RT was carried out with Expand Reverse Transcriptase (Roche) using 1 μg RNA as the template and the appropriate reverse primer. Reaction mixtures were incubated at 42 °C for 90 min. RT products were the subjected to PCR amplification, using Taq Gold polymerase (Roche) as we previously described with the appropriate primers. Samples were analyzed by electrophoresis on a 1% agarose gel.

2.11. qRT-PCR

cDNA libraries from *M. tuberculosis* H37Rv and GC1237 strains were constructed as follows. One μm of RNA was mixed with 25 pmol of random hexanucleotides primers (Sigma) and 50 units of Expand Reverse Trancriptase (Roche) in a final volume of 20 μl. Reaction mixtures were incubated at 65 °C for 10 min and then at 42 °C for 90 min. Expression of Rv2179c was measured and normalized with respect to the levels of *rrnaP1* mRNA by quantitative real-time PCR (qRT-PCR). qRT-PCR was carried out in a StepOne Plus (Applied Biosystems) instrument, using the cDNA generated by RT from 25 ng of RNA as a template, 1X Power SYBER green PCR master mix (Applied Biosystems), and the appropriated primers (supplementary material), each at a concentration of 250 nM. The PCR program involved an initial denaturation step for 10 min at 95 °C, followed by 40 cycles at 95 °C for 15 s and 60 °C for 1 min. The specificity of the PCR products was confirmed by the loss of fluorescence at a single temperature, when the double-stranded DNA melted to single-stranded DNA.

Please cite this article in press as: Alonso H, et al., Deciphering the role of IS*6110* in a highly transmissible *Mycobacterium tuberculosis* Beijing strain, GC1237, Tuberculosis (2011), doi:10.1016/j.tube.2010.12.007

3. Results

3.1. Analysis and representativity of the pBeloBAC11 library of M. tuberculosis GC1237 strain

We have constructed a BAC library of GC1237 strain. To characterize the BAC clones, we systematically subjected them to end-sequencing reactions. The sequence of the clones provided us the exact BAC positions on the physical map of the reference chromosome H37Rv. A total of 59 BACs named A_{1-12} to H_{1-12} were selected covering different regions of the genome of GC1237. These BACs represented around 75% of the chromosome of the studied strain (Figure 1A).

3.2. Characterization of GC1237 strain: study of RD and pks15/1 region and analysis of extra regions

We studied the regions RD105, RD207, RD181, RD150 and RD142 in order to classify GC1237 strain within one of the four subgroups of the Beijing family.[6] The results indicated that GC1237 has deleted RD181 and RD105, a robust marker for strains pertaining to the Beijing family, and RD207 concerning the Direct Repeat Locus. The findings regarding these deletions indicate that the studied strain belongs to subgroup four of the seven groups of *M. tuberculosis* described by Kong et al.[28] We further found the deletions of RD149 and RD152, associated with mobile genetic elements such as IS*6110*[6] (Figure 1B). GC1237 presents a copy of IS*6110* in place of RD152. Finally, the studied strain presented intact regions in the rest of the analyzed regions including RD110a, RD127, RD129, RD139BW, RD165, RD166 and RD182a.

We observed two extra regions in GC1237 absent in H37Rv. When we compared the findings in Bovilist database, we found that the extra regions corresponded to Mb3356-Mb3359c and Mb2047c-Mb2048c. A copy of IS*6110* was detected between the genes corresponding to Mb2047c-Mb2048c (Figure 1B).

The study of *pks15/1* region revealed an intact gene in GC1237 (data not shown) indicating that this strain could retain the ability to produce functional glycolipid PGL in contrast to the reference strain H37Rv.

3.3. Different strategies to localize all copies of IS6110 insertion sequence in GC1237 strain

We localized 19 different copies of IS*6110* in GC1237 applying two main strategies: LMPCR technique with BAC library and genomic DNA of the studied strain and PCR with designed primers based on the known points of the insertion sequence in the two reference Beijing strains 210 and W.[18] The presence of IS*6110* in the representative GC1237 BAC library was detected using PCR with the specific oligonucleotides Gab 1 and Gab 2 and ten of total of 59 BACs were positive for IS*6110* (data not shown). Using LMPCR the ten different copies were localized in GC1237 and this technique was further utilized with genomic DNA to detect the rest of the copies present in GC1237 by amplifying one or both ends of each copy of IS*6110* and its flanking sequences. As a second strategy, based on genome known locations of IS*6110* in 210 and W strains,[18] we designed primers to amplify the different regions that might include IS*6110* in the studied strain. This design was realized using *M. tuberculosis* H37Rv as genome reference. As a result of the two different strategies followed in this study we determined the exact location of the 19 copies of IS*6110* in GC1237 strain (Table 1).

The three Beijing strains share nine IS*6110* positions, these include the two insertions characteristic of the Beijing family, between *dnaA*:*dnaN*, and in the Direct Repeat Locus (Figure 1C). GC1237 strain contains eight copies of IS*6110* in intergenic regions

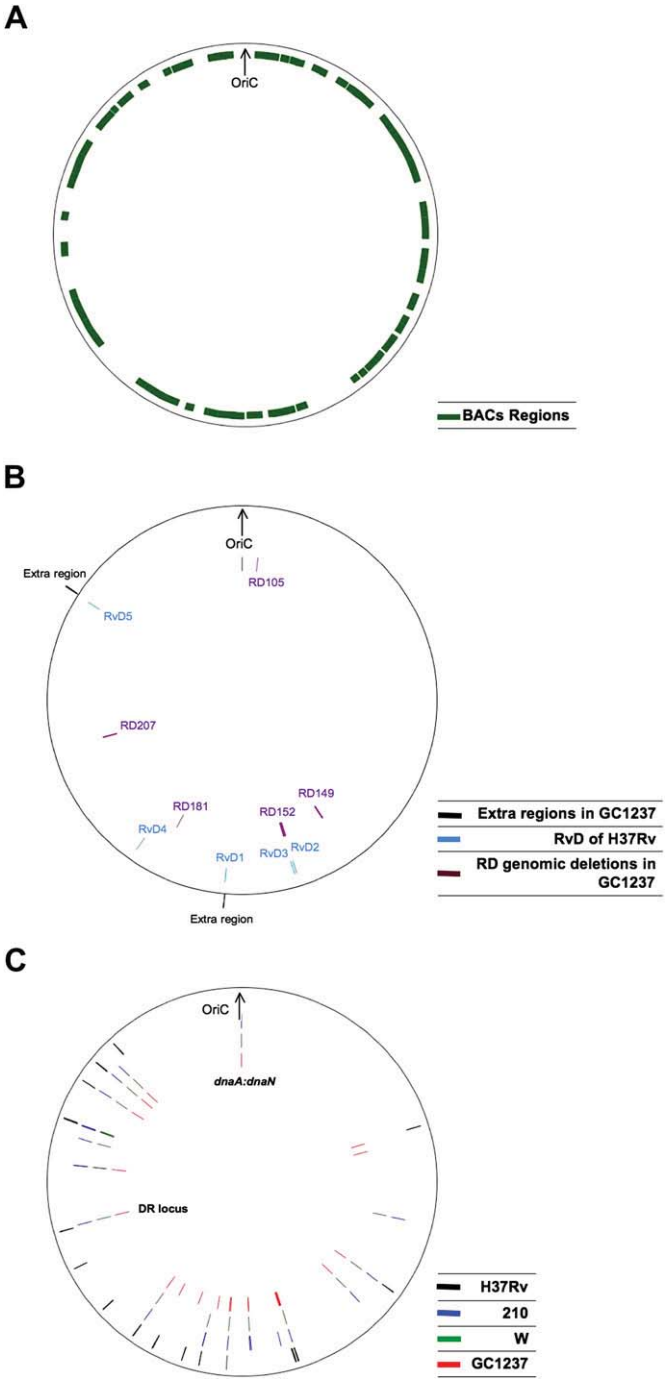


Figure 1. Circular plot of GC1237 genome prepared with DNAplotter. Graphic representation of *M. tuberculosis* GC1237 genome. (A) BACs regions. (B) Genomic regions absent in *M. tuberculosis* H37Rv (RvDs), lacking regions of difference (RD) and extra regions in GC1237 strain. (C) Location of IS*6110* in different Beijing strains GC1237, 210 and W. The DR locus and *dnaA*:*dnaN* region are marked.

and eleven in intragenic regions (Table 1). Of the eleven interrupted genes, *pip*, *ctpD*, *idsB* and Rv2180c are involved in bacterial metabolism and respiration, four correspond to genes with no predicted function and three are members of the PE and PPE family. We investigated the location of IS*6110* within Rv2180c in other *M. tuberculosis* strains including other Beijing strains (data not shown) and this point of insertion is unique to GC1237 strain.

Two IS*6110* within the *PPE34* were consecutive and in the same orientation but the second one lacks 60 bp at the 5′ end. One of the

Please cite this article in press as: Alonso H, et al., Deciphering the role of IS*6110* in a highly transmissible *Mycobacterium tuberculosis* Beijing strain, GC1237, Tuberculosis (2011), doi:10.1016/j.tube.2010.12.007

Table 1
Insertion sites of IS*6110* in *M. tuberculosis* GC1237, 210 and W.

GENES[a]	GC1237		210[d]	W[d]
	GC1237 genome[b]	BAC region[c]		
Rv0001(*dnaA*):Rv0002(*dnaN*)	+	–	+	+
Rv0794c:Rv0797 (IS*1547*)	+	B4		
Rv0840c (*pip*)	+	–		
Rv1135c		B11	+	+
Rv1371	+	C7	+	+
*Rv1469 (*ctpD*)	+	C10	+	+
Rv1754c		D4	+	+
Rv1754c-Rv1765c (RD152)	+	D4		
Rv1798:Rv1799 (*lppT*)		D6	+	
Rv1917c (*PPE34*)		D9	+	+
Rv1917c (*PPE34*)	++	D9		
Rv2016	+	D12	+	+
MT2080 (Mb2047c):MT2081 (Mb2048c)	+	D12		
Rv2077A:Rv2078	+	E1		
Rv2104c:Rv2107 (*PE22*)		E1	+	
Rv2104c:Rv2107 (*PE22*)		E1		+
Rv2107 (*PE22*):Rv2108 (*PPE36*)		E1	+	
*Rv2180c	+	–		
Rv2286c	+	E7		
Rv2352c (*PPE38*)		E8	+	+
Rv2353c (*PPE39*):Rv2356c (*PPE40*)	+	E8		
Rv2813-Rv2820c (RD207)	+	–	+	+
*Rv3018A (*PE27A*):Rv3019c (*esxR*)	+	–	+	+
Rv3019c (*esxR*):Rv3020c (*esxS*)		–	+	+
Rv3128c		G1	+	+
Rv3128c:Rv3129		G1		+
Rv3178:Rv3179		–	+	
Rv3179:Rv3180c		–	+	+
Rv3184:Rv3187		–	+	+
Rv3324A:Rv3327 (IS*1547*)	+	–	+	+
Rv3383c (*idsB*)	+	G7	+	+
*Rv3427c (IS*1532*):Rv3428c (IS*1532*)	+	G8	+	+

*Candidate locations to act as a promoter.
[a] Gene names come from the *M. tuberculosis* H37Rv genome map. One gene is listed when the location of insertion is in that gene and two genes are listed when the insertion of IS*6110* is between both.
[b] Sites are in order of appearance on the *M. tuberculosis* H37Rv genome map.
[c] Name of BAC.
[d] The plus sign indicates that the *M. tuberculosis* 210 and W contain a copy of IS*6110* in those regions.[18]

intergenic copies in GC1237, localized between *PPE39* and *PPE40* genes, is also present in H37Rv but the region is longer in the studied strain. Within the two *ipl* regions two copies of IS*6110* were located: one between Rv0794c and Rv0797, which is in the opposite orientation in H37Rv and other studied Beijing strains (data no shown), and the second one was located between Rv3324A:Rv3327 in the same orientation as reference H37Rv.

The NTF region of GC1237 does not contain any copies of IS*6110* which is an indication of being an ancestral strain different to 210 and W.[29,30]

3.4. Analysis of direct repeats of each copy of IS6110 in GC1237 genome

We analyzed the flanking regions of each of the nineteen copies of IS*6110* in GC1237 genome and detected the presence of direct repeats (DR) of 3–4 nucleotides at the extremities of sixteen IS*6110* sequences as result of IS*6110* transposition (Table 2). The other three copies analyzed were not flanked by DRs. Genomic reorganization was observed in two of these locations (RD152 and RD207), and this is probably a consequence of recombination between two adjacent copies of IS*6110*.[6] The last of the three IS*6110* elements without DR, which was localized between Rv0794c:Rv0797 was in the opposite orientation with respect to the IS*6110* in the reference genome

Table 2
Direct repeats of IS*6110* insertion sequence in GC1237 strain.

Genes of IS*6110* in GC1237 genome	Direct Repeat (DR)[a]
Rv0001 (*dnaA*):Rv0002 (*dnaN*)	ATT
Rv0794c:Rv0797	–
Rv0840c (*pip*)	ACG
Rv1371	GAGG
Rv1469 (*ctpD*)	CGT
Rv1754c-Rv1765c (RD152)	–
Rv1917c (PPE34)	TTA
Rv1917c (PPE34)	TTA
Rv2016	AGG
MT2080 (Mb2047c):MT2081 (Mb2048c)	GAA
Rv2077A:Rv2078	AGG
Rv2179c:Rv2180c	AGC
Rv2286c	ATC
Rv2353 (*PPE39*):Rv2356c (*PPE40*)	CCG
Rv2813-Rv2820c (RD207)	–
Rv3018A (*PE27A*):Rv3019c (*esxR*)	GCC
Rv3324A:Rv3327 (IS*1547*)	GGC
Rv3383c (*idsB*)	ATC
Rv3427c (IS*1532*):Rv3428c (IS*1532*)	CCCG

[a] The absence of Direct Repeat is indicated by minus.

H37Rv. The IS*6110* in H37Rv contains DR and the lack of DR in the studied IS*6110* could be explained by reorganization in this region in GC1237 genome. The two copies localized in *PPE34* are flanked by the direct repeat sequence TTA. These DRs were located in the 5′ end of the first IS*6110* and the 3′ end of the second one, probably due to simultaneous transposition of the two copies. The DR sequences differ among each other indicating the lack of transposition specificity.

3.5. Study of the orientation and distance of IS6110 to the neighboring genes

Different studies have indicated that IS*6110* could upregulate the expression of downstream genes. Previous studies have shown that when IS*6110* is inserted in the same orientation as, and close enough to, a downstream gene, IS*6110* could potentially function as a promoter.[14] We analyzed the orientation of each copy of the IS*6110* insertion sequence in GC1237 strain and the distance to the close gene in order to test the promoter function of IS*6110* and we observed that IS*6110* could act as a mobile promoter in four locations (Table 1). One of them is located 297 bp upstream of the Rv1468c gene, another 138 bp upstream of PE27A gene, the third one 41 bp upstream Rv3427c gene and the last one is located 31 bp upstream of the essential gene Rv2179c. We focused our study in the last one as this location is specific for GC1237 and the distance to the downstream gene is the shortest.

3.6. Detection of a transcript from IS6110 to aroG in GC1237 strain and from Rv2180c to aroG in H37Rv strain

To determine if a transcript extended from IS*6110* or Rv2180c into downstream genes Rv2179c and *aroG*, we extracted RNA from broth cultures of GC1237 and H37Rv strains and RT-PCR was performed using the corresponding oligonucleotides as previously described in this work. We obtained amplification products from Rv2180c to *aroG* in H37Rv (Figure 2A and C). These results indicate that the three genes are cotranscribed in an operon. In the case of GC1237, amplification products were obtained from 3′ end of IS*6110* to *aroG* suggesting that the 3′ end of IS*6110*, Rv2179c and *aroG* genes are cotranscribed (Figure 2B and D). All the RT-PCR products were sequenced and the Blast results verified their specificity. Our results indicated that the three genes are cotranscribed in an

Please cite this article in press as: Alonso H, et al., Deciphering the role of IS*6110* in a highly transmissible *Mycobacterium tuberculosis* Beijing strain, GC1237, Tuberculosis (2011), doi:10.1016/j.tube.2010.12.007

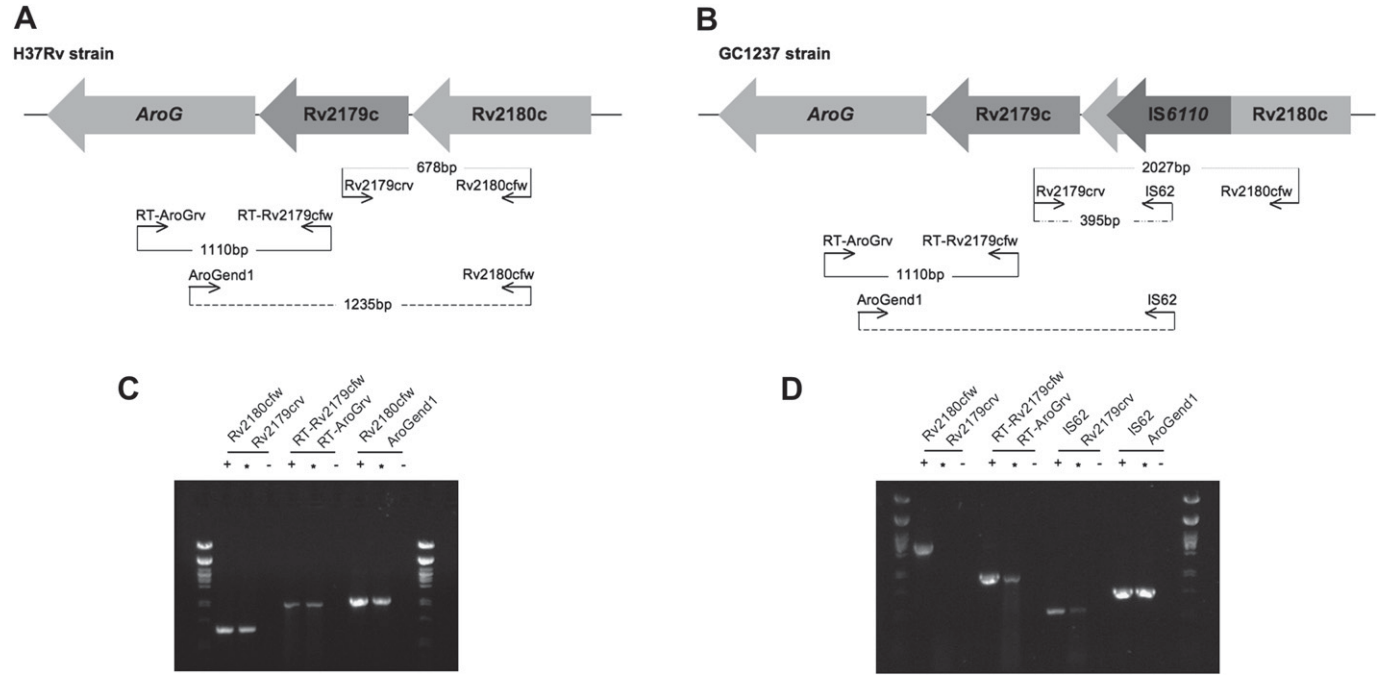


Figure 2. RT-PCR analysis of Rv2179c region from *M. tuberculosis* H37Rv and GC1237 strains. (A) Schematic diagram of Rv2179c region in H37Rv strain. (B) Schematic diagram of Rv2179c region in GC1237 strain. In both diagrams, the primers used for RT-PCR (Table 1) and the sizes of the fragments obtained with each pair of primers are indicated. The direction of transcription for Rv2179c is indicated by arrows. (C) RT-PCR analysis of this region in H37Rv. (D) RT-PCR analysis of this region in GC1237. In both RT-PCR the combination of primers is indicated above each set of reaction. Each set of the three reaction consists of a positive control PCR assay with genomic DNA as the template (+), an RT-PCR (*), and a negative control assay without reverse transcriptase (−).

operon, but these results do not exclude the possibility of independent promoters for each gene.

3.7. IS*6110* is acting as a promoter of Rv2179c gene both under extracellular and intracellular conditions

To verify that IS*6110* located upstream Rv2179c in GC1237 genome was functional as a promoter, we constructed the recombinant plasmids pFPV79c and pFPVIS79c containing the promoterless *gfp* gene preceded by the region upstream Rv2179c in H37Rv strain or in GC1237 strain, respectively. They were transformed into H37Rv obtaining FPV79c and FPVIS79c strains to compare GFP expression of both constructions in the same genetic background. As control, we used H37Rv transformed with pFPV27 which contains *gfp* gene without promoter (FPV27 strain). GFP expression was measured both by fluorometry and flow cytometry. Results of both techniques showed that the fluorescence of FPVIS79c strain is about 5-fold higher than the observed in FPV79c strain (Figure 3A). Moreover, data obtained by fluorometry during the growth curve indicated that the effect of IS*6110* on the regulation of GFP in this case was independent of bacterial growth phase (Figure 3B). Therefore, this result suggests that in GC1237, IS*6110* is acting as a promoter of Rv2179c gene.

We decided to study the promoter activity of IS*6110* upstream Rv2179c under intracellular conditions. For that issue, we infected murine immortalized alveolar macrophages (MH-S cell line) with the different GFP expressing strains and infected cells FMI was measured by FACS. Thus, to exclude non-infected cells, we also infected with non-fluorescent H37Rv, considering their FMI as negative fluorescence level (R1 in Figure 4A). On the other hand, cells contained in the region with a higher fluorescence level were established as the infected ones (R2 in Figure 4A). Moreover, data were referred to annexinV-negative cells, in order to analyze only live cells. As it is observed in Figure 4B cells infected with FPVIS79c strain showed a clear increase of their FMI values regarding the control at all times studied. These results are in agreement with the obtained data under extracellular conditions. Furthermore, GFP expression of FPVIS79c-infected cells tended to increase during the experiment, being this rise more dramatic at 72 h. In contrast, in the absence of IS*6110* this event failed to happen (Figure 4B).

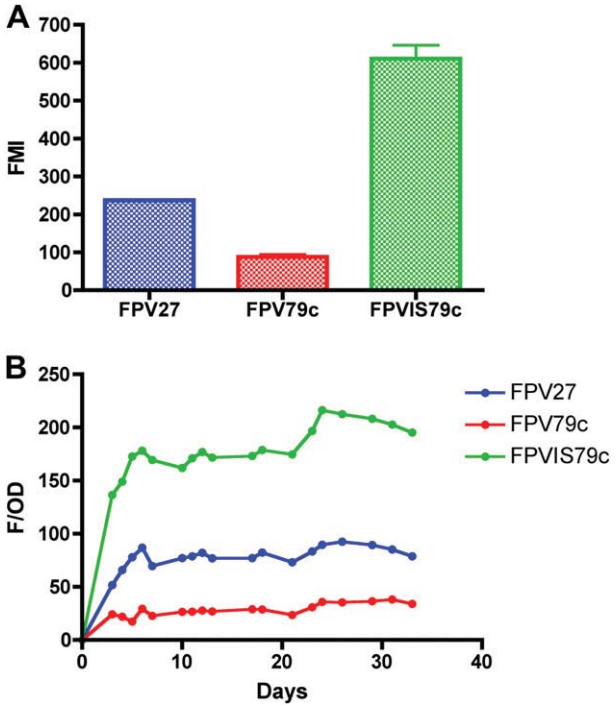


Figure 3. GFP expression in FPV79c and FPVIS79c strains. (A) Fluorescence median intensity (FMI) of FPV27, FPV79c and FPVIS79c strains was measured by flow cytometry at logarithmic phase. (B) Time course the evolution of the fluorescence of FPV27, FPV79c and FPVIS79c strains during the bacterial growth normalized with $O.D_{600}$.

Please cite this article in press as: Alonso H, et al., Deciphering the role of IS*6110* in a highly transmissible *Mycobacterium tuberculosis* Beijing strain, GC1237, Tuberculosis (2011), doi:10.1016/j.tube.2010.12.007

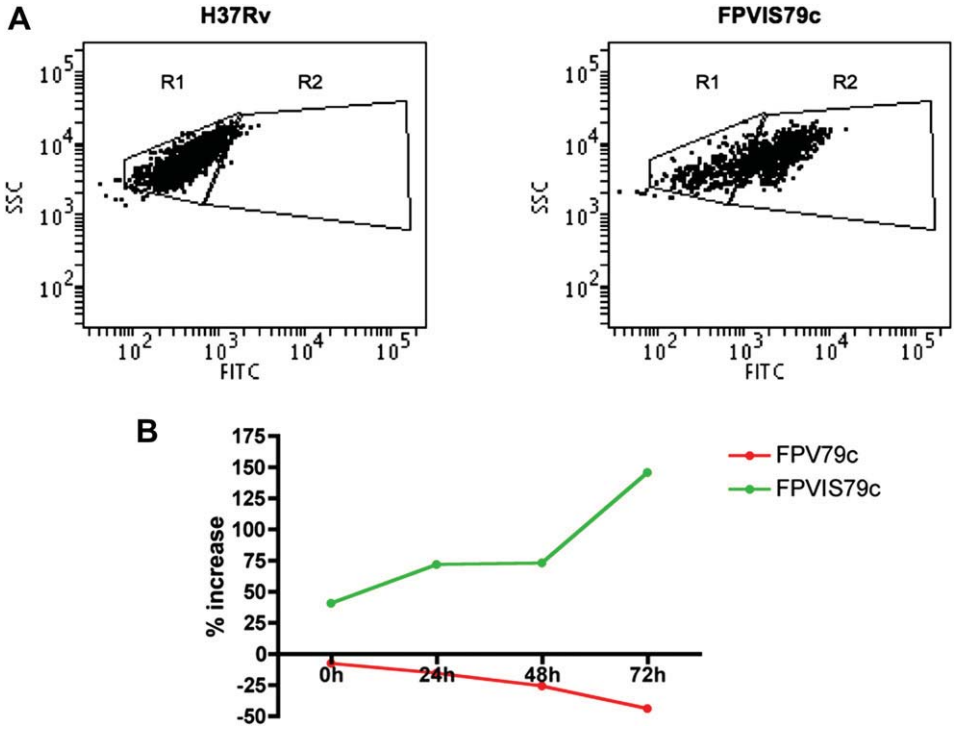


Figure 4. GFP expression in FPV79c and FPVIS79c strains in murine macrophages. (A) Histograms of H37Rv and FPVIS79c-infected cells. R1 indicates the non-fluorescent cells and R2 represents the fluorescent cells. (B) Time course of FMI of cells gated in R2, infected with FPV79c and FPVIS79c strains. The percentage of increment was calculated with respect to FMI values of FPV27 strain. Figure shows a representative experiment of three performed.

3.8. Rv2179c gene expression is increased in GC1237 strain both in broth and intracellular conditions

As it had been previously described that IS*6110* can act as a promoter of downstream genes in *M. tuberculosis*,[14] we compared the expression of Rv2179c gene in GC1237 and H37Rv strains. Thus, both strains were cultured to mid-exponential growth phase and RNAs were extracted to generate cDNAs. The expression of Rv2179c was measured by qRT-PCR and normalized to *rrnAP1* expression levels. Results demonstrated that Rv2179c expression was about 6-fold higher in GC1237 than in H37Rv (Figure 5A), which suggests that IS*6110* in this location is up-regulating gene expression.

To determine the effect of IS*6110* on Rv2179c expression during infection, bacterial RNA was obtained both from intracellular and broth growth conditions. The results showed that Rv2179c expression rose dramatically in GC1237 infected macrophages, being 6 and 10-fold higher at 0 h and 48 h post-infection, respectively (Figure 5B). On the contrary, lower increase was observed in the case of H37Rv, confirming that the presence of IS*6110* correlated with higher expression of Rv2179c in the studied strain both under intracellular and extracellular conditions.

As part of the three genes forming an operon we analyzed the expression of *aroG* by qRT-PCR, but we could not observed significant increment of this gene (data no shown). This result does not exclude the possibility of another regulation of *aroG*.

4. Discussion

The *M. tuberculosis* GC1237 strain has demonstrated an enhanced capacity to spread and transmit within a community. This strain belongs to one of the most virulent families, Beijing. The capacity of infection and the ability of this family to spread rapidly probably result from genetic advantages and unidentified mechanisms of virulence not yet thoroughly investigated.

Several reports have strongly suggested that the severity and clinical manifestations of TB depend on the immunogenicity and pathogenicity of the infecting *M. tuberculosis* strain; these properties vary considerably among strains.[31] The notion that some *M. tuberculosis* lineages, such as the W-Beijing genotype, are more virulent than others is supported by results obtained with animal and cellular models. Studies in both show that among the Beijing family strains there are hyper-virulent phenotypes that are associated with relapse and treatment failure in humans and distinct immune responses in animal models.[32] Furthermore, different studies have demonstrated that some W-Beijing strains can escape the immunoprotective efficacy of the current TB vaccine and thus act as “escape variants” of *Mycobacterium bovis* BCG vaccination.[33]

We decided to study most extensively the GC1237 strain because it has been responsible for a high number of TB cases since its introduction in the Gran Canary Island in 1990s.

Beijing strains are genetically homogeneous and conserved. They share some genetic markers, such as similar spoligotypes[1,29] characterized by the deletions of spacers 1–34 in the Direct Repeat Locus of *M. tuberculosis*; a copy of IS*6110* between *dnaA*:*dnaN* genes; and it is frequent to find at least a copy of IS*6110* in the NTF region in the strains belonging to Beijing family[34] and some members of this family may have a second insertion within this locus.[35] However, the absence of IS*6110* in NTF locus is characteristic of ancestral sublineages within the Beijing genotype.[29] Due to the course of evolution, some strains (“modern” sublineages) have acquired the insertion of IS*6110* in this region.[35] The clinical strain GC1237 presents an intact NTF locus indicating that GC1237 is an ancestral Beijing strain.

Comparative whole-genome hybridization of Beijing strains showed large sequence polymorphisms (LSPs) also known as regions of difference (RD), which subdivide the Beijing family into at least four phylogenetic subgroups, raising the possibility that there are phenotypic differences within the Beijing family. The subdivisions

Please cite this article in press as: Alonso H, et al., Deciphering the role of IS*6110* in a highly transmissible *Mycobacterium tuberculosis* Beijing strain, GC1237, Tuberculosis (2011), doi:10.1016/j.tube.2010.12.007

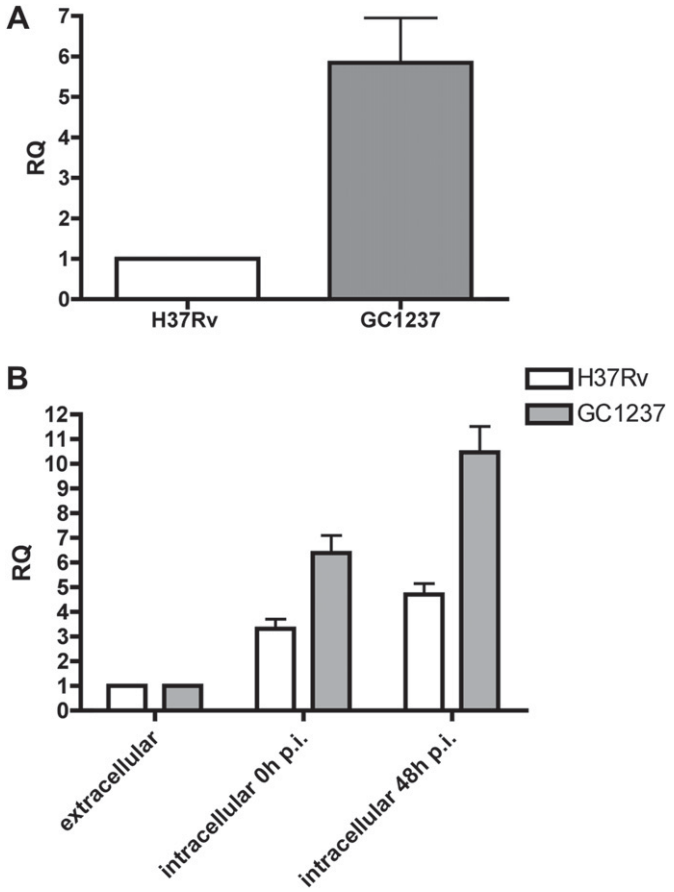


Figure 5. Quantification of Rv2179c in *M. tuberculosis* GC1237 and H37Rv by qRT-PCR. (A) Relative quantification (RQ) of Rv2179c expression in GC1237 with respect to H37Rv at logarithmic phase. (B) RQ of Rv2179c expression in intracellular conditions at 0 h and 48 h post-infection in GC1237 or H37Rv with respect to extracellular conditions. In both cases, the expression of Rv2179c was normalized to the levels of *rrnAP1* mRNA. Results are the means of three independent experiments; error bars indicate the standard deviations of the means.

are made on the basis of RD105 and RD207, and the variable appearance of RD181, RD150 and RD142 deletions.[6] Some authors have proposed an evolutionary pathway of the Beijing lineages based on the RD deletions and on the IS*6110* insertions in the NTF region indicating that ancient strains have neither deletion of RD181 nor insertion of IS*6110* in NTF region.[36] However, our results are in disagreement with this classification because GC1237 presents RD181 deleted but lacks IS*6110* in NTF locus. A recent study supports the view that genomic deletion of RD181 has occurred during the evolutionary process from ancient to modern strains proposing four genetic sublineages. According to this view, GC1237 falls in the genetic subgroup "ancient with deleted RD181", being the ancient subgroup the most predominant in Japan.[37] When we analyzed the RD152 in GC1237, we detected a copy of IS*6110* without flanking DR confirming a recombination event of two previous copies of IS*6110* resulting in the loss of this region. This finding is in agreement with different studies indicating that strains with a high number of IS*6110* copies (>14) have lost genomic regions more often than strains with only few copies,[38] as is in the case of GC1237. RD149 has also been associated with recombinations between mobile genetic elements[6] but IS*6110* is not implicated in the deletion of this region as we show the absence of IS*6110* in this region in GC1237. The two phenomenons regarding RD152 and RD149 have been observed in other Beijing strains studied in our laboratory. We also studied the deletion of other RD observed in clinical strains belonging to Beijing family from East Asia clade[6]; GC1237 has these RD intact. However, we do not rule out the possibility of the presence of specific deleted regions in GC1237. The high number of different combinations of deleted RD that we can find in Beijing strains shows that although the Beijing family is a well-defined group, there is genetic variability among these strains.

Another feature of GC1237 is the presence of two extra regions, Mb2047c-Mb2048c and Mb3356-Mb3357, with respect to H37Rv. These regions, present in other sequenced strains of *M. tuberculosis* and *M. bovis*, do not correspond exactly to the described RvD1 and RvD5, respectively. The two IS*6110* localized in these regions suggest possible implication in new genomic reorganizations with loss of these regions. By studying the flanking regions of IS*6110* compared with H37Rv, we have detected as many new regions as loss of other ones. These findings confirm the implication of IS*6110* in genomic reorganizations and its capacity to generate genomic plasticity.

The *pks*15/1 gene, which is implicated in *M. tuberculosis* virulence[4] is intact in GC1237 as well as the rest of the Beijing strains. This strain presents a single point mutation in Rv2952 gene implicated in the biosynthetic pathway of PGL as previously described by Huet et al.[39] In spite of this mutation the structural variants of PGL have no major impact on virulence.

When a strain of *M. tuberculosis* presents a high level of copies of IS*6110* insertion sequence the probability to localize all the copies by LMPCR technique is low. This technique was useful to determine the points of insertion of this sequence in *M. bovis*.[40] However, the number of copies of IS*6110* in *M. bovis* is very much minor (1–5) than the observed in *M. tuberculosis*. In this regard, the construction of BAC library of GC1237 strain has been a successful tool. The use of the different molecular strategies has allowed us to localize all IS*6110* copies inside the GC1237 genome. Except for some sequenced strains, to our knowledge this is the first time to localize all IS*6110* copies in a strain of *M. tuberculosis* containing a high number of this insertion sequence. Previous studies supported the idea that some insertion sites were prevalent in the low-copy number strains and suggested a separate evolutionary lineage for the low and the high-copy number strains of *M. tuberculosis*.[41] According to this, GC1237 does not have any copy of IS*6110* inserted at the same genomic position described for low-copy *M. tuberculosis* strains.

In sixteen of the nineteen copies, we observed DR of 3–4 nucleotides in the flanking regions. These DR indicate that the presence of IS*6110* was due to transposition events and the absence of DR in the other copies suggests rearrangements between IS*6110* elements. There are very few studies where authors report DR flanking IS*6110*.[26,40] In these studies the number of copies is low (<6) and all of them were flanked by DR, indicating that these IS*6110* are consequence of transposition. In our study, because of the high number of copies of IS*6110* we observed some copies flanked by DR and other copies without DR suggesting that the probability of rearrangement process between copies rises when the number of those increases producing more variability among strains.

Numerous studies indicate that IS*6110* insertion sequence has some preferential points of insertion (hot-spots) in the genome. The IS*1547* (*ipl* locus) is a site of intensive insertion of IS*6110*[42,43] and the studied strain corroborates this hot-spot with the insertion of a copy of IS*6110* in each *ipl* locus. On the other hand, mapping of IS*6110* insertion points in *M. tuberculosis* GC1237 has shown that no consensus target sequence was detected in the immediate vicinity of the nineteen analyzed copies. Obviously, the DR flanking IS*6110* also were different except for the sequence ATC that was found in two cases. All these support the lack of transposition specificity for IS*6110*, according with other authors.[44,45] The open reading frames represent 91% of *M. tuberculosis* genome[46] but in GC1237 the insertion of IS*6110* into coding regions occurred in 42% of the cases, suggesting that the transposition is

Please cite this article in press as: Alonso H, et al., Deciphering the role of IS*6110* in a highly transmissible *Mycobacterium tuberculosis* Beijing strain, GC1237, Tuberculosis (2011), doi:10.1016/j.tube.2010.12.007

relatively more frequent in intergenic regions. Our results agreed with other studies that found that 58% of discrete IS*6110* insertion sites occurred within coding regions in *M. tuberculosis*[47] and in *M. bovis* strains.[40] In this context, because of the insertion of IS*6110* in possible promoter regions the probability of IS*6110* influencing the expression of the neighboring genes could be increased. It is necessary to take into account that the locations of IS*6110* observed in the different genomes are the result of transposition events and the natural selection of viable mycobacteria because when insertion occurs in a necessary gene it could affect survival and we cannot observe it. In case of the interruption of a non-essential gene, the phenotype of the mycobacteria could be affected. It has been demonstrated the potential for intra- and inter-genic IS*6110* insertions to add plasticity to *M. tuberculosis* genome and influence fitness of the mycobacterium.

With the exception of the copy of IS*6110* located between *dnaA* and *dnaN* genes, we have not found any copy of IS*6110* in the quarter of *M. tuberculosis* GC1237 circular genome surrounding these genes (Figure 1C). Also, we observed the same in the two Beijing strains 210 and W. This result is in agreement with other authors when they have characterized IS*6110* integration loci.[47] This finding could indicate that in this part of the genome there are no specific sequences or regions recognized by the sequence or that these genes have an important role in *M. tuberculosis* and the interruption or modulation of their expression could result in no advantage for the bacteria.

The IS*6110* insertion sequence varies in number and position within the genome generating a high level of DNA polymorphism among strains. It is known that the insertion of IS*6110* can alter bacterial gene expression. Depending on the location, IS*6110* may cause loss of gene activity or can also upregulate the expression of downstream genes by acting as a mobile promoter. When IS*6110* is inserted in the same orientation as, and close enough to, a downstream gene it could potentially function as a promoter.[14,15] According to these previous studies we analyzed the orientation and the distance to the closest gene of each copy of IS*6110* in GC1237 strain. Thus, we obtained eight genes with an IS*6110* inserted upstream, in the right orientation and at a reasonable distance to work as a promoter these genes. However, when we compared the expression of these genes in GC1237 with reference to H37Rv strain, we only found a clear increase in the case of the gene Rv2179c, whose IS*6110* is located 31 bp upstream of it. To discern that Rv2179c upregulation observed in GC1237 strain was really due to the presence of IS*6110* and not to the different genetic background, we constructed pFPV79c and pFPVIS79c plasmids and we transformed them into H37Rv. Thus, we observed that GFP expression was clearly higher when IS*6110* was located upstream as we observed in GC1237 strain, showing the IS*6110* promoter activity at this location. Moreover, we infected macrophages with the GFP-expressing H37Rv strains explained above, finding that GFP fluorescence from cells infected with FPVIS79c strain was more likely to increase across the experiment, confirming that IS*6110* promoter activity results enforced inside host cells. Furthermore, we also studied mRNA from GC1237 and H37Rv strains both under intra and extracellular conditions, detecting in the case of GC1237 a dramatic increase of Rv2179c gene expression under intracellular conditions regarding to broth ones. Curiously, although far lower than in GC1237, Rv2179c gene was also upregulated in H37Rv intracellular bacteria, which could indicate that this gene has some role during infection process. A previous work showed a reinforcement of IS*6110* promoter activity under intracellular conditions.[14] However, in that work authors studied the expression of different genes and the presence or not of IS*6110* upstream in two different strains. Nevertheless, authors did not discern the possibility of a regulation due to the different genetic background of the strains studied. Our results clearly show that the promoter activity of IS*6110* on Rv2179c is only related to disposition of the sequence with respect to this gene. We also quantified the Rv2179c expression in macrophages by qRT-PCR demonstrating that the expression of this gene in GC1237 was about 6-fold higher than in extracellular conditions. When we compared Rv2179c expression in both strains in broth culture, we observed higher expression in GC1237 than in H37Rv. These results confirm the promoter activity of IS*6110* and suggest that this activity is upregulated inside macrophages.

On the other hand, the RT-PCR results indicated that in H37Rv strain the Rv2180c, Rv2179c and Rv2178c (*aroG*) genes are cotranscribed. Differently, in GC1237 we could not detect the cotranscript between Rv2180c and Rv2179c genes due to the insertion of IS*6110* in Rv2180c, but we detected the cotranscript from the 3′ end of IS*6110* to *aroG* indicating that this sequence acts as a new promoter for the Rv2179c gene and also for *aroG* gene which is an essential gene involved in biosynthesis of chorismate precursor of the three aromatic amino acids in *M. tuberculosis*.[48] There are no available data about the function of the protein encoded by Rv2179c gene. In fact, according to Tuberculist database, the product of this gene is a conserved essential hypothical protein and it is conserved in different mycobacterial strains.[49] From our results we can hypothesize that the over-expression of this gene could be advantageous for GC1237 at least in certain environments such as infecting macrophages but we do not rule out the possibility of not having an effect. Providing that this insertion is unique in this strain we do not discard its implication in specific fitness advantages in GC1237 and on the other hand, this point of insertion can be a useful tool in epidemiological studies because it allows us to identify and differentiate this strain among Beijing strains.

In summary, this study facilitates the location of all the copies of IS*6110* in GC1237 which will enable the search of the common locations of IS*6110* in Beijing strains that would be characteristic of this family. Knowledge on distribution of IS insertion sites in Beijing strains sheds more light in evolutionary processes involving this family. The location of all the copies will allow studying the implication of each one in the different strain properties. Some of the detected copies are interrupting genes, others are implicated in recombination events and we have demonstrated that IS*6110* is acting as a promoter in Rv2179c.

Acknowledgments

We thank Dessislava Marinova for her kind help in English revision and for her valuable comments on the manuscript. Henar Alonso is financially supported by Gobierno de Aragón (B034/09).

This work is part of the TB-VIR Project supported by the European Commission under the Health Cooperation of the 7th Framework Programme 200973 and Project of the Ministerio de Ciencia e Innovación BIO2008-01561.

Appendix. Supplementary material

The supplementary data associated with this article can be found in the on-line version at doi:10.1016/j.tube.2010.12.007.

Funding: None.

Competing interests: None Declared.

Ethical approval: Not required.

References

1. van Soolingen D, Qian L, de Haas PE, Douglas JT, Traore H, Portaels F, et al. Predominance of a single genotype of Mycobacterium tuberculosis in countries of east Asia. *J Clin Microbiol* 1995;**33**:3234–8.

2. Bifani PJ, Mathema B, Kurepina NE, Kreiswirth BN. Global dissemination of the Mycobacterium tuberculosis W-Beijing family strains. *Trends Microbiol* 2002; **10**:45–52.
3. Caws M, Thwaites G, Dunstan S, Hawn TR, Lan NT, Thuong NT, et al. The influence of host and bacterial genotype on the development of disseminated disease with Mycobacterium tuberculosis. *PLoS Pathog* 2008;**4**:e1000034.
4. Reed MB, Domenech P, Manca C, Su H, Barczak AK, Kreiswirth BN, et al. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature* 2004;**431**:84–7.
5. Constant P, Perez E, Malaga W, Laneelle MA, Saurel O, Daffe M, et al. Role of the pks15/1 gene in the biosynthesis of phenolglycolipids in the Mycobacterium tuberculosis complex. Evidence that all strains synthesize glycosylated p-hydroxybenzoic methyl esters and that strains devoid of phenolglycolipids harbor a frameshift mutation in the pks15/1 gene. *J Biol Chem* 2002;**277**: 38148–58.
6. Tsolaki AG, Gagneux S, Pym AS, Goguet de la Salmoniere YO, Kreiswirth BN, Van Soolingen D, et al. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of Mycobacterium tuberculosis. *J Clin Microbiol* 2005; **43**:3185–91.
7. Sinsimer D, Huet G, Manca C, Tsenova L, Koo MS, Kurepina N, et al. The phenolic glycolipid of Mycobacterium tuberculosis differentially modulates the early host cytokine response but does not in itself confer hypervirulence. *Infect Immun* 2008;**76**:3027–36.
8. Lopez B, Aguilar D, Orozco H, Burger M, Espitia C, Ritacco V, et al. A marked difference in pathogenesis and immune response induced by different Mycobacterium tuberculosis genotypes. *Clin Exp Immunol* 2003;**133**:30–7.
9. Zhang M, Gong J, Yang Z, Samten B, Cave MD, Barnes PF. Enhanced capacity of a widespread strain of Mycobacterium tuberculosis to grow in human macrophages. *J Infect Dis* 1999;**179**:1213–7.
10. Agerton TB, Valway SE, Blinkhorn RJ, Shilkret KL, Reves R, Schluter WW, et al. Spread of strain W, a highly drug-resistant strain of Mycobacterium tuberculosis, across the United States. *Clin Infect Dis* 1999;**29**:85–92. discussion 3–5.
11. Bifani PJ, Mathema B, Liu Z, Moghazeh SL, Shopsin B, Tempalski B, et al. Identification of a W variant outbreak of Mycobacterium tuberculosis via population-based molecular epidemiology. *Jama* 1999;**282**:2321–7.
12. Bifani PJ, Plikaytis BB, Kapur V, Stockbauer K, Pan X, Lutfey ML, et al. Origin and interstate spread of a New York City multidrug-resistant Mycobacterium tuberculosis clone family. *Jama* 1996;**275**:452–7.
13. Caminero JA, Pena MJ, Campos-Herrero MI, Rodriguez JC, Garcia I, Cabrera P, et al. Epidemiological evidence of the spread of a Mycobacterium tuberculosis strain of the Beijing genotype on Gran Canaria Island. *Am J Respir Crit Care Med* 2001;**164**:1165–70.
14. Safi H, Barnes PF, Lakey DL, Shams H, Samten B, Vankayalapati R, et al. IS6110 functions as a mobile, monocyte-activated promoter in Mycobacterium tuberculosis. *Mol Microbiol* 2004;**52**:999–1012.
15. Soto CY, Menendez MC, Perez E, Samper S, Gomez AB, Garcia MJ, et al. IS6110 mediates increased transcription of the phoP virulence gene in a multidrug-resistant clinical isolate responsible for tuberculosis outbreaks. *J Clin Microbiol* 2004;**42**:212–9.
16. Gutacker MM, Mathema B, Soini H, Shashkina E, Kreiswirth BN, Graviss EA, et al. Single-nucleotide polymorphism-based population genetic analysis of Mycobacterium tuberculosis strains from 4 geographic sites. *J Infect Dis* 2006; **193**:121–8.
17. Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, et al. Variable host-pathogen compatibility in Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* 2006;**103**:2869–73.
18. Beggs ML, Eisenach KD, Cave MD. Mapping of IS6110 insertion sites in two epidemic strains of Mycobacterium tuberculosis. *J Clin Microbiol* 2000;**38**: 2923–8.
19. Jacobs Jr WR, Kalpana GV, Cirillo JD, Pascopella L, Snapper SB, Udani RA, et al. Genetic systems for mycobacteria. *Methods Enzymol* 1991;**204**:537–55.
20. Barker LP, Brooks DM, Small PL. The identification of Mycobacterium marinum genes differentially expressed in macrophage phagosomes using promoter fusions to green fluorescent protein. *Mol Microbiol* 1998;**29**:1167–77.
21. van Soolingen D, de Haas PE, Hermans PW, van Embden JD. DNA fingerprinting of Mycobacterium tuberculosis. *Methods Enzymol* 1994;**235**:196–205.
22. Brosch R, Gordon SV, Billault A, Garnier T, Eiglmeier K, Soravito C, et al. Use of a Mycobacterium tuberculosis H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. *Infect Immun* 1998;**66**:2221–9.
23. Birnboim HC, Doly J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 1979;**7**:1513–23.
24. Staden R. The Staden sequence analysis package. *Mol Biotechnol* 1996;**5**:233–41.
25. Prod'hom G, Lagier B, Pelicic V, Hance AJ, Gicquel B, Guilhot C. A reliable amplification technique for the characterization of genomic DNA sequences flanking insertion sequences. *FEMS Microbiol Lett* 1998;**158**:75–81.
26. Mendiola MV, Martin C, Otal I, Gicquel B. Analysis of the regions responsible for IS6110 RFLP in a single Mycobacterium tuberculosis strain. *Res Microbiol* 1992;**143**:767–72.
27. Fontan P, Aris V, Ghanny S, Soteropoulos P, Smith I. Global transcriptional profile of Mycobacterium tuberculosis during THP-1 human macrophage infection. *Infect Immun* 2008;**76**:717–25.
28. Kong Y, Cave MD, Zhang L, Foxman B, Marrs CF, Bates JH, et al. Population-based study of deletions in five different genomic regions of Mycobacterium tuberculosis and possible clinical relevance of the deletions. *J Clin Microbiol* 2006;**44**:3940–6.
29. Kremer K, Glynn JR, Lillebaek T, Niemann S, Kurepina NE, Kreiswirth BN, et al. Definition of the Beijing/W lineage of Mycobacterium tuberculosis on the basis of genetic markers. *J Clin Microbiol* 2004;**42**:4040–9.
30. Mokrousov I, Ly HM, Otten T, Lan NN, Vyshnevskyi B, Hoffner S, et al. Origin and primary dispersal of the Mycobacterium tuberculosis Beijing genotype: clues from human phylogeography. *Genome Res* 2005;**15**:1357–64.
31. Malik AN, Godfrey-Faussett P. Effects of genetic variability of Mycobacterium tuberculosis strains on the presentation of disease. *Lancet Infect Dis* 2005;**5**:174–83.
32. Manca C, Tsenova L, Bergtold A, Freeman S, Tovey M, Musser JM, et al. Virulence of a Mycobacterium tuberculosis clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN-alpha /beta. *Proc Natl Acad Sci U S A* 2001;**98**:5752–7.
33. Kremer K, van-der-Werf MJ, Au BK, Anh DD, Kam KM, van-Doorn HR, et al. Vaccine-induced immunity circumvented by typical Mycobacterium tuberculosis Beijing strains. *Emerg Infect Dis* 2009;**15**:335–9.
34. Plikaytis BB, Marden JL, Crawford JT, Woodley CL, Butler WR, Shinnick TM. Multiplex PCR assay specific for the multidrug-resistant strain W of Mycobacterium tuberculosis. *J Clin Microbiol* 1994;**32**:1542–6.
35. Kurepina N, Likhoshvay E, Shashkina E, Mathema B, Kremer K, van Soolingen D, et al. Targeted hybridization of IS6110 fingerprints identifies the W-Beijing Mycobacterium tuberculosis strains among clinical isolates. *J Clin Microbiol* 2005;**43**:2148–54.
36. Dou HY, Tseng FC, Lin CW, Chang JR, Sun JR, Tsai WS, et al. Molecular epidemiology and evolutionary genetics of Mycobacterium tuberculosis in Taipei. *BMC Infect Dis* 2008;**8**:170.
37. Maeda S, Wada T, Iwamoto T, Murase Y, Mitarai S, Sugawara I, et al. Beijing family Mycobacterium tuberculosis isolated from throughout Japan: phylogeny and genetic features. *Int J Tuberc Lung Dis* 2010;**14**:1201–4.
38. Brosch R, Philipp WJ, Stavropoulos E, Colston MJ, Cole ST, Gordon SV. Genomic analysis reveals variation between Mycobacterium tuberculosis H37Rv and the attenuated M. tuberculosis H37Ra strain. *Infect Immun* 1999;**67**:5768–74.
39. Huet G, Constant P, Malaga W, Laneelle MA, Kremer K, van Soolingen D, et al. A lipid profile typifies the Beijing strains of Mycobacterium tuberculosis: identification of a mutation responsible for a modification of the structures of phthiocerol dimycocerosates and phenolic glycolipids. *J Biol Chem* 2009;**284**:27101–13.
40. Otal I, Gomez AB, Kremer K, de Haas P, Garcia MJ, Martin C, et al. Mapping of IS6110 insertion sites in Mycobacterium bovis isolates in relation to adaptation from the animal to human host. *Vet Microbiol* 2008;**129**:333–41.
41. Fomukong N, Beggs M, el Hajj H, Templeton G, Eisenach K, Cave MD. Differences in the prevalence of IS6110 insertion sites in Mycobacterium tuberculosis strains: low and high copy number of IS6110. *Tuber Lung Dis* 1997;**78**: 109–16.
42. Fang Z, Forbes KJ. A Mycobacterium tuberculosis IS6110 preferential locus (ipl) for insertion into the genome. *J Clin Microbiol* 1997;**35**:479–81.
43. Fang Z, Doig C, Kenna DT, Smittipat N, Palittapongarnpim P, Watt B, et al. IS6110-mediated deletions of wild-type chromosomes of Mycobacterium tuberculosis. *J Bacteriol* 1999;**181**:1014–20.
44. Philipp WJ, Poulet S, Eiglmeier K, Pascopella L, Balasubramanian V, Heym B, et al. An integrated map of the genome of the tubercle bacillus, Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium leprae. *Proc Natl Acad Sci U S A* 1996;**93**:3132–7.
45. Gordon SV, Heym B, Parkhill J, Barrell B, Cole ST. New insertion sequences and a novel repeated sequence in the genome of Mycobacterium tuberculosis H37Rv. *Microbiology* 1999;**145**(Pt 4):881–92.
46. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* 1998;**393**:537–44.
47. Sampson S, Warren R, Richardson M, van der Spuy G, van Helden P. IS6110 insertions in Mycobacterium tuberculosis: predominantly into coding regions. *J Clin Microbiol* 2001;**39**:3423–4.
48. Parish T, Stoker NG. The common aromatic amino acid biosynthesis pathway is essential in Mycobacterium tuberculosis. *Microbiology* 2002;**148**:3069–77.
49. Marmiesse M, Brodin P, Buchrieser C, Gutierrez C, Simoes N, Vincent V, et al. Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the Mycobacterium tuberculosis complex. *Microbiology* 2004;**150**:483–96.

Please cite this article in press as: Alonso H, et al., Deciphering the role of IS*6110* in a highly transmissible *Mycobacterium tuberculosis* Beijing strain, GC1237, Tuberculosis (2011), doi:10.1016/j.tube.2010.12.007

Journal of Clinical Microbiology

Rapid Test for Identification of a Highly Transmissible Mycobacterium tuberculosis Beijing Strain of Sub-Saharan Origin

María Isabel Millán-Lou, Henar Alonso, Patricia Gavín, Melissa Hernández-Febles, María Isolina Campos-Herrero, Rodolfo Copado, Fernando Cañas, Kristin Kremer, José Antonio Caminero, Carlos Martín and Sofía Samper

J. Clin. Microbiol. **2012, 50(2):516. DOI: 10.1128/JCM.06314-11.**
Published Ahead of Print 23 November 2011.

JCM
Journals.ASM.org

Rapid Test for Identification of a Highly Transmissible *Mycobacterium tuberculosis* Beijing Strain of Sub-Saharan Origin

María Isabel Millán-Lou,[a,b,c] Henar Alonso,[b,c] Patricia Gavín,[a,b,c] Melissa Hernández-Febles,[d] María Isolina Campos-Herrero,[d] Rodolfo Copado,[e] Fernando Cañas,[f] Kristin Kremer,[g] José Antonio Caminero,[d] Carlos Martín,[a,b,c] and Sofía Samper[a,b,c]

IIS Aragón, Hospital Universitario Miguel Servet, Zaragoza, Spain[a]; Grupo de Genética de Micobacterias, Universidad de Zaragoza, Zaragoza, Spain[b]; CIBER de Enfermedades Respiratorias, Zaragoza, Spain[c]; Hospital Universitario de Gran Canaria Dr. Negrín, Las Palmas de Gran Canaria, Spain[d]; Hospital José Molina Orosa, Lanzarote, Spain[e]; Hospital Universitario Insular de Gran Canaria, Las Palmas de Gran Canaria, Spain[f]; and National Institute for Public Health and the Environment, Bilthoven, The Netherlands[g]

The development of a rapid test to identify *Mycobacterium tuberculosis* Beijing isolates and specifically strain GC1237, coming from a sub-Saharan country, is needed due to its alarming wide spread on Gran Canaria Island (Spain). A rapid test that detects IS*6110* present between *dnaA* and *dnaN* in the Beijing strains and in a specific site for GC1237 (*Rv2180c*) has been developed. This test would be a useful tool in the surveillance of subsequent cases.

Beijing is a lineage of *Mycobacterium tuberculosis* that is dispersed worldwide predominating throughout East Asia and the former Soviet Union. This lineage may have a selective advantage either with higher virulence or transmissibility that led to clonal expansion. The prevalence of this family is low in Spain, and its presence in the population has been associated with the recent increase in the number of tuberculosis (TB) cases among immigrants (5, 9, 11, 16).

The introduction of a Beijing strain, strain GC1237, in Gran Canaria Island, Spain, by an African refugee (from Liberia), and its explosive spread in this community (27.1% of TB cases) over the next few years were reported in 2001 (6). Through review of the literature, an identical restriction fragment length polymorphism (RFLP) pattern of strain GC1237 was found in four isolates from Somalia, Ethiopia, Sierra Leone, and Liberia (6). Three Beijing isolates sharing the RFLP genotype and MIRU-VNTR were identified in Sweden from patients from Ethiopia and Eritrea (12). An *M. tuberculosis* Beijing strain responsible for an outbreak of streptomycin-resistant TB in Benin that affected at least 17 patients shared an identical mycobacterial interspersed repetitive-unit–variable-number tandem-repeat (MIRU-VNTR) pattern, except for two loci that authors failed to amplify (1). All these findings suggest that this clone originated in sub-Saharan Africa and that it is being introduced in countries with a low incidence of TB through immigration.

The purpose of this study was to develop a rapid diagnostic method based on multiplex PCR to identify the Beijing isolates and specifically the GC1237 genotype which could be applied in the laboratories located in the country of origin. The test could help in the subsequent prospective epidemiological studies to better control of the cases.

In order to carry this out, all the isolates from 2007 to 2008 collected in the three main hospitals of the Province of Las Palmas (Gran Canaria Island) were genotyped at the University of Zaragoza. One stored isolate of the original Beijing GC1237 strain was used as a control of the different techniques. First, all 292 isolates were screened by spoligotyping with a commercially available kit (Ocimum Biosolutions Ltd., Hyderabad, India) by the method of Kamerbeek et al. (14). Seventy of the 292 isolates belonged to the Beijing family. Second, IS*6110* RFLP was performed on the Beijing isolates detected by spoligotyping, if enough DNA were available (32 of the 70 isolates) according to the standardized protocol (19). Twenty-nine isolates (90.6%) had RFLP patterns similar to the pattern of GC1237, four of which presented an additional band. The other 3 isolates showed a clearly different Beijing pattern. Besides, genotyping based on 15-locus MIRU-VNTR for 3 different isolates of the GC1237 genotype was conducted as previously described (18). The three isolates showed the pattern 442333464465372 (MIRU-VNTR loci ordered according to their position on the *M. tuberculosis* H37Rv genome).

Subsequently, a multiplex PCR detecting two different locations of IS*6110* was designed for a rapid identification of isolates related to the GC1237 genotype. The first target recognizes the point of insertion of IS*6110* between *dnaA* and *dnaN*, which is a characteristic feature of all Beijing strains (15). To accomplish this, primer dnaA3 (5′-GGGCGGTTCAATTGGCTGT), was designed to anneal in the 3′ end of IS*6110* and the following sequence of the genome and used together with primer dnaA2 (5′-CCACCCACGACACCGCAT) which had been previously described (10). The second target detected a specific IS*6110* of strain GC1237. This target was chosen among the 19 copies of IS*6110* recently localized in the genome of GC1237 (3). Eleven of the insertion sites were shared with 210 strains, 2 were located in the variable PPE regions, which could present difficulties in the specificity of the PCR. The other 5 were in hot spots, detected in other non-Beijing clinical isolates or its flanking region was not presented in the reference strain H37Rv. Finally, the copy inserted into the *Rv2180c* gene was selected as the specific target for rapid diagnosis. To detect this target, primers Rv2179c-fw (fw stands for forward) (5′-CAATACGTGATCGCCGGGAC) and Rv2179c-rv (rv stands for reverse) (5′-CGAATTCCGTGGATACTGCTAG)

Received 4 November 2011 **Accepted** 13 November 2011
Published ahead of print 23 November 2011
Address correspondence to María Isabel Millán-Lou, mimillan.iacs@aragon.es.
M. I. Millán Lou and H. Alonso contributed equally to this article.

doi:10.1128/JCM.06314-11

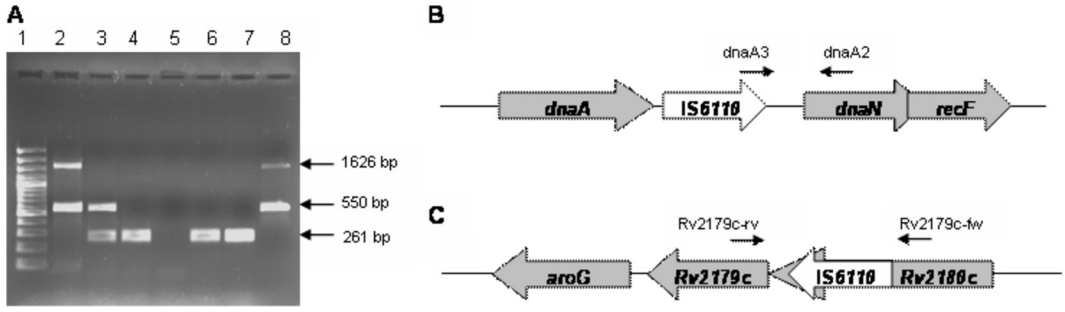


FIG 1 (A) Results of multiplex PCR visualized with agarose gel electrophoresis. Lane 1, molecular size markers (100-bp DNA ladder); lane 2, Beijing GC1237 genotype (1,626 bp and 550 bp); lane 3, a Beijing genotype different from strain GC1237 (550 bp and 261 bp); lane 4, strain H37Rv (261 bp); lane 5, negative PCR control; lanes 6, 7, and 8, different clinical isolates of *M. tuberculosis*. (B and C) Schematic representations of the positions and orientations of the two IS*6110* insertions used for the diagnostic test, the *dnaA-dnaN* region in the Beijing strains (B) and the *Rv2180c* region in strain GC1237 (C). The IS*6110* element is represented by the white arrow, with the arrowhead indicating the direction of transcription of the putative transposase. The primers are represented by the small black arrows above the schematic representations.

were designed (3). For this assay, a rapid DNA extraction was made as previously described (2). As a result of the multiplex PCR, two amplicons were expected for isolates with GC1237 genotypes of 1,626 bp and 550 bp, respectively. Two 550-bp fragments and a 261-bp fragment were expected for other Beijing isolates. A non-Beijing isolate was expected to produce only one fragment of 261 bp for the wild-type *Rv2180c* gene (Fig. 1).

To validate this technique, all 58 isolates of *M. tuberculosis* from one of the hospitals reported between 2007 and 2008 were tested. Nine isolates were found to belong to the Beijing lineage, and eight of these isolates exhibited the GC1237 pattern. Subsequently, the IS*6110* RFLP patterns of the Beijing isolates (9 isolates) were compared. The other 49 isolates presented a non-Beijing genotype. Of the isolates with a non-Beijing genotype, 8 isolates showed a high similarity to strain GC1237, and one had a different pattern, validating the diagnostic test, which showed 100% specificity (Fig. 2).

Finally, the multiplex PCR was used to identify the GC1237 genotype among 66 Beijing isolates. All the isolates amplified the specific target of Beijing family. Fifty-six presented the GC1237 genotype, and 10 presented a genotype of a different Beijing strain (Table 1). We confirmed these results by comparison with RFLP in 32 of the 66 isolates supporting the specificity of the test. A group of 50 "non-Beijing" isolates of TB patients diagnosed in the same period previously analyzed by spoligotyping were further analyzed, and all presented a non-Beijing genotype.

Due to the high incidence of this strain in the population studied, a rapid diagnostic method to discriminate non-Beijing subtypes from strain GC1237 would be useful for better control of TB on Gran Canaria island. Considering that the RFLP, spoligotyping, and MIRU methods are difficult to implement in the clinical setting, the described multiplex PCR could bring the possibility of *in situ* identification of GC1237 genotype isolates.

Until now, the incidence of this strain was retrospectively screened by spoligotyping in reference laboratories. This technique is a good method to detect those isolates of the Beijing lineage; however, information about clonality in this group was not provided. In addition, the spoligotyping method is better used for the study of 40 isolates at a time (13). RFLP analysis based on IS*6110* gives us this information, but a larger amount of DNA is required, it is time-consuming and technically demanding, and the results among laboratories are difficult to compare (19). The rapid proposed technique 24-locus MIRU-VNTR could be also a tool for detecting this genotype, but it requires 24 single PCRs or 8 multiplex PCRs. Several PCR-based methods for the identification of Beijing strains have also been developed targeting different

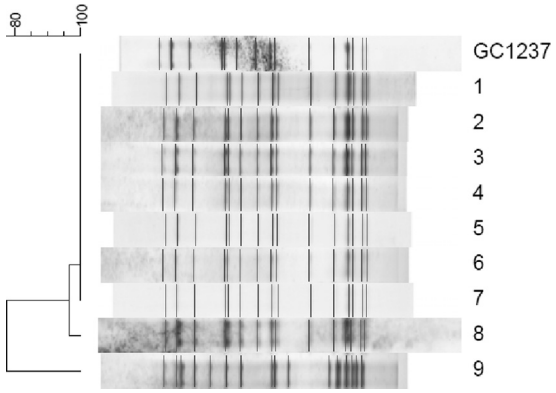


FIG 2 Dendrogram showing the IS*6110* RFLP fingerprints patterns of *M. tuberculosis* Beijing isolates used to verify the multiplex PCR technique. Isolates 1 to 8 shows high similarity to strain GC1237, and isolate 9 had a different pattern.

TABLE 1 Results of the multiplex PCR (validating and applying the method) for detection of the isolates with the GC1237 genotype

Purpose and method	No. of isolates with the following genotype:			Total no. of isolates tested by method
	Beijing GC1237	Beijing but not GC1237	Not Beijing	
Validation				
Spoligotyping	9[a]	9[a]	49	58
Multiplex PCR	8	1	49	58
RFLP	8	1		9
Test use				
Spoligotyping	66[a]	66[a]	50	116
Multiplex PCR	56	10	50	116

[a] Because spoligotyping does not discriminate between the Beijing GC1237 genotype and Beijing but not GC1237 genotype, these data correspond to the same isolates.

MIRU loci, specific deletions, or single-nucleotide polymorphisms (SNPs). However, most of these methods for identification are not specific, expensive, or difficult to perform (4, 7, 8, 17).

The results obtained in this work show that our multiplex PCR has demonstrated the specificity to identify the GC1237 genotype isolates. In contrast to spoligotyping and RFLP, this multiplex PCR can be used to identify one or more positive isolates. It is a quick, straightforward, and inexpensive technique. It can be completed in 1 day, and it does not require hybridization to a panel of species-specific probes, as RFLP does. Moreover, sample preparation is simple (simple heating of the bacteria), and trained personnel can interpret the gels with minimal difficulty. In contrast to 24-locus MIRU-VNTR, a single reaction is needed. It can be used by laboratories lacking a real-time PCR machine and easily implemented for a large-scale GC1237 strain screening. This designed PCR would allow its use in the clinical setting and facilitate the detection of new cases aiding a more detailed supervision of patients and discrimination from cases infected with other Beijing genotypes.

ACKNOWLEDGMENTS

This study was partially funded by the Fondo de Investigaciones Sanitarias (FIS09/0051), Instituto de Salud Carlos III (CM09/000123), and Gobierno de Aragón (B034/09) in Spain.

We thank Dessislava Marinova for her kind help revising the English and for valuable comments on the manuscript.

REFERENCES

1. **Affolabi D, et al.** 2009. Possible outbreak of streptomycin-resistant *Mycobacterium tuberculosis* Beijing in Benin. Emerg. Infect. Dis. **15**:1123–1125.
2. **Aldous WK, Pounder JI, Cloud JL, Woods GL.** 2005. Comparison of six methods of extracting *Mycobacterium tuberculosis* DNA from processed sputum for testing by quantitative real-time PCR. J. Clin. Microbiol. **43**: 2471–2473.
3. **Alonso H, et al.** 2011. Deciphering the role of IS*6110* in a highly transmissible *Mycobacterium tuberculosis* Beijing strain, GC1237. Tuberculosis **91**:117–126.
4. **Alonso M, et al.** 2011. A novel method for the rapid and prospective identification of Beijing *Mycobacterium tuberculosis* strains by high-resolution melting analysis. Clin. Microbiol. Infect. **17**:349–357.
5. **Alonso M, et al.** 2010. Characterization of *Mycobacterium tuberculosis* Beijing isolates from the Mediterranean area. BMC Microbiol. **10**:151.
6. **Caminero JA, et al.** 2001. Epidemiological evidence of the spread of a *Mycobacterium tuberculosis* strain of the Beijing genotype on Gran Canaria Island. Am. J. Respir. Crit. Care Med. **164**:1165–1170.
7. **Chen J, et al.** 2007. Deletion-targeted multiplex PCR (DTM-PCR) for identification of Beijing/W genotypes of *Mycobacterium tuberculosis.* Tuberculosis **87**:446–449.
8. **Chin PJ, Chiu CC, Jou R.** 2007. Identification of Beijing lineage *Mycobacterium tuberculosis* with combined mycobacterial interspersed repetitive unit loci 26, 31, and ETR-A. J. Clin. Microbiol. **45**:1022–1023.
9. **European Concerted Action on New Generation Genetic Markers and Techniques for the Epidemiology and Control of Tuberculosis.** 2006. Beijing/W genotype *Mycobacterium tuberculosis* and drug resistance. Emerg. Infect. Dis. **12**:736–743.
10. **Fennelly KP, et al.** 2004. Cough-generated aerosols of *Mycobacterium tuberculosis*: a new method to study infectiousness. Am. J. Respir. Crit. Care Med. **169**:604–609.
11. **Garcia de Viedma D, Chaves F, Iñigo J.** 2006. New route of importation of *Mycobacterium tuberculosis* Beijing genotype. Emerg. Infect. Dis. **12**: 169–170.
12. **Ghebremichael S, et al.** 2010. Drug resistant *Mycobacterium tuberculosis* of the Beijing genotype does not spread in Sweden. PLoS One **5**:e10893.
13. **Hillemann D, Warren R, Kubica T, Rüsch-Gerdes S, Niemann S.** 2006. Rapid detection of *Mycobacterium tuberculosis* Beijing genotype strains by real-time PCR. J. Clin. Microbiol. **44**:302–306.
14. **Kamerbeek J, et al.** 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J. Clin. Microbiol. **35**:907–914.
15. **Kremer K, et al.** 2004. Definition of the Beijing/W lineage of *Mycobacterium tuberculosis* on the basis of genetic markers. J. Clin. Microbiol. **42**: 4040–4049.
16. **Samper S, et al.** 2005. Systematic molecular characterization of multidrug-resistant *Mycobacterium tuberculosis* complex isolates from Spain. J. Clin. Microbiol. **43**:1220–1227.
17. **Sun JR, Lee SY, Dou HY, Lu JJ.** 2009. Using a multiplex polymerase chain reaction for the identification of Beijing strains of *Mycobacterium tuberculosis.* Eur. J. Clin. Microbiol. Infect. Dis. **28**:105–107.
18. **Supply P, et al.** 2006. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis.* J. Clin. Microbiol. **44**:4498–4510.
19. **van Embden JD, et al.** 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J. Clin. Microbiol. **31**:406–409.