

**UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS**

**FACULTAD DE CIENCIAS BIOLÓGICAS**

**UNIDAD DE POSGRADO**

**Firma genómica para genotipificación de**  
*Mycobacterium tuberculosis*

**TESIS**

Para optar el Grado Académico de Magíster en Biología  
Molecular

**AUTOR**

Luis José JARAMILLO VALVERDE

**ASESOR**

Ruth GARCÍA DE LA GUARDA

Lima – Perú

2017

A mis padres, Hilda y José Luis,  
que me dieron la vida y han estado  
conmigo en todo momento, por su  
amor, comprensión y entrega.

A mis queridas hermanas  
Betty e Hilda, por su amistad,  
apoyo, compañía y amor.

## AGRADECIMIENTOS

Quiero agradecer a mis padres, por educarme, enseñarme, amarme y por el apoyo en el desarrollo de la presente tesis. A mis hermanas, por confiar en mí, por su cariño y amistad. A mi abuelita Mila que desde el cielo, sé que intercede por mí ante Dios. A mis tíos y primos que siempre estuvieron conmigo demostrándome su cariño y apoyo.

Al Ph.D. Heinner Guio Chunga, investigador del Laboratorio de Biotecnología y Biología Molecular (LBBM) del Instituto Nacional de Salud (INS), por su generosidad al brindarme la oportunidad de ser parte del equipo que lleva a cargo diversos proyectos de investigación en desarrollo, por el asesoramiento temático en el campo de la tuberculosis y su experiencia científica en un marco de confianza, consejos, afecto y amistad.

Un agradecimiento muy especial a David Tarazona Jurado, por confiar plenamente en mi capacidad profesional, brindándome todo su apoyo y encaminándome en el campo de la bioinformática, por su amistad, dedicación, por su sincero y desinteresado apoyo en la elaboración de esta investigación y por sus conocimientos en busca del mejoramiento y desarrollo de la misma.

A mi asesora, Mag. Ruth García de la Guarda, por sus valiosas sugerencias, críticas y acertados aportes durante la revisión del proyecto, avances e informe final de tesis, además de consejos que me brindó desde los inicios de mi camino profesional.

A mis amigos: Marco Galarza, Kelly Lévano, Víctor Borda a quienes tuve la suerte de conocer y que apoyaron poniendo su valioso grano de arena durante la realización de la presente investigación. A Rose, quien con mucho cariño y amor me motivó para seguir adelante.

A todos aquellos profesores y amigos que de una u otra manera me dieron consejos y ánimos para seguir adelante. Muchas gracias a ellos también.

# CONTENIDO

	Pág.
ABREVIATURAS .....	I
ÍNDICE DE FIGURAS.....	III
ÍNDICE DE TABLAS.....	III
RESUMEN.....	IV
ABSTRACT .....	V
<b>1. INTRODUCCIÓN</b>	<b>1</b>
<b>2. ANTECEDENTES</b>	<b>6</b>
2.1. Patogenia de la tuberculosis .....	6
2.2. Clasificación taxonómica de MTB .....	7
2.3. Aspectos biológicos y estructurales de las micobacterias .....	7
2.4. Diagnóstico de la infección tuberculosa .....	11
2.5. Aspectos epidemiológicos de la tuberculosis .....	14
2.6. Epidemiología molecular de <i>Mycobacterium tuberculosis</i> .....	16
2.6.1. RFLP-IS6110 .....	18
2.6.2. <i>Spoligotyping</i> .....	19
2.6.3. MIRU-VNTR.....	21
2.6.4. SNP.....	23
2.7. Secuenciamiento de ADN.....	25
2.8. Firma genómica para genotipicar MTB.....	26
<b>3. HIPÓTESIS Y OBJETIVOS</b>	<b>28</b>
3.1 Hipótesis .....	28
3.2 Objetivos.....	28
3.2.1 Objetivo General .....	28
3.2.2 Objetivos Específicos .....	28
<b>4. MATERIALES Y MÉTODOS</b>	<b>29</b>
4.1. Selección de secuencias genómicas.....	29
4.2. Análisis filogenético de secuencias genómicas .....	30

4.3. Análisis de clúster de grupos ortólogos (COG) .....	31
4.4. Selección de grupo de SNPs .....	31
<b>5. RESULTADOS</b>	<b>33</b>
5.1. Determinación de genotipo por análisis filogenético de 48 genomas de MTB desconocidos.....	33
5.2. Identificación y selección del conjunto de SNPs.....	33
5.3. Distribución de SNPS en genotipos de <i>Mycobacterium tuberculosis</i> .....	36
5.4. Validación de la asignación de genotipos mediante el nuevo conjunto de 10 SNPs propuesto .....	40
<b>6. DISCUSIÓN</b>	<b>42</b>
<b>7. CONCLUSIONES</b>	<b>49</b>
<b>8. RECOMENDACIONES</b>	<b>50</b>
<b>9. REFERENCIAS BIBLIOGRÁFICAS</b>	<b>51</b>
<b>10. ANEXOS</b>	<b>77</b>
ANEXO I: 249 secuencias genómicas de <i>Mycobacterium tuberculosis</i> analizadas en este estudio.....	78
ANEXO II: Base de datos de SNPs, que incluye SNPs correspondientes a genes y grupos COG por cada genotipo de MTB.....	86
<b>11. PRODUCTIVIDAD CIENTÍFICA</b>	<b>105</b>

## ABREVIATURAS

ADN	Acido desoxirribonucleico
APP	Proporciones agar en placa
ARN	Ácido ribonucleico
BAAR	Bacilo alcohol ácido resistente
COG	Clúster de grupos ortólogos
CRISPR	Repeticiones palindrómicas cortas agrupadas reguladoras
DR	Repetición directa
IGR	Regiones intergénicas
IS	Secuencia de inserción
IE	Principio inclusión/exclusión
LJ	Löwenstein-Jensen
LPS	Polimorfismos de secuencias largas
MDR	Multidrogo resistente
MINSA	Ministerio de Salud
MIRU	Unidades repetitivas interespaciadas de micobacterias
MODS	Ensayo de susceptibilidad a drogas por observación microscópica
MTB	<i>Mycobacterium tuberculosis</i>
MUM	Encuentros únicos máximos
NGS	Secuenciamiento de próxima generación
nsSNP	Polimorfismos de un solo nucleótido no sinónimos
OK	Ogawa Kudoh
OMS	Organización Mundial de la Salud
OPS	Organización Panamericana de Salud
PRA	Análisis de restricción PCR
PCR	Reacción en cadena de la polimerasa
pb	Pares de bases

RFLP	Polimorfismo de longitud de fragmentos de restricción
SGC	Secuenciamiento de genomas completos
SIDA	Síndrome de inmunodeficiencia adquirida
SIT	Tipo Internacional de Espoligotipaje
SNP	Polimorfismo de un solo nucleótido
<i>Spoligotyping</i>	Tipificación de oligonucleótido espaciador
SSR	Repeticiones de secuencias cortas
VIH	Virus de la inmunodeficiencia humana
VNTR	Número variable de repeticiones en tándem
XDR	Extremadamente resistente
°C	Grados celsius
T°	Temperatura
µm	Micrometros

## ÍNDICE DE FIGURAS

	Pág.
<b>Figura 1:</b> Esquema de la pared celular de las micobacterias .....	9
<b>Figura 2:</b> Ubicación de marcadores moleculares en el cromosoma de <i>M. tuberculosis</i> para genotipificación: RFLP-IS6110, <i>Spoligotyping</i> y MIRU-VNTR.....	17
<b>Figura 3:</b> DRs y espaciadores utilizados en <i>Spoligotyping</i> .....	20
<b>Figura 4:</b> Posición de 41 <i>loci</i> MIRU en el cromosoma de <i>M. tuberculosis</i> H37Rv.....	22
<b>Figura 5:</b> Construcción de contigs y scaffolds .....	26
<b>Figura 6:</b> Diversidad Global de 249 secuencias analizadas de <i>M. tuberculosis</i> para este estudio .....	30
<b>Figura 7:</b> Árbol filogenético de Máxima parsimonia de 249 cepas de <i>M. tuberculosis</i> basados en SNPs .....	34
<b>Figura 8:</b> Diagrama de SNPs específicos para genotipos de <i>M. tuberculosis</i> : Beijing, CAS, EAI, Haarlem, LAM, TX, Ural, AFRI1 y AFRI2.....	35
<b>Figura 9:</b> Distribución relativa de SNPs de <i>M. tuberculosis</i> pertenecientes a ciertos Clúster de grupos Ortólogos (COG) .....	37

## ÍNDICE DE TABLAS

	Pág.
<b>Tabla 1:</b> Clasificación general de los genes de <i>Mycobacterium tuberculosis</i> .....	11
<b>Tabla 2:</b> Pasos para la selección de SNPs.....	38
<b>Tabla 3:</b> Conjunto de 10 SNPs propuestos para genotipificación de MTB.....	39
<b>Tabla 4:</b> Asignación del genotipo en 34 genomas de MTB utilizando el nuevo conjunto de 10 SNPs propuesto .....	41



## RESUMEN

*Mycobacterium tuberculosis* (MTB), agente causal de la tuberculosis, tiene una gran diversidad de genotipos entre ellos Beijing, CAS, EAI, Haarlem, LAM, X, Ural, T, AFRI1 y AFRI2. Sin embargo, la genotipificación puede ser costosa, consumir tiempo y en algunos casos, los resultados pueden variar dependiendo de la metodología utilizada. En este sentido, se propone un nuevo conjunto de 10 polimorfismos de un solo nucleótido (SNPs) utilizando un total de 249 genomas de MTB, que por primera vez se seleccionaron basándose en el principio de inclusión/exclusión (IE) utilizando datos de *Spoligotyping* y filogenia, seguido de la selección de SNPs no sinónimos presentes en clúster de grupos ortólogos (COG) más conservados de cada genotipo de MTB. La asignación del genotipo utilizando el nuevo conjunto de 10 SNPs fue validado con 34 genomas de MTB adicionales y los resultados mostraron una correlación del 100% con sus genotipos ya conocidos. Nuestro conjunto de 10 SNPs no ha sido reportado previamente y abarca los genotipos de MTB que son más prevalentes en todo el mundo. Este conjunto de SNPs proporciona una resolución superior a las técnicas actuales y podría ser utilizado en epidemiología molecular con marcadores resistentes a fármacos, determinando los factores de riesgo en la transmisión de la tuberculosis en diferentes poblaciones para desarrollar estrategias para el control de esta enfermedad.

**Palabras clave:** COG, genotipo, inclusión/exclusión, *Mycobacterium tuberculosis*, SNPs.

## ABSTRACT

*Mycobacterium tuberculosis* (MTB), causative agent of tuberculosis, has a vast diversity of genotypes including Beijing, CAS, EAI, Haarlem, LAM, X, Ural, T, AFRI1 and AFRI2. However, genotyping can be expensive, time consuming and in some cases, results may vary depending on methodology used. Here, we proposed a new set of 10 SNPs using a total of 249 MTB genomes, and selected by first the inclusion/ exclusion (IE) criteria using *Spoligotyping* and phylogenies, followed by the selection of the nonsynonymous SNPs present in the most conserved cluster of orthologous groups (COG) of each genotype of MTB. Genotype assignment of the new set of 10 SNPs was validated using an additional of 34 MTB genomes and results showed 100% correlation with their known genotypes. Our set of 10 SNPs have not been previously reported and cover the MTB genotypes that are more prevalent worldwide. This set of SNPs provides a higher resolution than current techniques and could be used in molecular epidemiology with drug resistant markers, determining the risk factors in the transmission of tuberculosis in different populations to develop strategies for the control of this disease.

**Key words:** COG, genotype, inclusion/exclusion, *Mycobacterium tuberculosis*, SNPs.

# 1. INTRODUCCIÓN

*Mycobacterium tuberculosis* (MTB), agente causal de la tuberculosis, representa una de las bacterias con mayor importancia epidemiológica en salud pública y que ha coevolucionado con nosotros desde nuestros orígenes comunes en África (Comas *et al.*, 2012). A pesar de los esfuerzos realizados por la Organización Mundial de la Salud (OMS), en los últimos años la tuberculosis continúa siendo el mayor problema global de salud pública en el mundo y es la segunda causa de muerte debida a un agente infeccioso después del Virus de Inmunodeficiencia Humana (VIH) (WHO, 2014).

De acuerdo a estimativos de la OMS, en el año 2014 se presentaron 9.6 millones de casos nuevos de TB en el mundo, con una incidencia mundial de 133 casos por 100 000 habitantes, de los cuales 1.2 millones (12%) eran VIH positivos; se presentaron 1.5 millones de muertes, de las cuales 390 000 ocurrieron en individuos VIH positivos; 480 000 casos de tuberculosis multidrogorresistente (MDR); sin embargo, solo fueron reportados 123 000 (WHO, 2015). Se estima que un 9.7% de las personas con tuberculosis MDR tiene tuberculosis extensivamente resistente (XDR) (WHO, 2015).

En la región de las Américas se estimaron 280 000 casos incidentes de tuberculosis, de los cuales 36 000 eran pacientes VIH positivos, 4 000 casos nuevos de tuberculosis MDR y solo 3 000 de ellos fueron tratados (WHO, 2015).

La aparición de cepas multidrogorresistentes y extremadamente resistentes a drogas (XDR), además del aumento de la incidencia del virus de la inmunodeficiencia humana (VIH) han contribuido a la difusión y a la reactivación de la enfermedad. Las cepas MDR son aquellas que presentan resistencia a rifampicina e isoniacida, dos de los principales medicamentos de primera línea utilizados para el control de la tuberculosis. El surgimiento de esta clase de cepas dirige potencialmente a la tuberculosis a ser una de las enfermedades incurables infecciosas más importantes (Comas *et al.*, 2011). Esto indudablemente se contrapone a los lineamientos de la OMS, que buscaba reducir la tasa de

prevalencia y mortalidad causada por TB hasta un 50% en el 2015 y su total eliminación como un problema de salud pública para el 2050 (WHO, 2010).

Además de la resistencia a medicamentos y la co-infección con el VIH, MTB ha desarrollado estrategias que hacen a la tuberculosis difícil de controlar. Por ejemplo, ha tenido gran éxito al evadir la respuesta inmune del hospedero, permitiéndole sobrevivir en el interior del fagosoma (Bai *et al.*, 2013) y es capaz de permanecer en el interior de los tejidos, en un estado de ninguna o muy baja replicación durante la vida de su hospedero. En un gran número de casos, la tuberculosis aparece como una reactivación de los bacilos en este estado de latencia (Rodríguez *et al.*, 2013).

Existe una alta diversidad genética de cepas de *Mycobacterium tuberculosis* a nivel mundial que causan la enfermedad, se han descrito como prevalentes en el mundo los genotipos: Haarlem (H), Latino América-Mediterráneo (LAM), T, Este de África-India (EAI), Central-Asia (CAS), X y Beijing. Cada nombre está atribuido al lugar de procedencia donde se identificó inicialmente.

Los estudios moleculares de MTB permiten la investigación de la epidemiología de tuberculosis, revelando perspectivas sobre la diferencias en el éxito de transmisión de las cepas; mientras que la observación y el análisis de la epidemiología puede comprobar hipótesis acerca de la biología de la cepa (Demay *et al.*, 2012). La información de la diversidad de MTB colectada en bases de datos mundiales proporciona una plataforma sólida para la investigación en filogenia y virulencia (Streicher *et al.*, 2007).

Desde el descubrimiento de polimorfismos de ADN en MTB, la genotipificación molecular de las cepas tiene un valor incalculable para el estudio de la epidemiología de la tuberculosis (Murray *et al.*, 2002). Los estudios epidemiológicos de tuberculosis en todo el mundo (Candia *et al.* 2007; Sharma *et al.* 2008; Valcheva *et al.*, 2008), se han basado en el supuesto de que los pacientes con cepas genotípicamente agrupadas están epidemiológicamente vinculadas y representan transmisiones recientes. En contraste, los pacientes infectados con diferentes tipos de cepas no se consideran indicativos de transmisión reciente. Por lo tanto, el uso

de marcadores genéticos polimórficos estables permitiría distinguir cepas no relacionadas; así como, identificar aislados de las mismas cepas. Además, estos métodos de caracterización empleados deben ser reproducibles, discriminatorios y fáciles de realizar.

La decodificación del genoma de *Mycobacterium tuberculosis* (Cole *et al.*, 1998) ha acelerado los avances en la diversidad genética y diagnóstico de la tuberculosis. Los genotipos de MTB se han establecido a través de la metodología *Spoligotyping* (Brudey *et al.*, 2006). La técnica se basa en la amplificación por PCR de repeticiones palindrómicas cortas agrupadas reguladoras (CRISPR) con cebadores marcados que reconocen las secuencias repetidas, seguido de una hibridación de los productos de PCR a una membrana que contiene sondas, patrones de hibridación diferencial permiten la separación de diferentes cepas (Streicher *et al.*, 2006). La diferencia entre las cepas de MTB se reporta debido a los espaciadores de CRISPR (van Embden *et al.*, 2000).

Se han propuesto varias técnicas además, basadas en la reacción en cadena de la polimerasa (PCR) para genotipificar cepas de MTB. El Polimorfismo de longitud de fragmentos de restricción (RFLP) – IS6110 es la técnica de referencia y la más utilizada en todo el mundo (van Embden *et al.*, 1993) debido a su alta capacidad de discriminación (Kremer *et al.*, 1999). Un enfoque más reciente es el análisis de múltiples *loci* que contienen un número variable de repeticiones en tándem (VNTR) de unidades repetitivas intercaladas de micobacterias (MIRUS), la cual se basa en el polimorfismo de 24 MIRU-VNTR *loci* y permite un análisis fácil de los datos que se pueden comparar con la base de datos mundial *SpolDB 4.0* (van Embden *et al.*, 1993; Kremer *et al.*, 1999; Kanduma *et al.*, 2003).

Recientemente, dos estudios han revelado que el uso de SNPs (polimorfismos de nucleótido simple) son marcadores filogenéticos altamente valorables (Filliol *et al.*, 2006; Gutacker *et al.*, 2006). Por lo tanto, los SNPs pueden ser marcadores útiles tanto para la identificación de especies del complejo MTB y la determinación de genotipos de aislados de *Mycobacterium tuberculosis*.

En la actualidad, las regiones que tienen una alta frecuencia de infección de tuberculosis tienen un particular genotipo de MTB (Mokrousov, 2008). Sin embargo, en países con una población heterogénea, los migrantes de diversas partes del mundo, como Estados Unidos (Dasgupta y Menzies, 2005), Perú (Wamoto *et al.*, 2012), Israel (Goldblatt *et al.*, 2014), países europeos (Odone *et al.*, 2014; Zammarchi *et al.*, 2014) e incluso China (Dong *et al.*, 2010) han informado de una mayor diversidad de familias de MTB. Esto es lo que hace compleja la distribución, migración, evolución y distribución de genotipos.

Los estudios de genotipificación en MTB han permitido determinar su estructura poblacional y evolución mediante la cuantificación de las asociaciones entre el genotipo de la cepa con fenotipos de resistencia a fármacos, además de determinar el genotipo prevalente y otros factores epidemiológicos incluyendo el origen geográfico. Esto mejora el desarrollo de estrategias para el control de la enfermedad, ya que contribuye al conocimiento de las rutas de transmisión, diferencia en el caso de recaídas si se trata de una reinfección exógena o una reactivación endógena, también diferencia transmisión reciente de reactivación e identifica la presencia de brotes.

En los últimos años, los estudios de epidemiología molecular y la genotipificación de las cepas de MTB, han basado sus esfuerzos en las técnicas de secuenciación de genomas completos (SGC) (Kato-Maeda *et al.*, 2013). Las tecnologías de secuenciación permiten identificar la dinámica de transmisión del patógeno, la resistencia a medicamentos, los factores de virulencia y determinar la evolución del genoma (Roetzer *et al.*, 2013).

La completa información genética que otorga el SGC (Roetzer *et al.*, 2013) junto con la disposición de la secuencia completa del genoma de MTB y de programas para la búsqueda de similitud entre secuencias, puede incrementar el conocimiento de la interacción patógeno-hospedero permitiendo la identificación de nuevos blancos de vacunas (Bai *et al.*, 2013; Reyes *et al.*, 2013) así como una diferenciación entre cepas que conduzca a reconocer el origen y los patrones de dispersión del patógeno (Ragheb *et al.*, 2013).

Se ha demostrado que SGC proporciona una resolución superior a la de MIRU-VNTR y *Spoligotyping* proporcionando todos los posibles objetivos genómicos a la vez, como la información de resistencia a fármacos, evolución del genoma y determinantes de virulencia. Sin embargo, los costos masivos de secuenciación siguen siendo caros en países con alta incidencia de TB, por lo tanto, el mejor enfoque es analizar un conjunto específico de SNPs asociados a un determinado genotipo. El objetivo de este estudio es proponer un nuevo conjunto específico de SNPs, basados en el principio de inclusión/exclusión, para la genotipificación de *Mycobacterium tuberculosis*, el cual será útil para clasificar las cepas MTB (genotipo) y determinar sus factores de riesgo en la transmisión de la tuberculosis en diferentes poblaciones desarrollando estrategias para el control de esta enfermedad.

## 2. ANTECEDENTES

### 2.1. Patogenia de la tuberculosis

La tuberculosis es una enfermedad sistémica, infecciosa y crónica que afecta mayormente al sistema respiratorio. Producida principalmente por especies del complejo MTB: *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Mycobacterium africanum* y *Mycobacterium microtti*, que pueden afectar a humanos (Brosch *et al.*, 2002). MTB fue identificado como el agente causal de la tuberculosis en 1882 por Robert Koch, quien logró aislarlo y cultivarlo a partir de pacientes con tuberculosis (Fenton y Vermeulen, 1996). La infección se adquiere habitualmente en etapas tempranas de la vida, condicionando a una infección aguda frecuentemente asintomática con desarrollo de inmunidad, formación de granuloma y calcificación, denominándose a esto tuberculosis latente (Daniel *et al.*, 1994). La tuberculosis afecta principalmente los sistemas respiratorio, digestivo, linfático y renal, atacando al estado general, de manera que si no es tratada oportuna y eficientemente puede llevar a la muerte a quien la padece. La tuberculosis pulmonar es la forma más frecuente e infectante; así como, de mayor importancia epidemiológica, por lo que es considerada como un problema de salud pública en nuestro país (Van Soolingen, 2001).

En la mayoría de los casos el mayor riesgo de infección con *Mycobacterium tuberculosis* depende de factores exógenos, tales como: (i) la epidemiología de TB en una región determinada, (ii) el ambiente social y (iii) la calidad en los sistemas de salud. Sin embargo, la probabilidad de desarrollar enfermedad activa es principalmente determinada por factores endógenos tales como: el estado nutricional, la salud general e inmunocompetencia que son determinantes críticos para el desarrollo de la infección. La mayoría de personas infectadas (aprox. 90%) no desarrollarán TB activa a lo largo de su vida y los bacilos se confinan en un microambiente inmunológico llamado "Granuloma". Sin embargo, pueden sobrevivir y persistir durante toda la vida del hospedero en un "estado de dormancia", lo que plantea una amenaza continua para el individuo infectado. Cualquier compromiso que afecte el sistema inmunológico tales como terapia inmunosupresora, infección por VIH, o edad avanzada, reaniman los bacilos latentes, causando la reactivación



de la tuberculosis, la cual es clínicamente indistinguible de una infección primaria (Kaufmann *et al.*, 2009).

La tuberculosis primaria es encontrada predominantemente en niños menores de 4 años, que parecen tener un déficit en el control de *Mycobacterium tuberculosis* y la reactivación de tuberculosis es encontrada en adultos que adquirieron el bacilo muchos años antes. Esta forma de tuberculosis es localizada en más del 80% en el pulmón, aunque otros órganos pueden también ser afectados (Kaufmann *et al.*, 2009).

## 2.2. Clasificación taxonómica de MTB

Dominio: Bacteria

Phylum: Actinobacteria

Clase: Actinobacteria

Subclase: Actinobacteridae

Orden: Actinomycetales

Suborden: Corynebacterineae

Familia: Mycobacteriaceae

Género: Mycobacterium

Especie: *Mycobacterium tuberculosis*

*Mycobacterium tuberculosis*. **Sanger Institute** (2007)

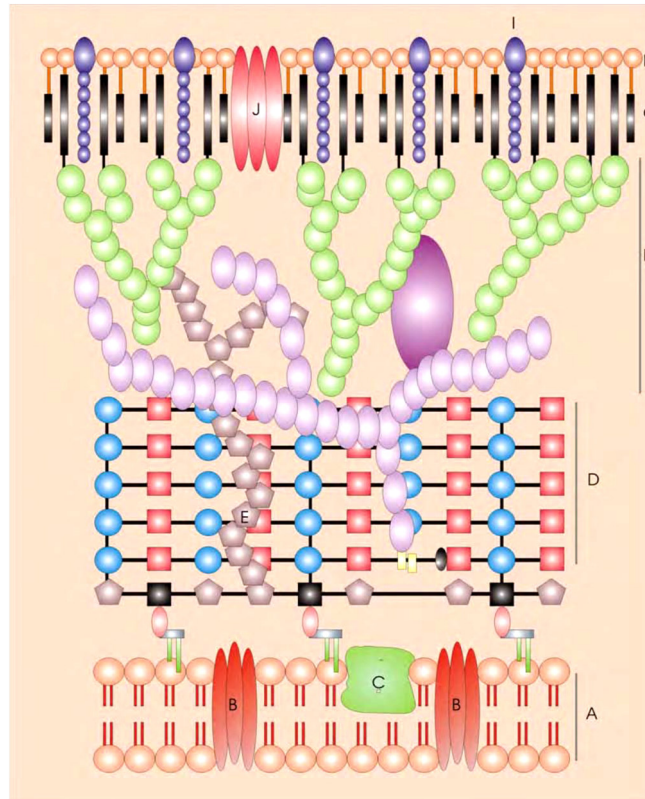
## 2.3. Aspectos biológicos y estructurales de las Micobacterias

Las micobacterias son bacilos largos de 3 a 5  $\mu\text{m}$  de longitud o curvos en forma de maza, inmóviles, no esporulados, con abundantes gránulos citoplasmáticos, resistentes a la tinción con colorantes comunes, resistentes a la decoloración con una mezcla de alcohol-ácido, algunos son aerobios, con crecimiento rápido y lento. Su estructura es rica en lípidos (20-60 %), oportunistas y saprofitas (Rodríguez, 2006).

MTB y otras micobacterias son biológicamente similares a las bacterias Gram+, aunque frente a la tinción de Gram las micobacterias son débilmente Gram+ o no se tiñen. Su membrana consiste en dos partes principales: la membrana plasmática y, alrededor de ella, la pared celular. La primera le otorga a la célula protección osmótica, además de transporte de iones y moléculas; en tanto que la segunda le brinda soporte mecánico y protección. La pared celular está constituida por tres capas: la capa interna, compuesta por peptidoglicano como en otras bacterias; la capa media, compuesta por el polisacárido arabinogalactano, cuyos extremos están esterificados con ácidos grasos de alto peso molecular, los ácidos micólicos, únicos para las micobacterias; y la capa exterior compuesta por una estructura glucolípida. La membrana plasmática tiene como característica la presencia de moléculas lipopolisacáridos, lipoarabinomananos, lipomananos y fosfatidil-inositol-manósidos (**Figura 1**) (Hong y Hopfinger, 2004).

Son bacilos ácido-alcohol resistentes por lo cual se tiñen mediante la coloración Ziehl-Neelsen, método basado en la coloración con fucsina y decoloración con HCl (Kasper *et al.*, 2005).

La mayoría de las micobacterias patógenas tienen un crecimiento lento, después de 3 a 4 semanas se pueden observar las colonias (Van Embden *et al.*, 1993). Las micobacterias son capaces de sobrevivir por largos periodos, siempre y cuando se encuentren aisladas de la luz solar ya que la luz ultravioleta y una temperatura mayor a 65 °C las inactiva, de igual manera poseen una mayor resistencia a los ácidos, álcalis y desinfectantes que el resto de las bacterias, además son resistentes a la deshidratación y la congelación (Dorransoro y Torroba, 2007).



**Figura 1:** Esquema de la pared celular de las micobacterias (Hong y Hopfinger, 2004).  
 A: Membrana plasmática, B: Proteína transversal, C: Proteína no transversal,  
 D: Peptidoglucano, E: Lipoarabinomano, F: Arabinogalactano, G: Ácidos micólicos,  
 H: Glicolípidos, I: Lípidos libres, J: Porina.

Aun con muchas características en común, las especies del género *Mycobacterium* muestran diversidad en otros aspectos. La mayoría vive y replica libremente en los ecosistemas y rara vez o nunca causa enfermedad. Solo unas pocas micobacterias pueden volverse patógenos exitosos en vertebrados superiores. En contraste, algunas especies de micobacterias no pueden crecer en el medio ambiente, son hospedador dependientes como *Mycobacterium leprae*, *Mycobacterium lepraemurium*, *Mycobacterium avium* subsp. *Paratuberculosis*, y los miembros del complejo *Mycobacterium tuberculosis*. Este complejo incluye MTB, junto con otras 7 especies estrechamente relacionadas: *Mycobacterium bovis*, *Mycobacterium africanum*, *Mycobacterium microti*, *Mycobacterium caprae*, *Mycobacterium pinnipedii*, *Mycobacterium canetti* y *Mycobacterium mungi* (Monteserin, 2016).

El genoma completo de MTB fue descrito por el Instituto Sanger en el año 1998. Comprende alrededor de 4 411 532 pb, que conforman aproximadamente 4 000 genes distribuidos equitativamente en el ADN; se conocen las funciones precisas o por lo menos supuestas en el 52 % de estos genes, para el 48 % restante se desconocen sus funciones. Entre el 3 y 4 % del genoma de la bacteria está conformado por secuencias de inserción (IS), siendo IS6110 la más empleada por su capacidad polimórfica para la clasificación de MTB (Cole, 2002).

*Mycobacterium tuberculosis* es considerado genéticamente monomórfico, porque tiene bajos niveles de diversidad genética y homoplasia (eventos de mutaciones independientes que resultan en el mismo genotipo entre aislamientos con diferentes ancestros) y muy raros eventos de recombinación homóloga. Sin embargo, en comparación con otras bacterias monomórficas, presenta sustancial variación en su ADN. Los polimorfismos identificados como *large sequence polymorphisms* (LPSs) y *single nucleotide polymorphisms* (SNPs) son filogenéticamente informativos y útiles para los estudios de población (Kato-Maeda *et al.*, 2011).

Además, el cromosoma posee otras regiones que son fuentes de variabilidad: los polimorfismos en las repeticiones cortas palindrómicas agrupadas regularmente espaciadas (*clustered regularly interspaced short palindromic repeats*, CRISPRs) y repeticiones en tándem de número variable (*variable number tandem repeats*, VNTR). Todos ellos han sido aplicados a estudios de epidemiología molecular (Kato-Maeda *et al.*, 2011).

Cerca de 200 genes del genoma de *Mycobacterium tuberculosis* codifican para proteínas enzimáticas del metabolismo de ácidos grasos, aproximadamente el 6% del total del genoma (**Tabla 1**). 100 de ellos están fuertemente implicados en la  $\beta$ -oxidación de ácidos grasos y tienen un alto contenido de guanina-citosina (G-C), el cual es constante a lo largo de todo su genoma (Smith, 2003).

**Tabla 1.** Clasificación general de los genes de *Mycobacterium tuberculosis*.

<b>Función</b>	<b>No. Genes</b>	<b>% Total</b>	<b>%de Capacidad Codificante</b>
<b>Metabolismo de lípidos</b>	225	5.7	9.3
<b>Vías de señales</b>	207	5.2	6.1
<b>Procesos celulares y pared celular</b>	517	13.0	15.5
<b>Estabilidad de ARNs</b>	50	1.3	0.2
<b>Elementos IS y bacteriófagos</b>	137	3.4	2.5
<b>Proteínas PE* y PPE **</b>	167	4.2	7.1
<b>Metabolismo intermediario y Resp.</b>	877	22.0	24.6
<b>Proteínas reguladoras</b>	188	4.7	4.0
<b>Virulencia, detoxificación</b>	91	2.3	2.4
<b>Funciones hipotéticamente conservadas</b>	911	22.9	18.4
<b>Proteínas de función desconocida</b>	607	15.3	9.9

\*, \*\*Proteínas ácidas ricas en glicina

Fuente: Smith, 2003

## **2.4. Diagnóstico de la infección tuberculosa**

El diagnóstico de tuberculosis se lleva a cabo a partir de la identificación de la bacteria en muestras provenientes del tracto respiratorio en caso de tuberculosis pulmonar, o de secreciones, líquidos o biopsias de otras partes del cuerpo (tuberculosis extrapulmonar) (OPS, 2006). Para la identificación de especies del complejo MTB se debe tener en cuenta: velocidad de crecimiento, características morfológicas de las colonias, y resultados de las pruebas bioquímicas como catalasa a T° ambiente, prueba de nitratos, prueba de niacina cuyos resultados en bacterias pertenecientes al complejo *Mycobacterium tuberculosis* son reacciones positivas (León *et al.*, 2002).

Uno de los métodos con más de 100 años para el diagnóstico inicial de tuberculosis es la baciloscopía, la cual busca la presencia de bacilos ácido-alcohol resistentes (BAAR), también se utiliza el cultivo de los especímenes clínicos en medio de Ogawa-Kudoh (OK) o Löwenstein-Jensen (LJ). La baciloscopía se basa en la capacidad de las micobacterias para formar complejos estables con ciertos colorantes de arylmetano como la fucsina, la cual penetra en la pared celular

uniéndose a los complejos micolil arabinogalactano. Este complejo retiene el colorante aún después de su exposición al ácido alcohol o ácidos minerales. Esta condición es requerida para que un microorganismo sea denominado ácido-alcohol resistente ((Raviglione y O'Brien, 2012).

Las coloraciones que demuestran la condición de ácido-alcohol resistencia de las micobacterias son el Ziehl-Neelsen y el Kinyoun (Fleischmann *et al.*, 2002). Es un método ventajoso para el diagnóstico de tuberculosis ya que es una técnica relativamente sencilla, de muy bajo costo, pero presenta problemas a nivel de sensibilidad, ya que en pacientes que no expulsan gran cantidad de bacilos no es posible detectarlos, en estos casos se realiza el cultivo, el cual sirve para diagnosticar los pacientes con baja carga bacilar, pero presenta el inconveniente que demora mucho el crecimiento y por lo tanto el diagnóstico, pero es muy útil para los pacientes que presentan resistencia a medicamentos, para la posterior realización de pruebas de sensibilidad a medicamentos antituberculosos (Warren *et al.*, 2004; Fujiki y Kato, 2008). El mayor inconveniente de la baciloscopia es su baja sensibilidad, 40 – 60 % como se indica en una revisión (Raviglione y O'Brien, 2012). Sin embargo ha surgido el método de baciloscopia concentrada con hipoclorito de sodio que aumenta la sensibilidad a 70 % aproximadamente, este método es de alta validez, casi semejante al cultivo, para el diagnóstico de tuberculosis en estos grupos de pacientes con radiografía anormal y BK negativo; siendo el costo y tiempo mínimo; así como, accesible a cualquier nivel básico de laboratorio (Barón *et al.*, 2012).

Otras metodologías que apoyan el diagnóstico de la tuberculosis son los rayos X, métodos inmunológicos tales como el Elisa, la realización de la prueba de la tuberculina (León *et al.*, 2002; Huebner *et al.*, 1993) que actualmente ha caído en desuso; métodos automatizados para la detección rápida de crecimiento bacteriano como BACTEC MGIT (CDC, 2006) que permite obtener resultados de pruebas de susceptibilidad a drogas anti-tuberculosas en un menor tiempo que el método de proporciones, donde BACTEC MGIT 960 presenta la ventaja considerable de ser un método no radioactivo (Tortoli *et al.*, 2002).

Las limitaciones que presentan los métodos tradicionales utilizados en el proceso de diagnóstico de infecciones por micobacterias han creado la necesidad de implementar nuevas estrategias para disminuir el tiempo necesario para el aislamiento e identificación de *Mycobacterium tuberculosis*. Hace 2 décadas, la irrupción y el posterior desarrollo de las técnicas de biología molecular en el mundo de la microbiología clínica han supuesto un gran avance en el diagnóstico y el tratamiento de las enfermedades infecciosas.

En la actualidad hay múltiples técnicas moleculares como PCR que amplifica la secuencia de inserción IS6110, característica del complejo MTB (Van Embden *et al.*, 1993). También el método PRA (PCR *Restriction Analysis*), en el cual se realiza una amplificación del gen *hsp 65* y el producto de PCR se corta con enzimas de restricción (Telenti *et al.*, 1993); la cuantificación de interferón gamma (Quantiferon) (Middlebrook *et al.*, 1977), ésta técnica se basa en la detección del IFN-gamma liberado como respuesta a la estimulación *in vitro* de las células T sensibilizadas, presentes en sangre periférica, con antígenos específicos de MTB (Lalvani *et al.*, 2001).

Actualmente en nuestro país se utilizan algunos métodos rápidos y moleculares para diagnóstico de tuberculosis MDR, como por ejemplo, la prueba molecular Genotype®MTBDRplus (Asencios *et al.*, 2012) diseñada para la detección de resistencia a isoniacida y rifampicina en un máximo de 72h a partir de esputo o de cultivo, en comparación con los métodos microbiológicos que demandan mucho tiempo; así también, el método de APP que demora 30 días a partir del cultivo para la obtención del resultado, mientras los métodos de BACTEC-460 y MODS (Caviedes *et al.*, 2000; Moore *et al.*, 2006) demoran entre 8 a 15 días respectivamente, desde la obtención de la muestra hasta la emisión de los resultados. Se debe señalar que la prueba Genotype®MTBDRplus requiere que la muestra de esputo sea baciloscopía positiva para poder utilizarla (Hain Lifescience, 2010), constituyéndose en una herramienta alternativa para la detección rápida de la resistencia de estas drogas antituberculosas.

## 2.5 Aspectos epidemiológicos de la tuberculosis

La tuberculosis es una de las enfermedades infecciosas de más amplia distribución en el mundo y constituye una de las primeras causas de muerte en pacientes con el Síndrome de Inmunodeficiencia Adquirida (SIDA), especialmente en países en vías de desarrollo como Perú. Un tercio de la población mundial está infectada con el *Mycobacterium tuberculosis*, cada año se enferman cerca de 8 millones de personas, de las cuales mueren dos millones, aproximadamente (Castiblanco y Willman, 2006) y es una de las primeras causas de muerte en países con alta prevalencia de esta enfermedad (De Paz *et al.*, 2012).

Aproximadamente unos 14 millones de personas presentan coinfección TB/VIH alrededor del mundo, siendo la TB la principal causa de muerte en pacientes con VIH (Bruchfeld *et al.*, 2015). La OMS estimó que, en 2014, se presentaron en el mundo 9.6 millones de casos nuevos de tuberculosis. Las regiones con mayor número de casos fueron: Asia con el 58% y África con el 28%. Los países con más casos fueron India (23%), Indonesia (10%) y China (10%). Las regiones que tuvieron el menor número de casos fueron: la región del Mediterráneo Oriental (8%), Europa (3%) y la región de las Américas (3%) (WHO, 2015).

De los 9.6 millones de casos de tuberculosis incidentes en el 2014, entre 1.1 y 1.3 millones (11% - 13%) se presentaron en personas VIH positivas. La mayor proporción de casos de TB coinfectados con VIH se presentó en la región de África con un 32% de casos. Durante el 2014 hubo un estimado de 1.5 millones de muertes por tuberculosis, de las cuales 1.1 millones se presentaron en personas VIH negativas (16 muertes por 100 000 habitantes), y 390 000 entre las personas VIH positivas (21 muertes por 100 000 habitantes). A nivel mundial, se estimó que el 3.3% de casos nuevos de tuberculosis y 20% de casos previamente tratados tuvieron tuberculosis MDR, las cuales cobraron alrededor de 190 000 vidas (WHO, 2015).



En el 2014 hubo un aumento considerable en la realización de las pruebas de susceptibilidad a medicamentos anti-tuberculosis con respecto a los años anteriores. 58% de pacientes tratados y 12% de nuevos casos fueron evaluados. Probablemente debido a la inclusión de las pruebas moleculares. Del total estimado de casos nuevos de tuberculosis MDR (480 000 casos), únicamente 123 000 (41%) fueron detectados y reportados. El inicio del tratamiento tuvo un incremento del 14% con respecto al 2013. Sin embargo, solo el 50% de pacientes fueron exitosamente tratados. Los casos de tuberculosis XDR se reportaron en 105 países. Se estima que un 9.7% de las personas con tuberculosis MDR tiene tuberculosis XDR (WHO, 2015).

En la Región de las Américas se estima que en el año 2011 se produjeron unos 268 400 casos de tuberculosis 100 menos que en el 2010. Alrededor de dos tercios (67%) de todos los casos nuevos de tuberculosis ocurrieron en América del Sur (subregión andina: 29%; otros países: 38%); 17% en México y Centroamérica; 11% en el Caribe, y 5,1% en América del Norte. Brasil ocupa actualmente el 17º lugar en todo el mundo y el primer lugar en las Américas en lo que respecta al número total de nuevos casos de tuberculosis (83 000). El 60% de todos los casos nuevos de tuberculosis que se registraron en la Región de las Américas correspondieron a Brasil, Perú, México y Haití. Más de un tercio de los 77 900 casos nuevos de tuberculosis registrados en los países andinos de América del Sur ocurrieron en Perú (30 000; 39%), que ocupa el segundo lugar en las Américas, seguido por Colombia (16 000; 21%). La mayor incidencia estimada de tuberculosis por 100 000 habitantes correspondió en el 2011 a Bolivia con 129 casos y Perú con 101 casos, donde había disminuido sustancialmente en los 20 últimos años (OPS, 2013).

En el Perú durante el año 2014 se presentaron 27 350 casos nuevos de tuberculosis de todas las formas. Desde el año 1997 hasta el año 2014 se han detectado más de 15 mil casos de tuberculosis MDR. El mayor número de casos de tuberculosis MDR se han reportado en los últimos 10 años (del 2005 en adelante) donde, el promedio reportado por año superó los 1 100 casos de tuberculosis MDR, con una tendencia creciente en los últimos 4 años. El mayor porcentaje de los casos de tuberculosis MDR, han sido reportados en

departamentos de la costa, Lima con el 83.1% y un 12 % en otros departamentos de la costa. Desde la detección del primer caso con tuberculosis XDR (1999) hasta el año 2014, se han diagnosticado más de 600 casos de tuberculosis XDR. En los últimos 7 años (2008-2014) se han detectado el 80% de los casos de tuberculosis XDR. La proporción de los casos de tuberculosis XDR entre los casos de MDR, se ha incrementado de manera acelerada. De menos de 2 casos de tuberculosis XDR por cada 100 casos de tuberculosis MDR antes del 2005, a 6 para el año 2014 (MINSA, 2016).

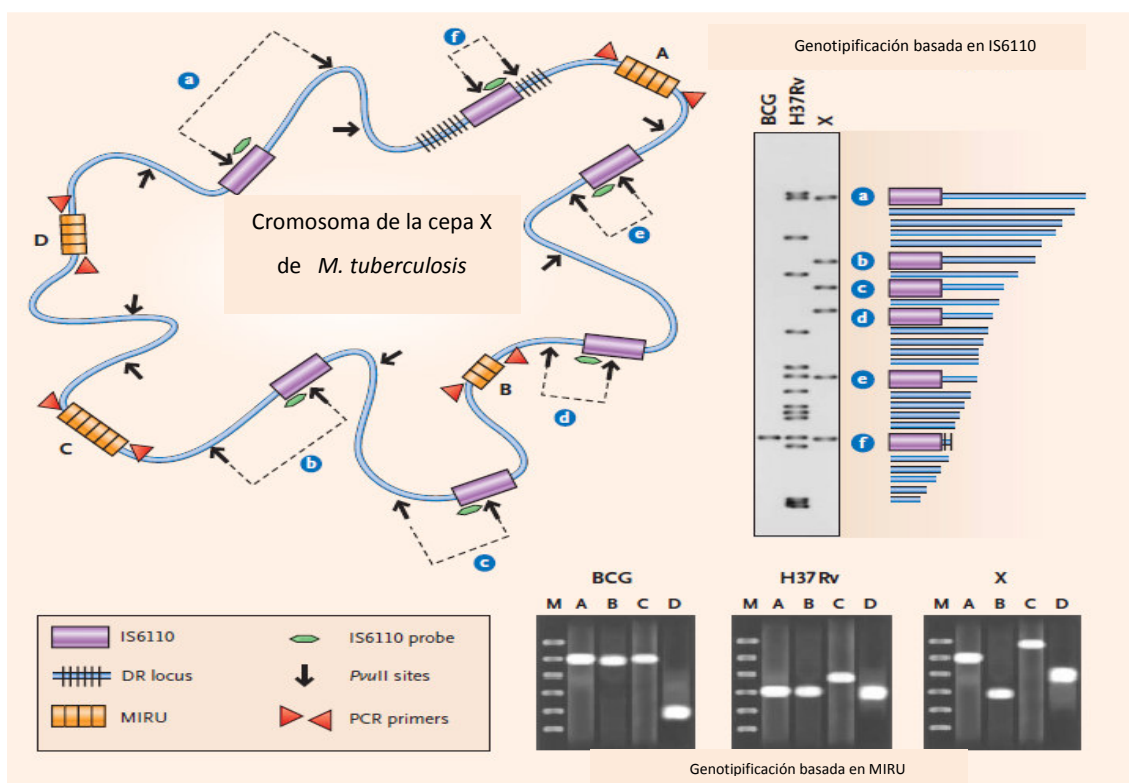
## **2.6. Epidemiología molecular de *Mycobacterium tuberculosis***

Actualmente la epidemiología molecular aplicada al estudio de enfermedades infecciosas, combina los métodos moleculares para la identificación de agentes patógenos con herramientas de la epidemiología convencional para el estudio de la distribución, comportamiento y factores de riesgo de una enfermedad (García *et al.*, 2003). Los estudios de genética y epidemiología molecular describen que el genoma de MTB posee regiones bien conservadas de ADN específicas de género y regiones hipervariables específicas de especie (Sola *et al.*, 2001).

Los métodos moleculares utilizados para el estudio y caracterización de miembros del complejo MTB han brindado un gran avance en el conocimiento de la dinámica de transmisión de esta enfermedad, han permitido la revelación de las bases genéticas implicadas en el desarrollo de resistencia a medicamentos y caracterización de brotes epidémicos en poblaciones e identificación de genotipos y familias del complejo (Van Embden *et al.*, 1993; Kamerbeek *et al.*, 1997; Supply *et al.*, 1997)

La secuenciación total del genoma de MTB ha permitido identificar diversos marcadores genéticos utilizados para la caracterización y diferenciación de cepas que infectan personas en un período de tiempo determinado, los cuales facilitan la caracterización de las cepas sin tener que realizar la secuenciación completa de las bacterias, constituyendo un ahorro en costos y tiempo (**Figura 2**) (Kwara *et al.*, 2003; Hawkey *et al.*, 2003).

El desarrollo de metodologías de *fingerprinting* ha permitido realizar estudios de transmisión global de la enfermedad, que permiten una comparación de cepas circulantes en distintas poblaciones y de esta manera establecer si hay un flujo genético de una población a otra, caracterizar brotes, hacer distinción entre transmisiones recientes o antiguas, detectar si hay contaminación cruzada en el laboratorio, detección de reinfección o recaída en pacientes con episodios secundarios de tuberculosis, identificación de poblaciones con alto riesgo de transmisión para la posterior implementación de la estrategia de control adecuada para contrarrestar la transmisión (Van Soolingen, 2001; Moström *et al.*, 2002; Peter *et al.*, 2003).



**Figura 2:** Ubicación de marcadores moleculares en el cromosoma de *Mycobacterium tuberculosis* para genotipificación: RFLP-IS6110, *Spoligotyping* y MIRU-VNTR (Barnes *et al.*, 2003).

Los métodos alternativos de identificación más comunes provienen del desarrollo de técnicas moleculares aplicadas a la identificación genotípica. Estas técnicas

están basadas en la amplificación de ADN bacteriano por PCR acompañadas por el estudio de polimorfismos generados por la digestión con enzimas de restricción, la hibridación con sondas específicas, o la secuenciación de cadenas nucleotídicas (Cortinas *et al.*, 2002).

En estudios realizados por Kremer *et al.* (1999) y Supply *et al.* (2001), se compararon cinco métodos de RFLP y ocho metodologías basadas en PCR. Se observó que, a pesar del potencial de las técnicas rápidas de amplificación para la caracterización de cepas de *Mycobacterium tuberculosis*, más de la mitad de los métodos utilizados tenían como limitación fundamental su reproducibilidad y/o poder de discriminación. De estos, los que mejores resultados aportaron fueron RFLP-IS6110, este último se basa en la amplificación del locus DR compuesto por repeticiones directas de 36 pb y la determinación de la presencia o ausencia de las secuencias espaciadoras que las separan (Kremer *et al.*, 1999).

Actualmente, la secuenciación del genoma completo de los microorganismos está siendo ampliamente utilizada para caracterizar genomas individuales, transcripción y variación genética de las poblaciones y las enfermedades (Bentley *et al.*, 2008). Esta técnica, constituye una herramienta importante para el estudio de la tuberculosis, proporcionando un alto poder de comprensión y discriminación para medir la diversidad genética y puede ser utilizada para análisis filogenéticos y clasificación de *Mycobacterium tuberculosis*. Adicionalmente, con la continua reducción de sus costos puede ser considerada próximamente como la “prueba de oro” para la genotipificación epidemiológica de *Mycobacterium tuberculosis* (Coscolla *et al.*, 2010).

### 2.6.1 RFLP-IS6110

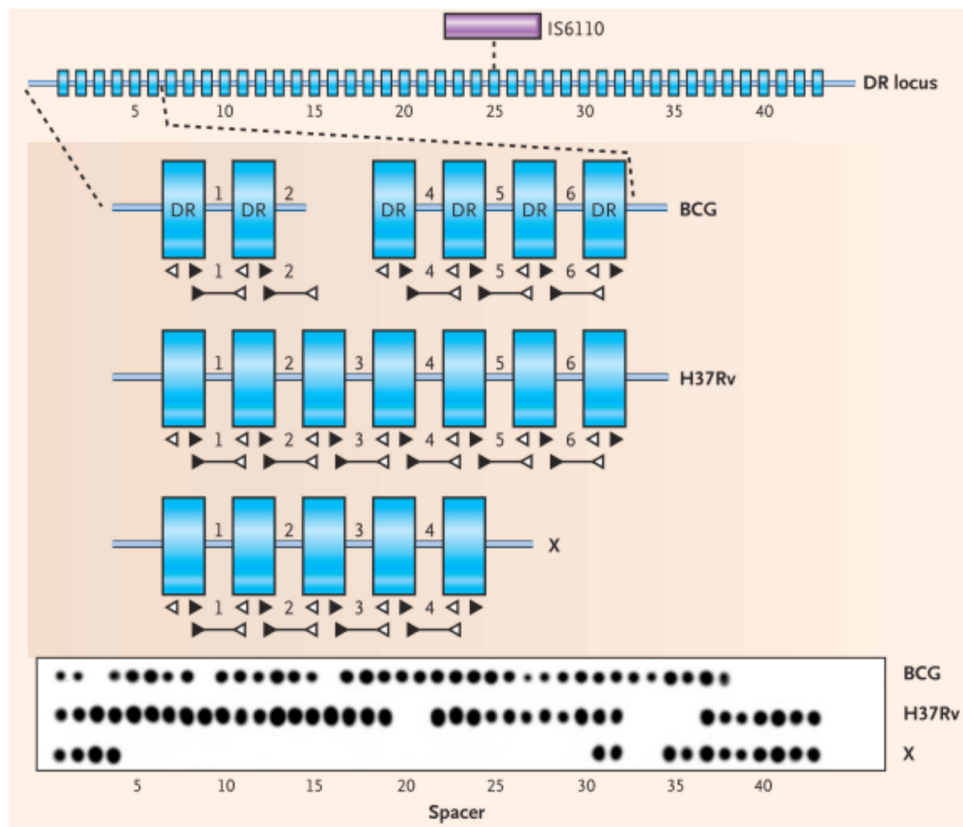
RFLP-IS6110 (Polimorfismo de Longitud de Fragmentos de Restricción) fue considerada la prueba de oro para genotipificar y caracterizar cepas de *Mycobacterium tuberculosis* debido a la variabilidad existente entre cepas, en el número de copias de la IS6110 y de su localización en el genoma de la micobacteria. Ésta metodología se basa en el polimorfismo de los fragmentos

largos de restricción que se obtienen después de cortar el ADN genómico con la enzima de restricción Pvu II. Se fundamenta en estudiar el número de veces que la secuencia de inserción IS6110 se repite en el genoma de la micobacteria (generalmente entre 1 y 25 veces) y aunque es una técnica muy reproducible, su mayor limitación es que no puede ser utilizada si la micobacteria tiene menos de 6 copias de esta secuencia en su cromosoma, en tal caso, se deben realizar otros métodos genotípicos. Adicionalmente, la genotipificación basada en IS6110 requiere de subcultivos de los aislados por algunas semanas para obtener suficiente ADN. Los aislados de pacientes infectados con cepas no relacionadas epidemiológicamente tienen patrones RFLP diferentes, mientras que los pacientes con cepas relacionadas epidemiológicamente por lo general tienen patrones RFLP idénticos (Barnes *et al.*, 2003).

### 2.6.2. *Spoligotyping*

Las cepas del complejo *Mycobacterium tuberculosis* contienen una única región en el cromosoma que consta de múltiples repeticiones directas (DRs) de 36 pb, éstas se encuentran intercaladas por espaciadores únicos que varían en tamaño de 35 a 41 pb (Groenen *et al.*, 1993) y es altamente polimórfico (Beggs *et al.*, 1996) (**Figura 3**). El número de DRs y la presencia o ausencia de espaciadores es variable entre las cepas. Su polimorfismo probablemente se debe a recombinaciones homólogas entre DRs cercanos o distantes y rearrreglos producidos por el elemento de inserción IS6110, el cual está presente en la mayoría de cepas de MTB (Groenen *et al.*, 1993). Esto puede llevar a la pérdida de sus espaciadores, las cuales son irreversibles (Shabbeer *et al.*, 2012).

El *Spoligotyping* (Espoligotipado), es una metodología basada en la amplificación de ADN de DRs y la presencia/ausencia de los espaciadores indican los diferentes patrones de hibridación del ADN amplificado. La hibridación es realizada con múltiples oligonucleótidos sintéticos de los espaciadores y son unidos covalentemente a una membrana de nylon. Para la genotipificación se utilizan sólo 43 secuencias espaciadoras de las 94 que se han encontrado. Para visualizar los 43 espaciadores amplificados, uno de los dos primers está marcado con biotina.



**Figura 3.** DRs y espaciadores utilizados en *Spoligotyping* (Barnes *et al.*, 2003).

Posteriormente, los puntos de hibridación se revelan con un conjugado de estreptavidina-peroxidasa, éste se une a la biotina, que al agregar el sustrato luminol forman un complejo luminoso. La reacción enzimática que se produce quema una placa radiográfica. El punto observado corresponde a cada espaciador hibridado, de esta forma se determinan los espaciadores presentes o ausentes en cada aislado evaluado (Acosta *et al.*, 2009).

El *Spoligotyping* es un método muy utilizado por su relativa simplicidad, rapidez y bajo costo. Los resultados se expresan en códigos binarios de 43 dígitos construidos por la presencia (1) o ausencia (0) de los espaciadores, que permiten crear bases de datos simples, con código binarios u octales (**Figura 3**). Estos resultados posteriormente son analizados filogenéticamente para determinar los clusters según su similitud genética. La codificación permite comparar fácilmente el genotipo de cepas aisladas en diferentes laboratorios en todo el mundo (Shabbeer

*et al.*, 2012). Los datos obtenidos se comparan con una base de datos internacional que tiene más de 11 000 SITs (*Spoligotype International Type*) de aislados obtenidos en más de 90 países, de esta manera se obtiene la identificación de los diferentes genotipos y sublinajes de cada aislado del complejo *Mycobacterium tuberculosis* (Filliol *et al.*, 2002).

La clasificación basada en *Spoligotyping CRISPR* tiene una base de datos con miles de cepas reportadas (Driscoll, 2009). Recientemente se ha vuelto más importante, ya que existe una correlación positiva entre la presencia de espaciador de emparejamientos de fago y fago-resistencia (Bolotin *et al.*, 2005; Pourcel *et al.*, 2005) como un posible mecanismo similar al ARN de interferencia resistente contra elementos génicos invasores (Makarova *et al.*, 2006), conteniendo restos y huellas de bacteriófagos encontrados previamente.

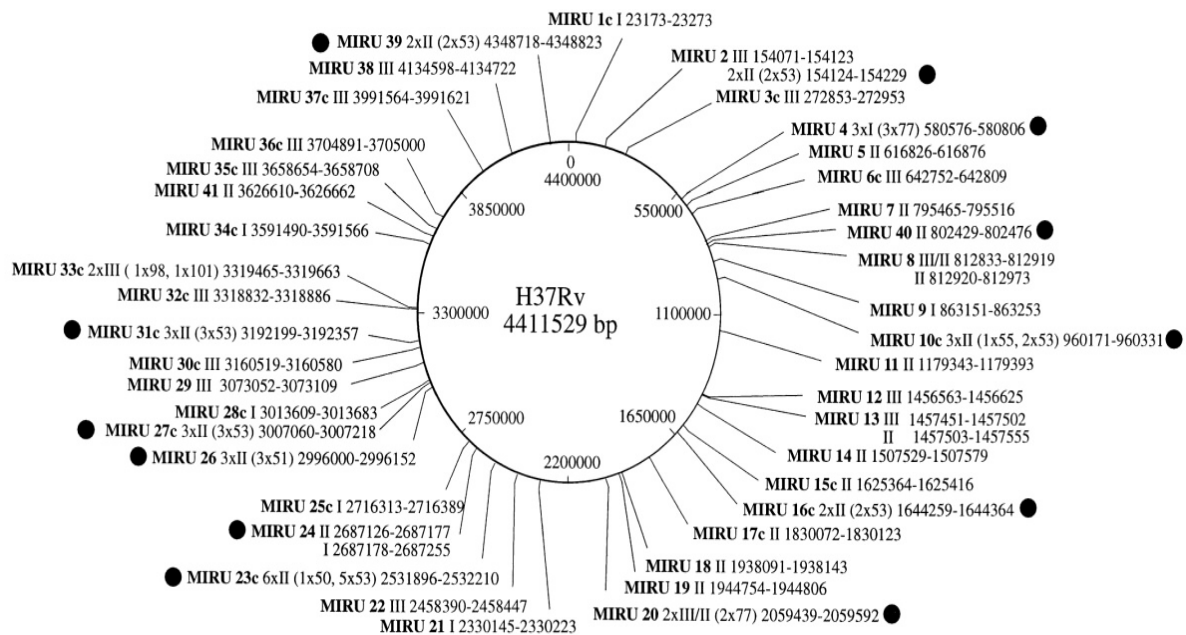
A pesar de las ventajas que tiene esta técnica, mencionadas previamente, su poder de discriminación es bajo, por lo tanto, necesita de otra técnica que permita una mayor diferenciación entre cepas (Barnes *et al.*, 2003).

### 2.6.3. MIRU-VNTR

Las Unidades Repetitivas Micobacterianas Interespaciadas - Número Variable de Repeticiones en Tándem están dispersas en miles de copias en los genomas eucariotas superiores. Los *loci* con repeticiones de secuencias cortas (SSRs) de 1-13 pb son llamados generalmente microsátélites, y los que tienen de 10-100 pb secuencias repetidas son llamados minisátélites. Muchos de estos *loci*, muestran hipervariabilidad en su número de repeticiones tanto en humanos, como en animales. Estos también son llamados *loci* de repeticiones en tándem de número variable (VNTR) y son una gran herramienta para el estudio de evolución y estudios de *pedigree* (Jeffreys *et al.*, 1985; Supply *et al.*, 2000). El tamaño de cada amplicón indica el polimorfismo del VNTR y se expresa como el número de copias que contiene. Las secuencias VNTR han emergido como marcadores para genotipificación de algunas especies bacterianas, especialmente para patógenos

genéticamente homogéneos tales como *Bacillus anthracis*, *Yersinia pestis* y los miembros del complejo *Mycobacterium tuberculosis* (Keim *et al.*, 1999; Klevytska *et al.*, 2001; Le Fleche *et al.*, 2001).

Supply *et al.* (2000), reportaron la identificación de minisatelites en el genoma de *Mycobacterium tuberculosis*. Estas estructuras están compuestas de secuencias repetidas llamadas MIRU de 40 a 100 pb y se encontraron dispersas en 41 sitios del cromosoma de la cepa H37Rv de *Mycobacterium tuberculosis* (**Figura 4**). Los MIRU han sido clasificados en tres grupos principales. Tipo I secuencias que contienen alrededor de 77 a 101pb, Tipo II de 46 a 53 pb y Tipo III de 58 a 101pb. Se estiman alrededor de 40 a 50 MIRU en el genoma de *Mycobacterium tuberculosis*; algunas contienen unidades repetidas idénticas y otras contienen repeticiones que varían ligeramente en secuencia y longitud (Supply *et al.*, 1997).



**Figura 4.** Posición de 41 *loci* MIRU en el cromosoma de *Mycobacterium tuberculosis* H37Rv (Supply *et al.*, 2000).

Inicialmente, se observó que 12 de estos *loci* presentaban polimorfismos en el número de copias MIRU entre aislados de *Mycobacterium tuberculosis* no



relacionados. Por lo tanto, los sistemas iniciales de tipificación VNTR para las cepas del complejo *Mycobacterium tuberculosis* hacían uso de dichos *loci* (12 *loci*) (Supply *et al.*, 2000). Sin embargo, no resultó ser suficientemente discriminatorio (Kremer *et al.*, 1999). Posteriormente se utilizaron 15 *loci* y en la actualidad se están utilizando 24 *loci* altamente polimórficos y con un mayor poder discriminatorio (Supply *et al.*, 2006).

La genotipificación basada en VNTR de diferentes clases de MIRU categoriza el número y tamaño de las copias de cada MIRU independiente, mediante la amplificación de múltiples *loci* usando primers específicos para las regiones que flanquean cada locus mediante PCR, y visualizando por electroforesis (**Figura 2**). En cada uno de los *loci* hay de 2 a 8 alelos, lo que da lugar a más de 20 millones de posibles combinaciones de ellos. El poder discriminatorio de la genotipificación por MIRU-VNTR es generalmente similar al RFLP - IS6110, con la diferencia que su análisis puede ser automatizado y de esta manera ser utilizado para la evaluación de un gran número de cepas, produciendo resultados digitales intrínsecos que pueden ser organizados en una base de datos. Un sitio web ha sido configurado para que una base de datos mundial de patrones MIRU pueda ser creada (Supply *et al.*, 2006). La tipificación por MIRU-VNTR ha mostrado una buena reproducibilidad y un aceptable poder de discriminación (Kremer *et al.*, 1999); es una técnica menos compleja que la genotipificación basada en RFLP-IS6110 y puede ser aplicada directamente a cultivos de *Mycobacterium tuberculosis* sin necesidad de purificar el ADN (Barnes *et al.*, 2003).

#### 2.6.4. **SNP**

Para tratar de resolver la clasificación de los genotipos de *Mycobacterium tuberculosis*, se ha intentado la genotipificación basándose en la secuenciación genómica (Roetzer *et al.*, 2013). Sin embargo, los costos de secuenciación masiva siguen siendo caros en los países donde la tuberculosis es más común. La clasificación de los genotipos ha sido reportada con diferentes métodos. Estudios recientes han propuesto un conjunto de SNPs (Polimorfismos de un solo nucleótido) asociados a un genotipo determinado; se reportó 45 SNPs utilizando 4 secuencias

genómicas (Filliol *et al.*, 2006), y 62 SNPs utilizando 1 601 genomas (Coll *et al.*, 2014). Sin embargo, estos reportes han mostrado limitaciones.

Homolka *et al.* (2012) estudiaron el genoma de 104 cepas pertenecientes al *Mycobacterium tuberculosis* de pacientes que vivían en Alemania, revelando una alta diversidad poblacional, comprendiendo cepas de varios genotipos filogenéticos, que comprenden las regiones del oeste de África, Beijing, Ghana y Uganda. La secuencia de los datos reveló la presencia de 61 polimorfismos de nucleótido simple (SNPs) de los cuales 41 fueron polimorfismos no-sinónimos y 20 sinónimos. 29 de los 61 SNPs fueron previamente encontrados como específicos de linajes o genotipos en la colección de referencia MIRU-VNTRplus ([www.MIRU-VNTRplus.org](http://www.MIRU-VNTRplus.org)).

Zhang *et al.* (2013) secuenciaron y analizaron 161 aislados con un alto rango de perfiles de resistencia a medicamentos, descubrieron 72 nuevos genes, 28 regiones intergénicas (IGR), 11 SNPs no sinónimos y 10 IGR SNPs con asociaciones fuertes a resistencia a antibióticos. Sugieren que los genes asociados con resistencia a drogas probablemente contienen esencialmente SNPs no sinónimos.

Illina *et al.* (2013) secuenciaron tres genomas de cepas MTB de Rusia de diferentes fenotipo: sensibles, MDR y XDR. Con base en el análisis filogenético, sus aislamientos pertenecían a diferentes genotipos: Beijing, Ural y LAM, lo que hizo imposible la comparación directa de sus genomas. Confirmaron SNPs únicos de XDR del aislado CTRI-4XDR, perteneciente a la familia Beijing cuando se comparó con SNPs de 14 cepas XDR rusas adicionales de la misma familia. Dentro de este grupo solo encontraron mutaciones específicas en los genes de reparación, replicación y sistema de recombinación (categoría L del grupo COG), probablemente los otros SNPs únicos descubiertos en CTRI-4XDR postulan que pueden tener un papel importante en la adaptación de este microorganismo a su alrededores y el escape al tratamiento por medicamentos antituberculosos.

En el Perú, Tarazona *et al.* (2014) reportan la secuencia de todo el genoma de una aislado sensible de *Mycobacterium tuberculosis* del Perú (INS-SEN) perteneciente al genotipo Latin American-Mediterráneo (LAM). El análisis funcional

reveló más mutaciones en la biosíntesis de metabolitos secundarios, el transporte, y catabolismo que para otras cepas LAM sensibles. Este estudio contribuyó a la comprensión de la diversidad genómica de *Mycobacterium tuberculosis* sensible a los fármacos.

Galarza *et al.* (2014) reportan la secuencia completa del genoma de *Mycobacterium tuberculosis* (INS-MDR) del Perú, una cepa tuberculosa multidrogoresistente (MDR) y genotipo Latin American-Mediterráneo (LAM). El análisis reportó mutaciones relacionadas con la resistencia a los medicamentos en los genes *rpoB* (D516V), *katG* (S315T), *kasA* (G269S) y *pncA* (Q10R). Este estudio sugiere que la cepa INS-MDR podría derivar de una expansión clonal relacionada a la cepa africana KZN 1435.

## **2.7. Secuenciamiento de ADN**

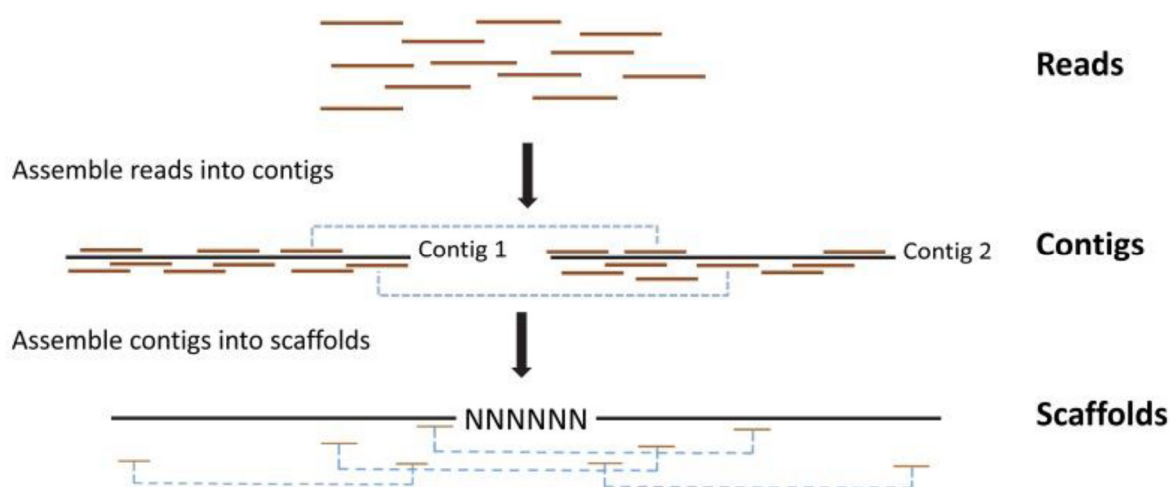
Las tecnologías de secuenciamiento de ADN, han ayudado a los biólogos proporcionando una amplia gama de aplicaciones, tal como estudios de la patogenicidad de genes, estudios comparativos y estudios evolutivos. La tecnología de secuenciamiento ideal debe ser rápida, precisa, fácil de operar y barata (Liu *et al.*, 2012).

En los pasados veinte años, las tecnologías de secuenciamiento han sufrido un tremendo desarrollo y han actuado como el motor de la era genómica, la cual es caracterizada por una vasta cantidad de datos y subsecuentemente un amplio rango de áreas de investigación (Liu *et al.*, 2012).

La primera tecnología de secuenciamiento introducida fue el método Sanger usando electroforesis de capilaridad. Esta tecnología ha sido aplicada en importantes proyectos de secuenciamiento y es considerada muy valiosa por la longitud y precisión de las lecturas. Después de esta tecnología han emergido las tecnologías de secuenciamiento de segunda generación (NGS). Dentro de este grupo encontramos Roche 454, Illumina GA y ABI SOLID, las cuales generan un número de lecturas de 3 a 4 órdenes de magnitud más que las producidas por

Sanger. Hoy en día estas tecnologías han sido aplicadas exitosamente a la identificación del sitio de unión de proteínas asociadas al ADN y secuenciamiento de RNA para el estudio de transcriptomas (Harismendy *et al.*, 2009).

El ensamblaje de referencia con cepa H37Rv pansensible consiste en el ensamble de lecturas dentro de largas secuencias contiguas llamadas contigs, seguido por el proceso del ordenamiento correcto de los contigs dentro de scaffolds (Sovic´, 2012). Este ensamble se realiza al analizar la similitud entre los extremos de dos contigs utilizando las lecturas pareadas (**Figura 5**).



**Figura 5:** Construcción de contigs y scaffolds. a) las lecturas son ensambladas en contigs y b) éstos a su vez son ensambladas en scaffolds (Johnson *et al.*, 2012).

La secuenciación de todo el genoma es actualmente el mejor método que proporciona una resolución superior. Sin embargo, sigue siendo costoso, especialmente en los países con alta incidencia de tuberculosis.

## 2.8. Firma genómica para genotipificar MTB

La firma genómica se refiere a la frecuencia característica de oligonucleótidos en un genoma o secuencia. Se ha observado que la firma genómica es similar en genomas filogenéticamente relacionados (Karlin y Burge, 1995).

La utilización de una firma genómica basada en SNPs es probable que sea una herramienta más exacta para estudios filogenéticos (Goujon *et al.*, 2000; Nakamura, 2001; Schork *et al.*, 2000). El análisis basado en SNPs es menos propenso a distorsión por presión selectiva en comparación con los marcadores genéticos polimórficos de secuencia larga como puede ser el caso de los espigotipos o MIRUS (Schork *et al.*, 2000). Además, los SNPs son selectivamente neutrales y deben acumularse a una velocidad uniforme y, por lo tanto, pueden usarse para medir la divergencia; es decir, pueden actuar como relojes moleculares.

Sólo se ha realizado un número limitado de estudios como firma genómica basados en SNPs para *M. tuberculosis* hasta la fecha. Sin embargo, estos estudios no han identificado sistemáticamente los grupos filogenéticos a través de modelos basados en el análisis de agrupaciones; así como, la composición y el número de agrupaciones de SNPs para *M. tuberculosis* no ha sido clara. Las relaciones entre las cepas y su país de origen tampoco fueron examinadas. Además, estudios anteriores no han aprovechado al máximo el poder de una firma genómica basada en SNPs para servir como un *Gold standard* para examinar la exactitud de los otros métodos de genotipificación de *M. tuberculosis*.

En nuestro estudio, correlacionamos el método de *Spoligotyping* y filogenia con un conjunto basado en SNPs del estudio comparativo de genomas de *M. tuberculosis*.

### **3. HIPÓTESIS Y OBJETIVOS**

#### **3.1. HIPÓTESIS**

La genotipificación de *Mycobacterium tuberculosis* utilizando un conjunto específico de SNPs basados en inclusión/exclusión tiene una correlación de 100% a las técnicas estándares.

#### **3.2. OBJETIVOS**

##### **3.2.1. Objetivo General**

- Proponer un nuevo conjunto específico de SNPs, basados en el principio de inclusión/exclusión, para la genotipificación de *Mycobacterium tuberculosis*

##### **3.2.2. Objetivos Específicos**

- Estandarizar la técnica de ensamblaje y anotación para los 249 genomas analizados de *Mycobacterium tuberculosis*.
- Estandarizar un conjunto de SNPs según el principio de inclusión/exclusión para la genotipificación de *Mycobacterium tuberculosis*.
- Determinar la eficiencia de la genotipificación de *Mycobacterium tuberculosis* en comparación con las técnicas estándares.

## 4. MATERIALES Y MÉTODOS

### 4.1. SELECCIÓN DE SECUENCIAS GENÓMICAS

En este estudio se utilizaron 249 secuencias genómicas de MTB (**Anexo I**): 125 de las secuencias genómicas tienen genotipos conocidos determinados por el método de *Gold standard (Spoligotyping)*, incluyendo Beijing (linaje 2), CAS (linaje 3), EAI (linaje 1 y 3), Haarlem (linaje 4), LAM (linaje 4), X (linaje 4), Ural (linaje 4), T (linaje 4), AFRI1 (linaje 5) y AFRI2 (linaje 6) basados en publicaciones y bases de datos públicas; 76 de las secuencias genómicas tienen genotipos conocidos determinados por análisis filogenético incluyendo linaje 1 (EAI), linaje 2 (Beijing), linaje 3 (CAS y EAI), linaje 4 (Haarlem, LAM, T, X), linaje 5 (AFRI1) y el linaje 6 (AFRI2); y 48 de las secuencias genómicas tienen genotipos desconocidos.

Estas secuencias genómicas se obtuvieron de los dominios públicos de las bases de datos internacionales del Instituto Nacional de Salud de USA (NIH), (<ftp://ftp.ncbi.nlm.nih.gov/genomes/>), del Archivo Europeo de nucleótidos (ENA; <http://www.ebi.ac.uk/ena/>) y de la Enciclopedia de genes y genomas Kyoto (KEGG) ([http://www.genome.jp/en/gn\\_ftp.html](http://www.genome.jp/en/gn_ftp.html)). Las secuencias de MTB analizadas fueron de 35 países (Alemania, Canadá, China, Colombia, El Salvador, Estados Unidos, Etiopía, Filipinas, Gambia, Ghana, Guatemala, Holanda, India, Irán, Italia, Japón, Kazajstán, Malasia, México, Mongolia, Nepal, Nicaragua, Panamá, Perú, Puerto Rico, Reino Unido, Rusia, Sierra Leona, Sudáfrica, Suecia, Tailandia, Tanzania, Uganda, Uruguay y Vietnam) lo que nos permite obtener una diversidad global y representatividad de MTB (**Figura 6**).

Adicionalmente, 34 secuencias genómicas de MTB con genotipos conocidos: Beijing (n = 12), LAM (n = 07), X (n = 06), CAS (n = 04), Haarlem (n = 04) y EIA (n=01) fueron utilizadas para confirmar el genotipo utilizando nuestro nuevo conjunto propuesto de 10 SNPs.

Los datos analizados fueron mapeados con el genoma de referencia H37Rv (número de acceso: NC\_000962.3) utilizando *BWA* (Li y Durbin, 2009). *SAMtools* (Li *et al.*, 2012), *kSNP* (Gardner *et al.*, 2015) y *ParSNP* se utilizaron para identificar SNPs. Estos programas excluyen el análisis de SNPs en regiones no secuenciadas.



**Figura 6:** Diversidad Global de 249 secuencias analizadas de *Mycobacterium tuberculosis* para este estudio (Fuente: Elaboración propia).

#### 4.2. ANÁLISIS FILOGENÉTICO DE LAS SECUENCIAS GENÓMICAS

Se realizaron ensayos filogenéticos en 173 genomas de MTB (125 con genotipos determinados por *Spoligotyping* y 48 con genotipos desconocidos) utilizando dos diferentes programas de alineamiento: *kSNP* (Gardner y Hall, 2013) y *ParSNP* (Treangen *et al.*, 2014). Cada programa utiliza diferentes algoritmos para seleccionar SNPs, construye las filogenias de máxima verosimilitud y compara las topologías. *kSNP* identifica SNPs basado en el análisis *k-mers* sin utilizar un genoma de referencia. Este programa requiere un *k*-valor para el tamaño del *mer*, que es calculado por la herramienta *Kchooser*. Además, con el fin de contar los *k-mers* en los genomas, *kSNP* utiliza el software *jellyfish* (Marcais y Kingsford, 2011), luego compara estos *mers* en todos los genomas para encontrar SNPs y finalmente crea una matriz de SNPs para hacer árboles filogenéticos. *ParSNP* utiliza encuentros únicos máximos (MUM) para generar una alineación genómica múltiple



y un árbol de SNPs. Este MUM para dos genomas es un índice genético que considera el nivel de conservación del ADN del genoma nuclear y la proporción de ADN compartida por estos genomas. Después del proceso de alineamiento, los SNPs del genoma nuclear son seleccionados y utilizados para construir un árbol filogenético.

#### **4.3. ANÁLISIS DE CLUSTER DE GRUPOS ORTÓLOGOS (COG)**

Para determinar los grupos COGs más variables y conservados para MTB, se realizó un análisis de COG en 249 genomas de MTB que incluyeron los genotipos Beijing, CAS, EAI, Haarlem, LAM, X, Ural, AFRI1, AFRI2 y desconocidos. Se realizó un análisis comparativo basado en H37Rv para determinar los SNPs, utilizando el programa *SAMTools* (Li *et al.*, 2009). A continuación, se generó una base de datos SNP que incluía SNPs correspondientes a genes y grupos COG.

#### **4.4. SELECCIÓN DEL GRUPO DE SNPs**

La asignación de genotipo se realizó en dos etapas: (I) utilizando el criterio de inclusión/exclusión en genotipos experimentales (Espoligotipos) y predicción filogenética y; (II) utilizando grupos COG.

(I) Criterio de inclusión/exclusión (IE) mediante datos de *Spoligotyping* y filogenias: Primero, se procesaron 125 genomas de MTB con información experimental de genotipos. Cada secuencia del genoma se transformó en códigos binarios, como presencia (1) o ausencia (0) de una mutación en *loci*, basados en H37Rv. Luego, los genomas se agruparon de acuerdo con sus genotipos (Beijing, CAS, EAI, Haarlem, LAM, X, Ural, AFRI1 y AFRI2). Utilizando el criterio IE, los *loci* compartidos con más de un genotipo fueron descartados. En segundo lugar, para reforzar la selección de SNP, se incluyeron 124 genomas de MTB adicionales (76 con genotipos determinados por análisis filogenético y 48 de genotipos desconocidos con genotipos predichos por análisis filogenético). Finalmente, los SNPs compartidos en más de un genotipo de MTB de Beijing, CAS, EAI, Haarlem,

LAM, X, Ural, AFRI1 y AFRI2 fueron descartados. Se identificaron un total de 2 123 SNPs específicos para cada genotipo de MTB reportándolos como SNPs-IE.

(II) Selección de SNPs no sinónimos basados en COG (*clúster* de grupos ortólogos): Luego, los SNPs fueron organizados en grupos funcionales COG (Tatusov *et al.*, 2001). Los genotipos de los genomas de MTB han mostrado una distribución de locus variante en COG a partir de la más variable a la más conservada (S, R, I, Q, E, C, F, L, N, G, H, K, P, J, T, O, F, D, V, U y A). En este caso, se seleccionó el locus de la familia de genes más conservada a la más variable. Luego, se seleccionó un locus asociado con un SNP no sinónimos para cada uno de los 9 genotipos MTB.

## 5. RESULTADOS

### 5.1. DETERMINACIÓN DE GENOTIPO POR ANÁLISIS FILOGENÉTICO DE 48 GENOMAS DE MTB DESCONOCIDOS

El árbol filogenético se construyó con 249 genomas de MTB, utilizando *kSNP* y *ParSNP* (**Figura 7**). Se identificaron ocho clados con buen soporte: AFRI1, EAI, CAS, Beijing, Haarlem, X, LAM y T. Sin embargo, el clado AFRI2 sigue siendo parafilético con cepas en diferentes niveles de la bifurcación. De los 48 genomas de MTB desconocidos, 03 son AFRI1, 16 son EAI, 03 son CAS, 22 son Beijing, 15 son Haarlem, 03 son X y 15 son LAM. Una cepa de Haarlem (Haarlem / NITR2) determinada por *Spoligotyping* se ubicó en el clado X mediante ambas metodologías (*kSNP* y *ParSNP*). Existen limitaciones en la asignación de genotipos por análisis filogenético ya que a menudo requieren información global de las secuencias genómicas para una óptima aproximación de genotipo.

### 5.2. IDENTIFICACIÓN Y SELECCIÓN DEL CONJUNTO DE SNPs

Para el desarrollo del sistema de asignación de SNPs y genotipos, primero se utilizó el principio de inclusión/exclusión. Se analizaron 125 genomas de MTB con genotipos conocidos determinados mediante *Spoligotyping*, como se describe en la sección de Metodología. El genotipo T no se analizó debido a que la secuencia genómica de H37Rv cae en este genotipo y se utilizó en el proceso de mapeo genómico. Además, 124 genomas de MTB se incluyeron con genotipos determinados por análisis filogenético.

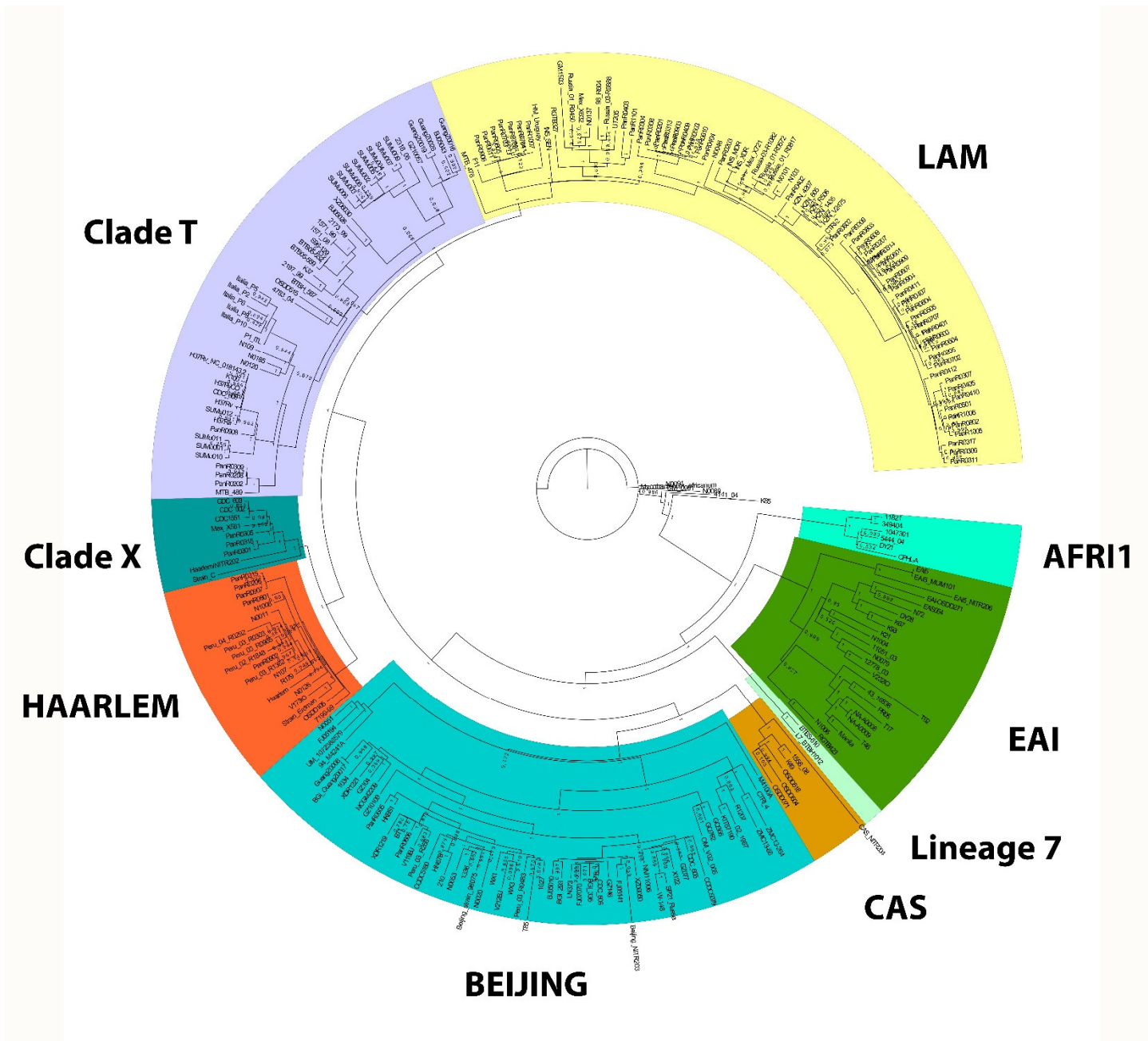
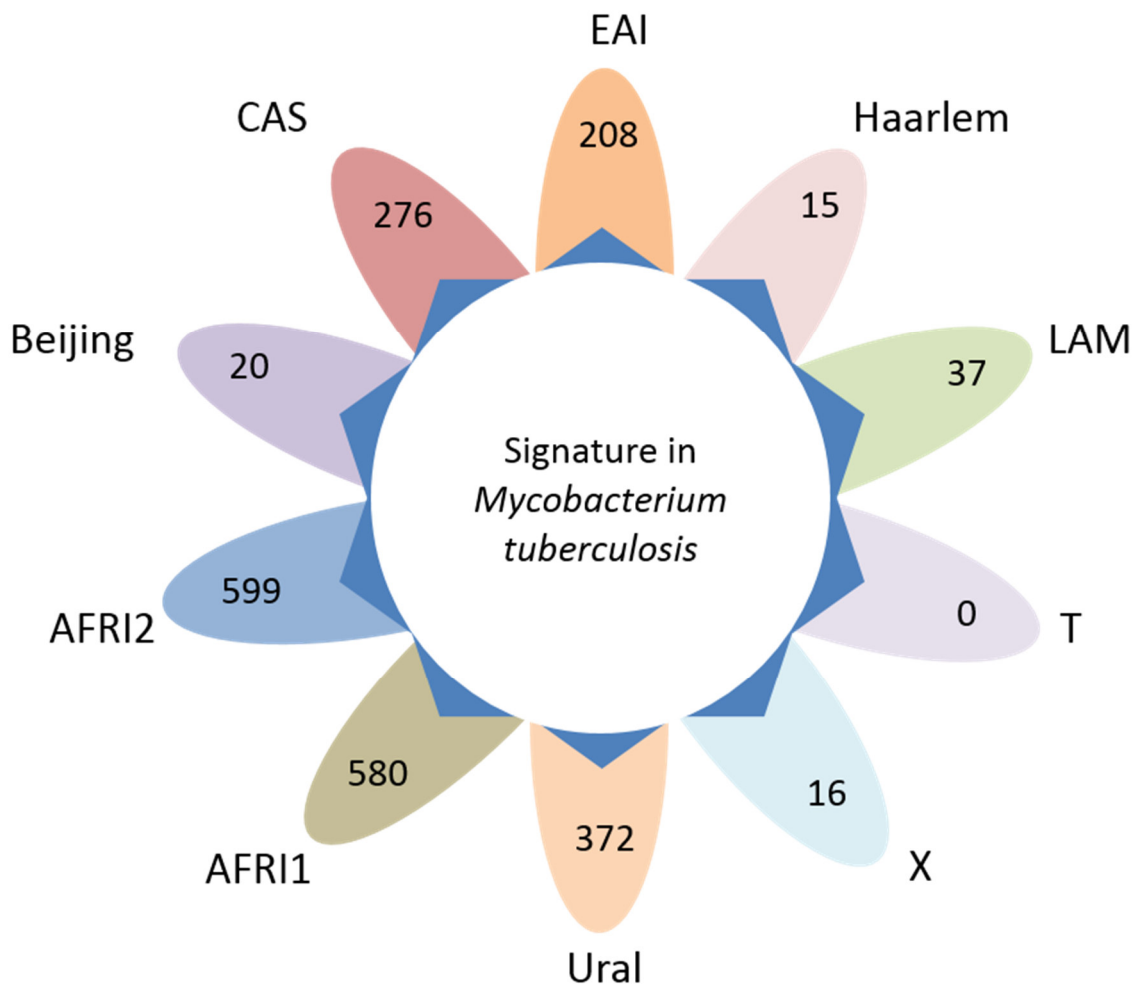


Figura 7: Árbol filogenético de Máxima parsimonia de 249 cepas de *Mycobacterium tuberculosis* basados en SNPs

En una base de datos, se han integrado los 7 649 SNP de 249 secuencias genómicas de MTB que incluyeron los genotipos Beijing, CAS, EAI, Haarlem, LAM, X, Ural, AFRI1 y AFRI2; el número de SNPs no específicos por genotipo fue de 458, 854, 674, 624, 424, 412, 886, 1 735 y 1 582, respectivamente. Después de la aplicación de los criterios de inclusión/exclusión, los SNP se organizaron en un llamado Grupo A, que contiene SNPs únicos para cada genotipo, y en un Grupo B, que contiene los SNPs compartidos entre los diferentes genotipos. A continuación, se siguió una exclusión manual de SNPs compartidos. El resultado fue un nuevo grupo de SNPs para los genotipos Beijing (n = 20), CAS (n = 276), EAI (n = 208), Haarlem (n = 15), LAM (n = 37), X (n = 16), Ural (n = 372), AFRI1 (n = 580) y AFRI2 (n = 599) (**Figura 8**).



**Figura 8:** Diagrama de SNPs específicos para genotipos de *Mycobacterium tuberculosis*: Beijing, CAS, EAI, Haarlem, LAM, TX, Ural, AFRI1 y AFRI2

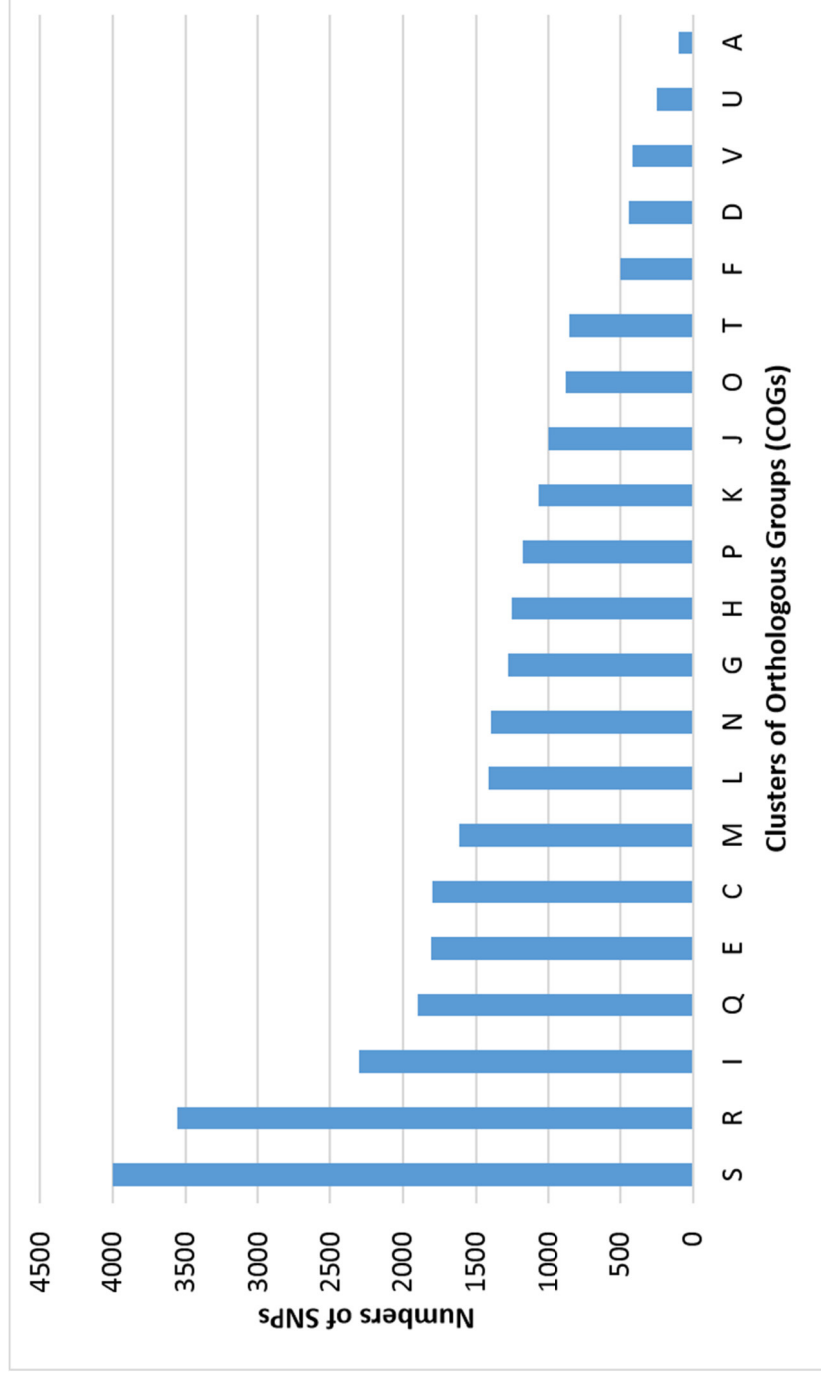
Finalmente, esto fue seguido por la selección de los SNPs no sinónimos presentes en el grupo COG más conservado de cada uno de los genotipos de MTB. Antes del análisis, se determinaron las COGs más variables y más conservadas para cada genotipo de MTB (**Figura 9**). Luego, se generó una base de datos de SNPs, que incluye SNPs correspondientes a genes y grupos COG, mostrando 2, 36, 20, 4, 2, 2, 40, 65, 67 SNPs para Beijing, CAS, EAI, Haarlem, LAM, X, Ural , AFRI1 y AFRI2 respectivamente (**Anexo II**). Se seleccionaron 10 SNPs no sinónimo para los 9 genotipos de MTB analizados (**Tabla 2 y 3**).

### **5.3. DISTRIBUCIÓN DE SNPS EN GENOTIPOS DE *Mycobacterium tuberculosis***

El genotipo Beijing contiene 16 SNPs no sinónimos que están presentes en los genes FadD34, LipU, GcvH, AccD6, FadE20, Mas, GuaB2, Ltp2, LysS y Prox.

El genotipo CAS contiene 161 SNPs no sinónimos distribuidos en los genes RecF, PonA1, Icd2, FbpC, familia PE, YrbE1B, Mce, FadE4, familia PE-PGRS, EccD3, Tam, FadD30, YrbE2B, VapC6, SerA2, Cyp51, CysK2, Pdc, EchA6, BetP, RimJ, LpqU, KdpA, LytB2, DapD, LpqX, Cys, GlgP, IrtA, PyrC, FadD25, DnaE1, MmpL6, IlvA, HisB, Cya, InfC, PknF, Cyp144, Aao , FadB5, LppF, PfkB, pPM1, PafC, FadD15, Cyp124, UspC, MmpL9, MbtG, GgtB, PepN, AccD1, Apt, CeoB, HsdM, HsdS, PepR, Smc, PpsA, NrdE, CstA, MoaR1, DevR, FadE24, NuoF, ManB, OtsB2, FdxB, FtsH, PonA2, DnaZ, Embc, EccB1 y EccD1.

Adicionalmente, 16 SNPs no sinónimos se encontraron en los genes de la familia PPE.



**Figura 9:** Distribución relativa de SNPs de *Mycobacterium tuberculosis* pertenecientes a ciertos Clúster de grupos Ortólogos (COG): Función desconocida (S), predicción general (R), Lípido de transporte y metabolismo (I), Biosíntesis de metabolitos secundarios, transporte y catabolismo (Q), Transporte de aminoácidos y metabolismo (E), Producción y conversión de energía (C), Biogénesis de la pared celular/membrana (M), Replicación, recombinación y reparación (L), Motilidad celular (N), Transporte de hidrato de carbono y metabolismo (G), Transporte de coenzima y metabolismo (H), Transcripción (K), Transporte de iones inorgánicos y metabolismo (P), Traducción, estructura ribosomal y biogénesis (J), Modificación post-traducciona, recambio proteico, y chaperones (O), Mecanismos de transducción de señal (T), Transporte de nucleótidos y metabolismo (F), Control del ciclo celular, división celular, partición de cromosomas (D), Mecanismos de defensa (V), Tráfico intracelular, secreción, y transporte vesicular (U), Procesamiento del ARN y modificación (A).

**Tabla 2:** Pasos para la selección de SNPs

Etapas		Genotipo	Beijing	CAS	EAI	Haarlem	LAM	X	Ural	AFRI1	AFRI2	T	Total
(1)	Integración de 7649 SNPs	Según Gagneux, 2013	L2	L3	L1 L3	L4	L4	L4	L4	L5	L6	L4	
		Secuencias genómicas de MTB	52	06	23	19	78	09	01	06	05	50	249
(2)	Selección por criterio IE	SNPs Totales	458	854	674	624	424	412	886	1735	1582		7649
		SNPs – IE	69	310	220	33	93	56	396	737	621		2535
(3)	Selección de SNPs no sinónimos presentes en grupos COG mas conservados	SNPs - IE*	20	276	208	15	37	16	372	580	599		2123
		Conjunto de SNPs basados en grupo COG **	03	29	22	01	03	01	36	68	79		243
		Grupo COG en MTB más conservado	F	U	U	F	D	K	A	A	A		-
Conjunto de SNPs propuesto ***			01	01	01	01	01	01	01	01	02		10

IE, genotipo diferencial. \*, se eliminaron los *loci* pertenecientes a otras cepas de diferentes genotipos y poco comunes entre genotipos. \*\*, conjunto de SNPs basados en el grupo COG (A→K). \*\*\*, conjunto de SNPs basados en el grupo COG menos variable por genotipo.



**Tabla 3:** Conjunto de 10 SNPs propuestos para genotipificación de MTB

<b>Genotipo Diferencial</b>	<b>COG</b>	<b>Gen (Locus)</b>	<b>Posición y SNP propuesto</b>
<b>Beijing</b>	Transporte de nucleótidos y metabolismo	GuaB2 (Rv3411c)	3830349 (Ala391Thr: GCG ->ACG)
<b>CAS</b>	Tráfico intracelular, secreción y transporte vesicular	(Rv3921c)	4409954 (Ala39Gly:GCC-> GGC)
<b>EAI</b>	Tráfico intracelular, secreción y transporte vesicular	SecE1 (Rv0638)	734116 (Met127Thr: ATG->ACG)
<b>Haarlem</b>	Transporte de nucleótidos y metabolismo	Hpt (Rv3625c)	4063682 (Leu61Met: CTG->ATG)
<b>LAM</b>	Control del ciclo celular, división celular, partición cromosómica	Smc (Rv2922c)	3236230 (Arg526Leu:CGT->CTT)
<b>X</b>	Transcripción	(Rv2618)	2946570 (Gly63Asp:GGC->GAC)
		(Rv3625c)	4063682 (Leu61Met:CTG->ATG)
<b>Ural</b>	Procesamiento y modificación del ARN	(Rv1097c)	1225462 (Asp228Gly:GAC->GGC)
<b>AFRI1</b>	Procesamiento y modificación del ARN	(Rv3439c)	3858894 (Leu266Phe:CTT->TTT)
<b>AFRI2</b>	Procesamiento y modificación del ARN	(Rv1097)	1226021 (Gln42Lys:CAG->AAG)
		(Rv3689)	4130604 (Ser42Asn:AGC->AAC)

El genotipo EAI contiene 129 SNPs no sinónimos en los genes OpcA, NdhA, TopA, ProA, Phet, FadD1, HtdX, PapP3, ThiX, Lrp, IleS, EspK, Ppm1, Dxs2, LipO, MoxR2, FadD1, PhoU, GadB, MoaC1, PgsA1, EccC5, Cyp1351A1, SdhC, PimB, FodD, GyrA, AccA2, LrpA, LysX, Pks8, PyrG, PstS1, AglA, LppE, MoeY, AccD2, FadD22, EngA, YrbE1A, OtsB1, SecE1, Lat, FadD11, Blal, DsbF, PimB, NarK1, EchA12, DppA, DnaK, GlpX, FadD11, Cya, Mas, FadE31, FadD2, EmbR, AhpD, EccE5, AtsB, AldC y YrbE2B. Los SNPs incluidos en la familia Fad están implicados en el metabolismo de lípidos.

El genotipo Haarlem contiene 9 SNPs no sinónimos en los genes RpoC, OtsB1, Topa, NarH, HpT, Cut3, FadD17, MtbC, Pks6 y PurF. Haarlem también está asociada con la resistencia a los medicamentos y, además, tenía una rápida expansión clonal (Parwati *et al.*, 2010; Ramazanzadeh *et al.*, 2015).

El genotipo LAM contiene 18 SNPs no sinónimos en los genes MceD1, RecC, Cyp130, ArgS, Mce3F, Smc, CstA y Tgs3. El genotipo X contiene 9 SNPs no sinónimos en los genes CelA1, Cyp138, EchA1, ThiO, CtpH, Pdc, PyrB, Pks5, FadB5 y PepR. Los estudios en otros genotipos de *M. tuberculosis* siguen siendo desarrollados.

#### **5.4. VALIDACIÓN DE LA ASIGNACIÓN DE GENOTIPOS MEDIANTE EL NUEVO CONJUNTO DE 10 SNPs PROPUESTO**

Utilizando otros 34 genomas de MTB con genotipos conocidos Beijing (n=12), LAM (n=07), X (n=06), CAS (n=04), Haarlem (n=04) y EAI (n=01), probamos la capacidad de nuestro nuevo conjunto de 10 SNPs para asignar genotipos. Obtuvimos una correlación del 100% de la asignación de genotipos con todas las secuencias genómicas evaluadas (**Tabla 4**).

**Tabla 4.** Asignación del genotipo en 34 genomas de MTB utilizando el nuevo conjunto de 10 SNPs propuesto

	<b>Cepa</b>	<b>Número de accesión</b>	<b>Genotipo de referencia</b>	<b>Genotipo asignado por nuestro conjunto de 10 SNPs</b>
1	13-2459	LDNL00000000	Beijing	Beijing
2	5351	JXXH01000000	Beijing	Beijing
3	96075	CP009426	Beijing	Beijing
4	B9741	LVJJ01000000	Beijing	Beijing
5	BEIJING-L 323	CP010873	Beijing	Beijing
6	BeijingDS 6701	JOKR01000001	Beijing	Beijing
7	E186hv	JXAW00000000	Beijing	Beijing
8	KT-0133	JUFG00000000	Beijing	Beijing
9	MTBR209	LATO00000000	Beijing	Beijing
10	W06	LHCK00000000	Beijing	Beijing
11	ZT272	LGTJ00000000	Beijing	Beijing
12	TBR-103XDR	JRJT01000001	Beijing	Beijing
13	tahitMT11	CVMX01000001	Haarlem	Haarlem
14	TBR-102	JRJS00000000	Haarlem	Haarlem
15	TKK_03_0101	GCF_000651975.1	Haarlem	Haarlem
16	TKK_03_0103	GCF_000651995.1	Haarlem	Haarlem
17	TBR-152	JRJQ00000000	LAM	LAM
18	TKK_04_0029	GCF_000673435.1	LAM3	LAM
19	TKK_04_0038	GCF_000673275.1	LAM4	LAM
20	TKK_04_0039	GCF_000673295.1	LAM4	LAM
21	TKK_04_0043	GCF_000673075.1	LAM4	LAM
22	TKK_04_0044	GCF_000673335.1	LAM3	LAM
23	TBR-175	JRJR00000000.1	LAM	LAM
24	TKK_04_0120	GCF_000654175.1	EAI	EAI
25	TKK-01-0028	GCF_000660665.1	X	X
26	TKK_02_0027	GCF_000672095.1	X	X
27	TKK_03_0063	GCF_000651695.1	X	X
28	TKK_03_0099	GCF_000651935.1	X	X
29	TKK_03_0150	GCF_000652255.1	X	X
30	TKK_05SA_0021	GCF_000653515.1	X	X
31	TKK_03_0037	GCF_000651475.1	CAS	CAS
32	TKK_04_0139	GCF_000656875.1	CAS	CAS
33	TKK_04_0148	GCF_000656935.1	CAS	CAS
34	TKK_05SA_0050	GCF_000653755.1	CAS	CAS

## 6. DISCUSIÓN

La tuberculosis sigue siendo un problema de salud importante en todo el mundo, con un tercio de la población infectada con *Mycobacterium tuberculosis*. A pesar de que el genoma de MTB está altamente conservado, se ha demostrado que diferentes cepas presentan diversas características fenotípicas que pueden influir en la infección y la enfermedad de la tuberculosis. Es importante clasificar las cepas de MTB (genotipo) y determinar sus factores de riesgo en la transmisión de la tuberculosis en diferentes poblaciones para desarrollar estrategias para el control de la tuberculosis. La secuenciación de todo el genoma es actualmente el mejor método que proporciona una resolución superior. Sin embargo, sigue siendo costosa, especialmente en países con alta incidencia de tuberculosis.

El uso de SNPs está ganando rápidamente importancia para la genotipificación del complejo MTB, principalmente debido a las tecnologías disponibles originalmente desarrolladas para otros organismos. Sin embargo, es importante tener en cuenta que, al igual que el análisis de delección genómica, la genotipificación mediante SNP no proporciona la resolución necesaria para ser aplicable en los estudios epidemiológicos moleculares clásicos de la transmisión de la tuberculosis (Kato-Maeda *et al.*, 2011).

Sin embargo, la genotipificación mediante SNPs es ideal para clasificar las cepas del complejo MTB en genotipo, ya que exhibe niveles insignificantes de homoplasia; es decir, eventos de mutaciones independientes que resultan en el mismo genotipo entre aislamientos con diferentes ancestros (Comas *et al.*, 2009). Es importante destacar que, en contraste con el descubrimiento de SNP a través de la secuenciación del ADN (Baker *et al.*, 2004; Hershberg *et al.*, 2008), la genotipificación mediante SNPs es de uso limitado en los estudios filogenéticos debido al problema conocido como "sesgo de descubrimiento filogenético" (Pearson *et al.*, 2004). Este sesgo se refiere al hecho de que si sólo se utilizan unos pocos genomas para identificar los SNPs para la posterior genotipificación, la filogenia resultante será sesgada, porque la diversidad genética entre las cepas no incluidas en el descubrimiento inicial de SNPs no se detectará. Como consecuencia, estas cepas caerán automáticamente en posiciones intermedias en

el árbol filogenético, dando lugar a los problemas conocidos como “Colapso de rama” o “filogenia lineal” (Achtman, 2008).

Recientemente, se ha desarrollado métodos de genotipificación mediante SNPs para la detección de los principales genotipos del complejo MTB (Stucki *et al.*, 2012). En contraste con la mayoría de los otros métodos disponibles, los SNPs usados en esta investigación han sido identificados por la comparación de muchos genomas mundiales representativos del complejo MTB.

La información reportada en esta investigación permite reconocer los genotipos más prevalentes que se encuentran circulando a nivel global, lo cual será de gran utilidad al sector salud, para establecer programas de control y prevención oportuna de casos de tuberculosis. Dentro de los más prevalentes, el genotipo LAM se cree que se originó en África (Viegas *et al.*, 2010; Thumamo *et al.*, 2012; Homolka *et al.*, 2008), y que llegó a América durante la colonización del continente, por eso su alta frecuencia en Latino América (Cerezo *et al.*, 2011; Taype *et al.*, 2012; Aristimuño *et al.*, 2006). El genotipo Haarlem es típico de Europa (Jagielski *et al.*, 2010; Ghebremichael *et al.*, 2008), pero se ha reportado con mayor frecuencia en América Central (Rosales *et al.*, 2010). El genotipo T tiene una distribución global (Molina *et al.*, 2010; Taype *et al.*, 2012). El genotipo asiático (Beijing), tiene tendencia a desarrollar multidrogorresistencia más frecuente que los otros genotipos en zonas endémicas propias del genotipo, como es el caso de China (Aguilar *et al.*, 2010; Rindi *et al.*, 2009), debiéndose posiblemente a la adaptabilidad al ambiente establecido.

La llegada de las tecnologías de secuenciación de nueva generación ha permitido nuevos análisis de SNPs basados en estudios filogenéticos (Filliol *et al.*, 2006; Coll *et al.*, 2014). Sin embargo, las limitaciones cuando múltiples secuencias están conectadas y analizadas con métodos estándar podrían desviar la verdadera filogenia (Jeffroy *et al.*, 2006; Liu *et al.*, 2008). Los errores producidos en el secuenciamiento de nucleótidos pueden llevar a posteriores errores en el ensamble de las secuencias y por lo tanto se puede formular datos erróneos de la secuencia genómica (Macmanes y Eisen, 2013). Esta dificultad se ha convertido en un problema en todos los proyectos de ensamble de secuencias y los investigadores

han empezado a diseñar software que corrijan estos errores en las lecturas (Medvedev *et al.*, 2011).

Hasta la fecha, hay más de 3 633 secuencias genómicas de MTB subidas en las diferentes bases de datos mundiales, pero todavía no hay datos sólidos para determinar todos los genotipos de *Mycobacterium tuberculosis*. Aunque el genoma de MTB es altamente conservado en su estructura y polimorfismos en comparación con las demás bacterias, éste presenta familias de genes polimórficos. Con el tiempo, *Mycobacterium tuberculosis* ha sufrido adaptaciones genéticas que permitieron obtener variaciones genéticas específicas para cada genotipo (Nicol *et al.*, 2008).

Los genes de las familias PE y PPE son altamente polimórficas y tienen un papel importante en la regulación de las células dendríticas y efecto inmune de macrófagos (Mukhopadhyay *et al.*, 2011), secreción y estabilización de la pared celular (Simeone *et al.*, 2015), y su variabilidad. Igualmente, la variación en los extremos c-terminal de las proteínas PE\_PGRS ha sido ligada a variación antigénica (Chaturvedi *et al.*, 2010). El papel de estos genes en la evasión inmune (Tiwari *et al.*, 2012) y en la variación antigénica se atribuye a las sustituciones, inserciones y deleciones presentes en los genes que cambian las propiedades de las proteínas que codifican (Kohli *et al.*, 2012). La gran cantidad de regiones únicas encontradas en nuestros aislamientos concuerdan con este postulado. Sin embargo, en el genotipo CAS, los genes de las familias PE y PPE aparecen más conservados que en otros genotipos.

Karboul *et al.* (2008) evaluaron el papel de la recombinación en los cambios de las secuencias de los genes de la familia PE/PPE. Los autores concluyeron que la variabilidad encontrada podría ser un mecanismo de adaptación en respuesta a las condiciones inmunes del hospedero. En relación con esta idea, los estudios de Talarico *et al.* (2008) sobre la variación de la secuencia de tres genes pertenecientes a la familia PE\_PGRS (PE\_PGRS16, PE\_PGRS26 y PE\_PGRS33), encontraron que la variación en los tres genes fue alta, apoyando la noción de estos genes como antígenos variable de superficie, que contribuyen a la evasión inmune, por lo tanto la variabilidad de los genes PPE y PE\_PGRS de los genomas utilizados

en este estudio, pueden contribuir a la virulencia de cada aislamiento. Futuras investigaciones se pueden centrar en el análisis funcional de los genes PPE/PE\_PGRS mencionados aquí para comprender el rol específico que tienen en MTB.

Los genes de la familia Fad son glicolípidos fenólicos de la superficie celular (Simeone *et al.*, 2015) necesarias para la síntesis de micobactina (Duckworth *et al.*, 2012) y captación de hierro (Luo *et al.*, 2005). Estos genes son la clave para la supervivencia a largo plazo debido a su uso como fuente de energía (Dunphy *et al.*, 2010). Hay pruebas que demuestran que las personas infectadas con cepas del genotipo EAI tienen mayores polimorfismos en el gen de la familia Fad, lo que podría explicar por qué las personas infectadas presentan una transmisión reducida o menor manifestación clínica de la tuberculosis, en comparación con otras cepas (Albanna *et al.*, 2011).

Los genes de la familia Mce juegan un papel importante en la internalización de *Mycobacterium tuberculosis* en células de mamífero y promueven la supervivencia de los macrófagos (Li *et al.*, 2015). Sin embargo, reportes indican que los SNPs no sinónimos (nsSNPs) cambian su función (Pasricha *et al.*, 2011). Los genotipos de MTB que tienen nsSNPs para estos genes son Ural, AFRI1 y AFRI2.

Los genotipos AFRI1 y AFRI2 tienen el mayor número de SNPs relacionados con el procesamiento y modificación de ARN en nuestra base de datos. También, se analizaron las mutaciones en las bombas de flujo transmembrana y transporte que pertenecen al grupo COG V (mecanismo de defensa), ya que pueden estar directamente no-asociados al genotipo LAM, Haarlem y X. Entre ellos, hemos identificado un nuevo SNP situado en el alelo del promotor TetR / acrR como un marcador informativo para reconocer cepas de los genotipos EAI y CAS.

El genotipo LAM tiene mayor grado de polimorfismos en el grupo COG relacionado con la biosíntesis, el transporte y el catabolismo de los metabolitos secundarios. Las adaptaciones genéticas de MTB corresponden a un patrón genético que, según nuestros datos, incluye una variación genética de las familias

de genes en la respuesta inmunitaria y la tasa de transmisibilidad de la tuberculosis entre individuos infectados en una región específica del mundo.

Los distintos hallazgos en la estructura genómica encontrados en este estudio apoyan los resultados de otros estudios en donde se observa que la tasa de mutación difiere entre cepas (Ford *et al.*, 2013; Merker *et al.*, 2013) y que la estructura genomas entre aislamientos es más variable de lo que se pensaba (Fleischmann *et al.*, 2002).

Así también, reportamos que ningún genotipo de *M. tuberculosis* estuvo relacionado con el país de origen de la muestra de los pacientes, incluso en países con alta incidencia también aumenta el número de genotipos circulantes.

Cada genotipo en MTB ha logrado obtener funciones especializadas basadas en mecanismos de adaptación o supervivencia, lo que ha generado una marca en su estructura genética principalmente por cambios en sus Espoligotipos. Uno de los mecanismos de supervivencia para la resistencia viral es *CRISPR*, un sistema inmune adaptativo en procariontes incluyendo MTB. Este sistema puede proporcionar una inmunidad adaptativa secuencia específica (Rath *et al.*, 2015) contra los fagos, plásmidos en las bacterias a través de repetición/espaciadores derivados de ARNs cortos (Makarova *et al.*, 2011), e incluso contra los macrófagos (Huang *et al.*, 2016). Esta es la razón por la cual el método *Gold standard* para la genotipificación de MTB durante mucho tiempo es el *Spoligotyping*, basado en *CRISPR*.

En años recientes, la genotipificación por MIRU-VNTR (número variable de repeticiones en tándem) ha sustituido al *Spoligotyping* como el método de genotipificación de elección para laboratorios de salud pública y estudios epidemiológicos. Así también, está disponible una herramienta en línea para convertir valores MIRU a un genotipo y familia de Espoligotipo apropiada (Weniger *et al.*, 2010). Futuros trabajos pueden considerar el uso de la metodología MIRU para definir hasta sublinajes; sin embargo, la determinación *in silico* del genotipo a partir de la secuenciación de alto rendimiento de cortas lecturas por SNPs se convertirá en la herramienta computacional más amigable.



Los SNPs aquí propuestos serán análogos al método de *Gold standard* para genotipificación de *Mycobacterium tuberculosis*, ya que utilizamos datos experimentales de *Spoligotyping*, criterios de inclusión/exclusión, filogenias y grupos COG para identificar estos SNPs.

De esta manera, en condiciones óptimas, *Spoligotyping* y MIRU-VNTR requieren de 1 a 6 días para dar resultados (Gori *et al.*, 2005), tiempo extenso para la determinación del genotipo en una muestra. Por otro lado, estudios genómicos previos identifican grupos de SNPs basados en estudios filogenéticos, por lo que sufren limitaciones del método filogenético estándar en secuencias múltiples.

En este estudio se propone un método sencillo para la genotipificación de cepas de MTB mediante el análisis de un conjunto de sólo 10 SNPs seleccionados basados en el criterio de inclusión/exclusión (IE) utilizando datos de *Spoligotyping* y filogenias, seguido de la selección de SNPs no sinónimos presentes en los clúster de grupos ortólogos (COG) más conservados.

La asignación del genotipo utilizando el nuevo conjunto de 10 SNPs fue validado con 34 genomas de MTB adicionales y los resultados mostraron una correlación del 100% con sus genotipos ya conocidos. Nuestro conjunto de 10 SNPs no ha sido reportado previamente y abarca los genotipos de MTB que son prevalentes en todo el mundo.

Sin embargo, el papel de esta diversidad de genotipos para la tuberculosis clínica sigue siendo en gran parte difícil de alcanzar, en parte debido a la dificultad de vincular los genotipos de la cepa de tuberculosis con fenotipos clínicos relevantes. Además de diferencias a nivel de ADN, las cepas de MTB, han demostrado que difieren en su transcriptoma (Homolka *et al.*, 2012). Sin embargo, no se conoce nada sobre la variación a nivel de ARNs reguladores, que son muy abundantes en este organismo (Arnvig *et al.*, 2011).

De manera similar, la epigenética es un campo emergente que recientemente se ha propuesto como una vía para descubrir los biomarcadores derivados del hospedero para la tuberculosis (Esterhuyse *et al.*, 2012). Sin embargo, poco se ha

hecho sobre la epigenética en MTB, y se desconoce si la variabilidad se reflejará en el nivel epigenético también. Hoy se dedica mucho esfuerzo al desarrollo de nuevos diagnósticos, fármacos y vacunas contra la tuberculosis (Young *et al.*, 2008).

Basado en las pruebas existentes de otros patógenos bacterianos, el estudio de la diversidad genética de MTB debe considerarse durante el desarrollo de nuevas herramientas y estrategias para mejorar el control de la tuberculosis. Cada vez son más, los enfoques (Comas *et al.*, 2011), no sólo para determinar si y cómo la diversidad de MTB es importante para el control de la enfermedad, sino también para comprender mejor la biología y la epidemiología de una de las enfermedades de salud pública más importantes.

Con el aumento de los genomas subidos en las bases de datos globales, los estudios futuros obtendrán un mejor apoyo en la asignación de genotipos basados en SNPs, que puede ser incorporado en ensayos de diagnóstico y asociaciones genotipo-fenotipo. La adición de la firma genómica junto con marcadores de resistencia a fármacos mejorará el control de la TB.

## 7. CONCLUSIONES

- Se propone el empleo de un conjunto de 10 SNPs seleccionados mediante la combinación de los criterios de inclusión/exclusión basado en datos de *Spoligotyping* y análisis filogenético y además de la selección de los SNPs no sinónimos presentes en los clúster de grupos ortólogos (COG) más conservados, utilizando todas las secuencias genómicas de 249 aislados de MTB.
- La asignación del genotipo utilizando el nuevo conjunto de 10 SNPs fue validado con 34 genomas de MTB adicionales y los resultados mostraron una correlación del 100% con sus genotipos ya conocidos. Nuestro conjunto de 10 SNPs no ha sido reportado previamente y abarca los genotipos de MTB que son prevalentes en todo el mundo.
- El nuevo conjunto de 10 SNPs propuesto tiene la flexibilidad de asociarse con las bases de datos globales donde los genotipos se basan en el método *Spoligotyping*, el actual *Gold standard*, determinando su alta eficiencia de genotipificación.
- Los 10 SNPs seleccionados se consideran como la firma genómica para genotipificación de *Mycobacterium tuberculosis*.

## 8. RECOMENDACIONES

- Se recomienda continuar los estudios epidemiológicos moleculares sobre tuberculosis como una herramienta de vigilancia dentro de la población mundial.
- El análisis y la discusión de los métodos moleculares de genotipificación siempre deben de complementarse con los datos epidemiológicos de cada uno de los pacientes.
- Se recomienda estandarizar el proceso de ensamblaje del genoma de *Mycobacterium tuberculosis* con la cepa de referencia H37Rv, la cual es pansensible a antibióticos antituberculosos y cuyo patrón de genotipificación constante es reportado por las bases de datos mundiales.
- Al utilizar bases de datos internacionales se deben tener en cuenta las diferentes variables e información de los pacientes con el fin de establecer relaciones entre dichas variables con los genotipos particulares.
- Implementar la automatización de este método de genotipificación para hacer más eficiente y económico el proceso.
- Desarrollar un estudio sobre epidemiología molecular de cepas de *Mycobacterium tuberculosis* que circulan en el Perú usando el conjunto de 10 SNPs relacionados y compararlos con el método *Gold standard*.

## 9. REFERENCIAS BIBLIOGRÁFICAS

- ACHTMAN, M. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol.* 2008;62:53–70.
- ACOSTA, R.; ESTRADA G.; MILIÁN F. Genotyping methods for *Mycobacterium bovis*. Review. *Técnica Pecuaria en México.* 2009;47(4):389-412.
- AGUILAR, L.; HANEKOM, M.; MATA, D.; PITTIUS, N.; HELDEN, P.; WARREN, R.; HERNANDEZ, R. *Mycobacterium tuberculosis* strains with the Beijing genotype demonstrate variability in virulence associated with transmission. *Tuberculosis.* 2010;90:319-325
- ALBANNA, A.; REED, M.; KOTAR, K.; FALLOW, A.; MCINTOSH, F.; BEHR, M.; MENZIES, D. Reduced transmissibility of East African Indian strains of *Mycobacterium tuberculosis*. *PLoS One.* 2011;6(9): p.e25075.
- ARISTIMUÑO, L.; ARMENGOL, R.; CEBOLLADA, A.; ESPAÑA, M.; GUILARTE, A.; LAFOZ, C.; LEZCANO, M.; REVILLO, M.; MARTÍN, C.; RAMÍREZ, C.; RASTOGI, N.; ROJAS, J.; DE SALAS, A.; SOLA, C.; SAMPER, S. Molecular characterization of *Mycobacterium tuberculosis* isolates in the First National Survey of Anti-tuberculosis Drug Resistance from Venezuela. *BMC Microbiology.* 2006;6:90.
- ARNVIG, K.; COMAS, I.; THOMSON, N.; HOUGHTON, J.; BOSHOFF, H.; CROUCHER, N.; ROSE, G.; PERKINS, T.; PARKHILL, J.; DOUGAN, G.; YOUNG, D. Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of *Mycobacterium tuberculosis*. *PLoS Pathog.* 2011;7:e1002342.
- ASENCIOS, L.; GALARZA, M.; QUISPE, N.; VÁSQUEZ, L.; LEO, E.; VALENCIA, E.; RAMÍREZ, J.; ACURIO, M.; SALAZAR, R.; MENDOZA, A.; CÁCERES, O.

Prueba molecular Genotype® MTBDRplus, una alternativa para la detección rápida de tuberculosis multidrogorresistente. Rev Peru Med Exp Salud Pública. 2012;29(1):92-8.

- BAI, X.; FELDMAN, N.; CHMURA, K.; OVRUTSKY, A.; SU, W.; GRIFFIN, L.; PYEON, D.; MCGIBNEY, M.; STRAND, M.; NUMATA, M.; MURAKAMI, S.; GAIDO, L.; HONDA, J.; KINNEY, W.; OBERLEY, R.; VOELKER, D.; ORDWAY, D.; CHAN, E. Inhibition of Nuclear Factor-Kappa B Activation Decreases Survival of *Mycobacterium tuberculosis* in Human Macrophages. PLoS One. 2013,8(4):e61925.
- BAKER, L.; BROWN, T.; MAIDEN, M.; DROBNIIEWSKI, F. Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. Emerg Infect Dis. 2004;10:1568–1577.
- BARNES, P.; CAVE, M. Molecular epidemiology of tuberculosis. The New England journal of medicine. 2003;349(12):1149-56.
- BARÓN, B.; SOTO, V.; ROSARIO, M. Validez del método concentrado en esputo con hipoclorito de sodio para el diagnóstico de tuberculosis en pacientes con radiografía anormal y baciloscopía negativa. Rev Cuerpo Médico HNAAA. 2012;5(2):15-18.
- BEGGS, M.; CAVE, M.; MARLOWE, C.; CLONEY, L.; DUCK, P.; EISENACH, K. Characterization of *Mycobacterium tuberculosis* complex direct repeat sequence for use in cycling probe reaction. Journal of clinical microbiology. 1996;34(12):2985-9.
- BENTLEY, D.; BALASUBRAMANIAN, S.; SWERDLOW, H.; SMITH, G.; MILTON, J.; BROWN, C.; HALL, K.; EVERS, D.; BARNES, C.; BIGNELL, H.; BOUTELL, J.; BRYANT, J.; CARTER, R.; KEIRA, R.; COX, A.; ELLIS, D.; FLATBUSH, M.; GORMLEY, N.; HUMPHRAY, S.; IRVING, L.; KARBELASHVILI, M.; KIRK, S.; LI, H.; LIU, X.; MAISINGER, K.; MURRAY, L.; OBRADOVIC, B.;

OST, T.; PARKINSON, M.; PRATT, M.; RASOLONJATOVO, I.; REED, M.; RIGATTI, R.; RODIGHIERO, C.; ROSS, M.; SABOT, A.; SANKAR, S.; SCALLY, A.; SCHROTH, G.; SMITH, M.; SMITH, V.; SPIRIDOU, A.; TORRANCE, P.; TZONEV, S.; VERMAAS, E.; WALTER, K.; WU, X.; ZHANG, L.; ALAM, M.; ANASTASI, C.; ANIEBO, I.; BAILEY, D.; BANCARZ, I.; BANERJEE, S.; BARBOUR, S.; BAYBAYAN, P.; BENOIT, V.; BENSON, K.; BEVIS, C.; BLACK, P.; BOODHUN, A.; BRENNAN, J.; BRIDGHAM, J.; BROWN, R.; BROWN, A.; BUERMANN, D.; BUNDU, A.; BURROWS, J.; CARTER, N.; CASTILLO, N.; CHIARA, E.; CATENAZZI, M.; CHANG, S.; NEIL, R.; CRAKE, N.; DADA, O.; DIAKOUMAKOS, K.; DOMINGUEZ, B.; EARNSHAW, D.; EGBUJOR, U.; ELMORE, D.; ETCHIN, S.; EWAN, M.; FEDURCO, M.; FRASER, L.; FUENTES, K.; SCOTT, W.; GEORGE, D.; GIETZEN, K.; GODDARD, C.; GOLDA, G.; GRANIERI, P.; GREEN, D.; GUSTAFSON, D.; HANSEN, N.; HARNISH, K.; HAUDENSCHILD, C.; HEYER, N.; HIMS, M.; HO, J.; HORGAN, A.; HOSCHLER, K.; HURWITZ, S.; IVANOV, D.; JOHNSON, M.; JAMES, T.; HUW, T.; KANG, G.; KERELSKA, T.; KERSEY, A.; KHREBTUKOVA, I.; KINDWALL, A.; KINGSBURY, Z.; KOKKO, P.; KUMAR, A.; LAURENT, M.; LAWLEY, C.; LEE, S.; LEE, X.; LIAO, A.; LOCH, J.; LOK, M.; LUO, S.; MAMMEN, R.; MARTIN, J.; MCCAULEY, P.; MCNITT, P.; MEHTA, P.; MOON, K.; MULLENS, J.; NEWINGTON, T.; NING, Z.; LING, B.; NOVO, S.; O'NEILL, M.; OSBORNE, M.; OSNOWSKI, A.; OSTADAN, O.; PARASCHOS, L.; PICKERING, L.; PIKE, A.; PIKE, A.; CHRIS, D.; PLISKIN, D.; PODHASKY, J.; QUIJANO, V.; RACZY, C.; RAE, V.; RAWLINGS, S.; CHIVA, A.; ROE, P.; ROGERS, J.; ROBERT, M.; ROMANOV, N.; ROMIEUA, R.; ROURKE, N.; RUEDIGER, S.; RUSMAN, E.; SANCHES, R.; SCHENKER, M.; SEOANE, J.; SHAW, R.; SHIVER, M.; SHORT, S.; SIZTO, N.; SLUIS, J.; SMITH, M.; ERNEST, J.; SPENCE, E.; STEVENS, K.; SUTTON, N.; SZAJKOWSKI, L.; TREGIDGO, C.; TURCATTI, G.; VANDEVONDELE, S.; VERHOVSKY, Y.; VIRK, S.; WAKELIN, S.; WALCOTT, G.; WANG, J.; WORSLEY, G.; YAN, J.; YAU, L.; ZUERLEIN, M.; ROGERS, J.; MULLIKIN, J.; HURLES, M.; MCCOOKE, N.; WEST, J.; OAKS, F.; LUNDBERG, P.; KLENERMAN, D.; DURBIN, R.; SMITH, A. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008;456(7218):53-9.

- BOLOTIN, A.; QUINQUIS, B.; SOROKIN, A.; EHRLICH, S. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*. 2005;2551-2561.
- BROSCHE, R.; GORDON, S.; MARMIESSE, M.; BRODIN, P.; BUCHRIESER, C.; EIGLMEIER, K.; GARNIER, T.; GUTIERREZ, C.; HEWINSON, G.; KREMER, K.; PARSONS, L.; PYM, A.; SAMPER, S.; VAN SOOLINGEN, D.; COLE, S. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl Acad. Sci. USA*. 2002; 99(6):3684–3689.
- BRUCHFELD, J.; CORREIA, M.; KALLENIIUS, G. Tuberculosis and HIV Coinfection. *Cold Spring Harbor perspectives in medicine*. 2015;5(7):a017871.
- BRUDEY, K.; DRISCOLL, J.; RIGOUTS, L.; PRODINGER, W.; GORI, A.; AL-HAJOJ, S.; ALLIX, C.; ARISTIMUÑO, L.; ARORA, J.; BAUMANIS, V.; BINDER, L.; CAFRONE, P.; CATALDI, A.; CHEONG, S.; DIEL, R.; ELLERMEIER, C.; EVANS, J.; FAUVILLE, M.; FERDINAND, S.; GARCIA DE VIEDMA, D.; GARZELLI, C.; GAZZOLA, L.; GOMES, H.; GUTTIEREZ, M.; HAWKEY, P.; VAN HELDEN, P.; KADIVAL, G.; KREISWIRTH, B.; KREMER, K.; KUBIN, M.; KULKARNI, S.; LIENS, B.; LILLEBAEK, T.; HO, M.; MARTIN, C.; MARTIN, C.; MOKROUSOV, I.; NARVSKAĪA, O.; NGEOW, Y.; NAUMANN, L.; NIEMANN, S.; PARWATI, I.; RAHIM, Z.; RASOLOFO, V.; RASOLONAVALONA, T.; ROSSETTI, M.; RÜSCH, S.; SAJDUDA, A.; SAMPER, S.; SHEMYAKIN, I.; SINGHM, U.; SOMOSKOVI, A.; SKUCE, R.; VAN SOOLINGEN, D.; STREICHER, E.; SUFFYS, P.; TORTOLI, E.; TRACEVSKA, T.; VINCENT, V.; VICTOR, T.; WARREN, R.; YAP, S.; ZAMAN, K.; PORTAELS, F.; RASTOGI, N.; SOLA, C. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international *spoligotyping* database (*SpolDB4*) for classification, population genetics and epidemiology. *BMC Microbiol*. 2006;6:23.
- CANDIA, N.; LOPEZ, B.; ZOZIO, T.; CARRIVALE, M.; DIAZ, C.; RUSSOMANDO, G.; DE ROMERO, N.; JARA, J.; BARRERA, L.; RASTOGI, N.; RITACCO, V. First



insight into *Mycobacterium tuberculosis* genetic diversity in Paraguay. BMC Microbiol. 2007;7:75.

- CASTIBLANCO, C; WLLMAN, R. Coinfección de tuberculosis en pacientes con VIH/SIDA: un análisis según las fuentes de información en Colombia. Asociación Colombiana de Infectología. 2006;10(4):232-42.
- CAVIEDES, L.; LEE, T.; GILMAN, R.; SHEEN, P.; SPELLMAN, E.; LEE, E.; BERG, D.; MONTENEGRO-JAMES, S. Rapid, efficient detection and drug susceptibility testing of *Mycobacterium tuberculosis* in sputum by microscopic observation of broth cultures. The Tuberculosis Working Group in Peru. J Clin Microbiol. 2000;38(3):1203-8.
- CENTERS FOR DISEASE CONTROL AND PREVENTION (CDC). Revised definition of extensively drug-resistant tuberculosis. MMWR Morb Mortal Wkly Rep. 2006;55:1176.
- CEREZO, I.; JIMÉNEZ, Y.; HERNANDEZ, J.; ZOZIO, T.; MURCIA, M.; RASTOGI, N. A first insight on the population structure of *Mycobacterium tuberculosis* complex as studied by *Spoligotyping* and MIRU-VNTRs in Bogotá, Colombia. Infect. Gen. and Evol. 2011.
- CHATURVEDI, R.; BANSAL, K.; NARAYANA, Y.; KAPOOR, N.; SUKUMAR, N.; TOGARSIMALEMATH, S.; CHANDRA, N.; MISHRA, S.; AJITKUMAR, P.; JOSHI, B.; KATOCH, V.; PATIL, S.; BALAJI, K. The multifunctional PE\_PGRS11 protein from *Mycobacterium tuberculosis* plays a role in regulating resistance to oxidative stress. J Biol Chem. 2010;285(40):30389-403.
- COLE, S.; BROSCHE, R.; PARKHILL, J.; GARNIER, T.; CHURCHER, C.; HARRIS, D.; GORDON, S.; EIGLMEIER, K.; GAS, S.; BARRY, C.; TEKAIA, F.; BADCOCK, K.; BASHAM, D.; BROWN, D.; CHILLINGWORTH, T.; CONNOR, R.;

DAVIES, R.; DEVLIN, K.; FELTWELL, T.; GENTLES, S.; HAMLIN, N.; HOLROYD, S.; HORNSBY, T.; JAGELS, K.; KROGH, A.; MCLEAN, J.; MOULE, S.; MURPHY, L.; OLIVER, K.; OSBORNE, J.; QUAIL, M.; RAJANDREAM, M.; ROGERS, J.; RUTTER, S.; SEEGER, K.; SKELTON, J.; SQUARES, R.; SQUARES, S.; SULSTON, J.; TAYLOR, K.; WHITEHEAD, S.; BARRELL, B. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*. 1998;393(6685):537-544.

- COLE, S. Comparative and functional genomics of the *Mycobacterium tuberculosis* complex. *Microbiology*. 2002;148(Pt 10):2919-28.
- COLL, F.; MCNERNEY, R.; GUERRA, J.; GLYNN, J.; PERDIGÃO, J.; VIVEIROS, M.; PORTUGAL, I.; PAIN, A.; MARTIN, N.; CLARK, T. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat Commun*. 2014;5:4812.
- COMAS, I.; GAGNEUX, S. The past and future of tuberculosis research. *PLoS Pathog*. 2009;5:e1000600.
- COMAS, I.; BORRELL, S.; ROETZER, A.; ROSE, G.; MALLA, B.; KATO, M.; GALAGAN, J.; NIEMANN, S.; GAGNEUX, S. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet*. 2011;44(1):106-10.
- COMAS, M.; COSCOLLA, S.; GAGNEUX, S. Two new rapid SNP-typing methods for classifying *Mycobacterium tuberculosis* complex into the main phylogenetic lineages. *PLoS One*. 2012;7(7):e41253.
- CORTINAS, M.; FERNANDEZ, M.; VALETA, M.; URIARTE, M.; MOGDSY, M. Caracterización genotípica de 80 cepas del género *Mycobacterium* en Uruguay. *Revista Médica del Uruguay*. 2002;18(3):230-8.

- COSCOLLA, M.; GAGNEUX, S. Does *M. tuberculosis* genomic diversity explain disease diversity? Drug discovery today Disease mechanisms. 2010;7(1):e43-e59.
- DANIEL, T.; SIPPOLA, A.; OKWERA, A.; KABENGERA, S.; HATANGA, E.; AISU, T.; NYOLE, S.; BYEKWASO, F.; VJECHA, M.; FERGUSON, L.; KATAAHA, P.; MUGERWA, R. Reduced sensitivity of tuberculosis serodiagnosis in patients with AIDS in Uganda. Tuberc Lung Dis. 1994;75:33-37.
- DASGUPTA, K.; MENZIES, D. Cost-effectiveness of tuberculosis control strategies among immigrants and refugees. Eur Respir J. 2005;25(6):1107-1116.
- DEMAY, C.; LIENS, B.; BURGUIÈRE, T.; HILL, V.; COUVIN, D.; MILLET, J.; MOKROUSOV, I.; SOLA, C.; ZOZIO, T.; RASTOGI, N. SITVITWEB—a publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. Infect. Genet. Evol. 2012;12:755–766.
- DE PAZ, D.; POTES, L.; QUIÑONEZ, E.; MARTINEZ, L.; CEPEDA, M.; GUARÍN, N.; VELEZ, J.; ROSSO, F.; GARCIA, J. Resistencia a fármacos antituberculosis en pacientes coinfectados con tuberculosis y virus de la inmunodeficiencia humana, en un hospital de referencia de 2007 a 2010 en Cali (Colombia). Infectio. 2012;16(3):161–5.
- DONG, H.; LIU, Z.; LV, B.; ZHANG, Y.; LIU, J.; ZHAO, X.; LIU, J.; WAN, K. Spoligotypes of *Mycobacterium tuberculosis* from different Provinces of China. J Clin Microbiol 2010;48(11): 4102-4106.
- DORRONSORO, I.; TORROBA, L. Microbiología de la tuberculosis. Anales del Sistema Sanitario de Navarra. 2007;30:67-84.

- DRISCOLL, J. *Spoligotyping* for molecular epidemiology of the *Mycobacterium tuberculosis* complex. *Methods Mol Biol.* 2009;551:117-128.
- DUCKWORTH, B.; NELSON, K.; ALDRICH, C. Adenylating enzymes in *Mycobacterium tuberculosis* as drug targets. *Curr Top Med Chem.* 2012;12(7):p.766-96.
- DUNPHY, K.; SENARATNE, R.; MASUZAWA, M.; KENDALL, L.; RILEY, L. Attenuation of *Mycobacterium tuberculosis* functionally disrupted in a fatty acyl-coenzyme A synthetase gene *fadD5*. *J Infect Dis.* 2010;201(8):p.1232-9.
- ESTERHUYSE, M.; LINHART, H.; KAUFMANN, S. Can the battle against tuberculosis gain from epigenetic research? *Trends Microbiol.* 2012;20:220–226
- FENTON, M.; VERMEULEN, M. Immunopathology of Tuberculosis: Roles of Macrophages and Monocytes. *Infect Immun.* 1996;64(3):683-690.
- FILLIOL, I.; DRISCOLL, J.; VAN SOOLINGEN, D.; KREISWIRTH, B.; KREMER, K.; VALÉTIÉDIE, G.; ANH, D.; BARLOW, R.; BANERJEE, D.; BIFANI, P.; BRUDEY, K.; CATALDI, A.; COOKSEY, R.; COUSINS, D.; DALE, J.; DELLAGOSTIN, O.; DROBNIEWSKI, F.; ENGELMANN, G.; FERDINAND, S.; GASCOYNE, D.; GORDON, M.; GUTIERREZ, M.; HAAS, W.; HEERSMA, H.; KÄLLENIUS, G.; KASSA, E.; KOIVULA, T.; LY, H.; MAKRISTATHIS, A.; MAMMINA, C.; MARTIN, G.; MOSTRÖM, P.; MOKROUSOV, I.; NARBONNE, V.; NARVSKAYA, O.; NASTASI, A.; NIOBE, S.; PAPE, J.; RASOLOFO, V.; RIDELL, M.; ROSSETTI, M.; STAUFFER, F.; SUFFYS, P.; TAKIFF, H.; TEXIER, J.; VINCENT, V.; DE WAARD, J.; SOLA, C.; RASTOGI, N. Global distribution of *Mycobacterium tuberculosis* spoligotypes. *Emerging infectious diseases.* 2002;8(11):1347-9.
- FILLIOL, I.; MOTIWALA, A.; CAVATORE, M.; QI, W.; HAZBÓN, M.; BOBADILLA, M.; FYFE, J.; GARCÍA, L.; RASTOGI, N.; SOLA, C.; ZOZIO, T.; GUERRERO, M.; LEÓN, C.; CRABTREE, J.; ANGIUOLI, S.; EISENACH, K.; DURMAZ, R.;

JOLOBA, M.; RENDÓN, A.; SIFUENTES, J.; PONCE DE LEÓN, A.; CAVE, M.; FLEISCHMANN, R.; WHITTAM, T.; ALLAND, D. 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.* 2006;188:759-772.

- FLEISCHMANN, R.; ALLAND, D.; EISEN, J.; CARPENTER, L.; WHITE, O.; PETERSON, J.; FRASER, C. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol.* 2002;184(19):5479-90.
- FORD, C.; SHAH, R.; MAEDA, M.; GAGNEUX, S.; MURRAY, M.; COHEN, T.; JOHNSTON, J.; GARDY, J.; LIPSITCH, M.; FORTUNE, S. *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat Genet.* 2013;45(7):784-90.
- FUJIKI, A.; KATO, S. Human resource capacity building on TB laboratory work for TB control program--through the experience of international TB laboratory training course for TB control at the Research Institute of Tuberculosis, JATA, Japan. *Kekkaku.* 2008;83(6):475-9.
- GAGNEUX, S. Genetic diversity in *Mycobacterium tuberculosis*. *Curr Top Microbiol Immunol.* 2013;374:1-25.
- GALARZA, M.; TARAZONA, D.; BORDA, V.; AGAPITO, J.; GUIO, H. Evidence of clonal expansion in the genome of a multidrug-resistant *Mycobacterium tuberculosis* clinical isolate from Peru. *Genome Announc.* 2014;2(1)00089-14.

- GARCÍA, B.; MONDRAGÓN, R.; LUNA, J. Internalization of *Mycobacterium tuberculosis* by macropinocytosis in non-phagocytic cells. *Microb Pathog.* 2003;35(2):49-55.
- GARDNER, S.; HALL, B. When whole-genome alignments just won't work: *kSNP* v2 software for alignment-free SNP discovery and phylogenetics of hundreds of microbial genomes. *PLoS one.* 2013;8(12):e81760.
- GARDNER, S.; SLEZAK, T.; HALL, B. *kSNP3.0*: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics.* 2015;31(17):2877-8.
- GHEBREMICHAEL, S.; PETERSSON, R.; KOIVULA, T.; PENNHAG, A.; ROMANUS, V.; BERGGREN, I.; PETRINI, B.; HOFFNER, S.; KÄLLENIUS, G. Molecular epidemiology of drug-resistant tuberculosis in Sweden. *Microbes and Infection.* 2008;10:699-705.
- GOLDBLATT, D.; RORMAN, E.; CHEMTOB, D.; FREIDLIN, P.; CEDAR, N.; KAIDAR, H.; DVEYRIN, Z.; MOR, Z. Molecular epidemiology and mapping of tuberculosis in Israel: do migrants transmit the disease to locals?. *Int J Tuberc Lung Dis.* 2014;18(9):1085-1091.
- GORI, A.; BANDERA, A.; MARCHETTI, G.; DEGLI, A.; CATOZZI, L.; NARDI, G.; GAZZOLA, L.; FERRARIO, G.; VAN EMBDEN, J.; VAN SOOLINGEN, D.; MORONI, M.; FRANZETTI, F. *Spoligotyping and Mycobacterium tuberculosis.* *Emerg Infect Dis.* 2005;11(8):p.1242-8.
- GOUJON, C.; SCHNEIDER, V.; GROFTI, J.; MONTIGNY, J.; JEANTILS, V.; ASTAGNEAU, P.; ROZENBAUM, W.; LOT, F.; FROCRAIN-HERCHKOVITCH, C.; DELPHIN, N.; LE GAL, F.; NICOLAS, J.; MILINKOVITCH, M.; DENY, P. Phylogenetic analyses indicate an atypical nurse-to-patient transmission of human immunodeficiency virus type 1. *J. Virol.* 2000;74:2525–2532.

- GROENEN, P.; BUNSCHOTEN, A.; VAN SOOLINGEN, D.; VAN EMBDEN, J. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Molecular microbiology*. 1993;10(5):1057-65.
- GUTACKER, M.; MATHEMA, H.; SOINI, E.; SHASHKINA, B.; KREISWIRTH, N.; GRAVISS, E.; MUSSER, J. Single-nucleotide polymorphism based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. *J. Infect. Dis.* 2006;193:121–128.
- HAIN LIFESCIENCE. GmbH GenoType MTBDRplus 1.0 product insert. Hain Lifescience GmbH, Nehren, Germany. 2010.
- HARISMENDY, O.; NG, P.; STRAUSBERG, R.; WANG, X.; STOCKWELL, T.; BEESON, K.; SCHORK, N.; MURRAY, S.; TOPOL, E.; LEVY, S.; FRAZER, K. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol.* 2009;10(3):R32.
- HAWKEY, P.; SMITH, E.; EVANS, J.; MONK, P.; BRYAN, G.; MOHAMED, H.; BARDHAN, M.; PUGH, R. Mycobacterial interspersed repetitive unit typing of *Mycobacterium tuberculosis* compared to IS6110-based restriction fragment length polymorphism analysis for investigation of apparently clustered cases of tuberculosis. *J Clin Microbiol.* 2003;41(8):3514-20.
- HERSHBERG, R.; LIPATOV, M.; SMALL, P.; SHEFFER, H.; NIEMANN, S.; HOMOLKA, S.; ROACH, J.; KREMER, K.; PETROV, D.; FELDMAN, M.; GAGNEUX, S. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol.* 2008;6:e311.
- HOMOLKA, S.; POST, E.; OBERHAUSER, B.; GARAWANI, A.; WESTMAN, L.; DAFAR, F.; RÜSCH, S.; NIEMANN, S. High genetic diversity among

*Mycobacterium tuberculosis* complex strains from Sierra Leone. BMC Microbiology. 2008;8:103.

- HOMOLKA, S.; PROJAHN, M.; FEUERRIEGEL, S.; UBBEN, T.; DIEL, R.; NÜBEL, U.; NIEMANN, S. High Resolution Discrimination of Clinical *Mycobacterium tuberculosis* Complex Strains Based on Single Nucleotide Polymorphisms. PLoS ONE. 2012;7(7),e39855.
- HONG, X.; HOPFINGER, A. Molecular modeling and simulation of *Mycobacterium tuberculosis* cell wall permeability. Biomacromolecules. 2004;5(3):1066-77.
- HUANG, Q.; LUO, H.; LIU, M.; ZENG, J.; ABDALLA, A.; DUAN, X.; LI, Q.; XIE, J. The effect of *Mycobacterium tuberculosis* CRISPR-associated Cas2 (Rv2816c) on stress response genes expression; morphology and macrophage survival of *Mycobacterium smegmatis*. Infect Genet Evol. 2016;40:p.295-301.
- HUEBNER, R.; SCHEIN, M.; BASS, J. The tuberculin skin test. Clin Infect Dis. 1993;17(6):968-75.
- ILINA, E.; SHITIKOV, E.; IKRYANNIKOVA, L.; ALEKSEEV, D.; KAMASHEV, D.; MALAKHOVA, M.; PARFENOVA, T.; AFANAS'EV, M.; ISCHENKO, D.; BAZALEEV, N.; SMIRNOVA, T.; LARIONOVA, E.; CHERNOUSOVA, L.; BELETSKY, A.; MARDANOV, A.; RAVIN, N.; SKRYABIN, K.; GOVORUN, V. Comparative genomic analysis of *Mycobacterium tuberculosis* drug resistant strains from Russia. PLoS One. 2013;Epub2013Feb20.
- JAGIELSKI, T.; AUGUSTYNOWICZ, E.; ZOZIO, T.; RASTOGI, N.; ZWOLSKA, Z. Spoligotype-Based Comparative Population Structure Analysis of Multidrug-Resistant and Isoniazid-Monoresistant *Mycobacterium tuberculosis* Complex Clinical Isolates in Poland. Journal of Clinical Microbiology. 2010;48:3899-3909.



- JEFFREYS, A.; WILSON, V.; THEIN, S. Hypervariable 'minisatellite' regions in human DNA. *Nature*. 1985;314(6006):67-73.
- JEFFROY, O.; BRINKMANN, H.; DELSUC, F.; PHILIPPE, H. Phylogenomics: the beginning of incongruence? *Trends Genet.* 2006;22(4):p.225-31.
- JOHNSON, M.; CARPENTER, E.; TIAN, Z.; BRUSKIEWICH, R.; BURRIS, J.; CARRIGAN, C.; CHASE, M.; CLARKE, N.; COVSHOFF, S.; DEPAMPHILIS, C.; EDGER, P.; GOH, F.; GRAHAM, S.; GREINER, S.; HIBBERD, J.; JORDON, I.; KUTCHAN, T.; LEEBENS, J.; MELKONIAN, M.; MILES, N.; MYBURG, H.; PATTERSON, J.; PIRES, J.; RALPH, P.; ROLF, M.; SAGE, R.; SOLTIS, D.; SOLTIS, P.; STEVENSON, D.; STEWART, C.; SUREK, B.; THOMSEN, C.; VILLARREAL, J.; WU, X.; ZHANG, Y.; DEYHOLOS, M.; WONG, G. Evaluation Methods for Isolating Total RNA and Predicting the Success of Sequencing Phylogenetically Diverse Plant Transcriptomes. *PLoS One*. 2012;7(11):e50226.
- KAMERBEEK, J.; SCHOOLS, L.; KOLK, A.; VAN AGTERVELD, M.; VAN SOOLINGEN, D.; KUIJPER, S.; BUNSCHOTEN, A.; MOLHUIZEN, H.; SHAW, R.; GOYAL, M.; VAN EMBDEN, J. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol.* 1997;35(4):907-14.
- KANDUMA, E.; MCHUGH, T.; GILLESPIE, S. Molecular methods for *Mycobacterium tuberculosis* strain typing: a users guide. *J. Appl. Microbiol.* 2003;94:781–791.
- KARBOUL, A.; MAZZA, A.; GEY VAN PITTIUS, N.; HO, J.; BROUSSEAU, R.; MARDASSI, H. Frequent homologous recombination events in *Mycobacterium tuberculosis* PE/PPE multigene families: potential role in antigenic variability. *J Bacteriol.* 2008;190(23):7838-46.
- KARLIN, S y BURGE, C. Dinucleotide relative abundance extremes: a genomic signature. *Trends in Genetics.* 1995;11(7):283-290.

- KASPER, D.; BRAUNWALD, E.; FAUCI, A. Tuberculosis. Harrison's Principles of Internal Medicine. 2005:953-66.
- KATO-MAEDA, M.; GAGNEUX, S.; FLORES, L.; KIM, E.; SMALL, P.; DESMOND, E.; HOPEWELL, P. Strain classification of *Mycobacterium tuberculosis*: congruence between large sequence polymorphisms and spoligotypes. Int J Tuberc Lung Dis. 2011;15(1):131-3.
- KATO-MAEDA, M.; HO, C.; PASSARELLI, B.; BANAEI, N.; GRINSDALE, J.; FLORES, L.; ANDERSON, J.; MURRAY, M.; ROSE, G.; KAWAMURA, L.; POURMAND, N.; TARIQ, M.; GAGNEUX, S.; HOPEWELL, P. Use of whole genome sequencing to determine the microevolution of *Mycobacterium tuberculosis* during an outbreak. PLoS One. 2013;8(3):e58235.
- KAUFMANN, S.; WALKER, B. AIDS and Tuberculosis. Federal Republic of Germany: WILEY-VCH Verlag GmbH & Co. KGaA. 2009;322 p.
- KEIM, P.; KLEVYTSKA, A.; PRICE, L.; SCHUPP, J.; ZINSER, G.; SMITH, K.; HUGH, M.; OKINAKA, R.; HILL, K.; JACKSON, P. Molecular diversity in *Bacillus anthracis*. Journal of applied microbiology. 1999;87(2):215-7.
- KLEVYTSKA, A.; PRICE, L.; SCHUPP, J.; WORSHAM, P.; WONG, J.; KEIM, P. Identification and characterization of variable-number tandem repeats in the *Yersinia pestis* genome. Journal of clinical microbiology. 2001;39(9):3179-85.
- KOHLI, S.; SINGH, Y.; KHUSHBU, S.; MITTAL, A.; NASREEN, Z.; SEYES, E. Comparative genomic and proteomic analyses of PE/PPE multigene family of *Mycobacterium tuberculosis* H37Rv and H37Ra reveal novel and interesting differences with implications in virulence. Nucleic Acids Res. 2012;40(15):7113–7122.

- KREMER, K.; VAN SOOLINGEN, D.; FROTHINGHAM, R.; HAAS, W.; HERMANS, P.; MARTÍN, C.; PALITTAPONGARNPIM, P.; PLIKAYTIS, B.; RILEY, L.; YAKRUS, M.; MUSSER, J.; VAN EMBDEN, J. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J Clin Microbiol. 1999;37(8):2607-18.
- KWARA, A.; SCHIRO, R.; COWAN, L.; HYSLOP, N.; WISER, M.; ROAHEN, S.; KISSINGER, P.; DIEM, L.; CRAWFORD, J. Evaluation of the epidemiologic utility of secondary typing methods for differentiation of *Mycobacterium tuberculosis* isolates. J Clin Microbiol. 2003;41(6):2683-5.
- LALVANI, A.; PATHAN, A.; MCSHANE, H.; WILKINSON, R.; LATIF, M.; CONLON, C.; PASVOL, G.; HILL, A. Rapid detection of *Mycobacterium tuberculosis* infection by enumeration of antigen-specific T cells. Am J Respir Crit Care Med. 2001;163(4):824-8.
- LE FLECHE, P.; HAUCK, Y.; ONTENIENTE, L.; PRIEUR, A.; DENOEUDE, F.; RAMISSE, V.; SYLVESTRE, P.; BENSON, G.; RAMISSE, F.; VERGNAUD, G. A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*. BMC microbiology. 2001;1-2.
- LEÓN, C.; GUERRERO, M.; NARANJO, N.; GARZÓN, M.; SIERRA, C. Descripción de la población tuberculosa colombiana (1999-2000). Infectio. 2002;6:84.
- LI, H.; DURBIN, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60.

- LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNELL, T.; RUAN, J.; HOMER, N.; MARTH, G.; ABECASIS, G.; DURBIN, R. The Sequence Alignment/Map format and *SAMtools*. *Bioinformatics*. 2012;25(16):2078-9.
- LI, J.; CHAI, Q.; ZHANG, Y.; LI, B.; WANG, J.; QIU, X.; LIU, C. *Mycobacterium tuberculosis* Mce3E Suppresses Host Innate Immune Responses by Targeting ERK1/2 Signaling. *J Immunol*. 2015;194(8):p.3756-67.
- LIU, L.; PEARL, D.; BRUMFIELD, R.; EDWARDS, S. Estimating species trees using multiple-allele DNA sequence data. *Evolution*. 2008;62(8):p.2080-91.
- LIU, L.; LI, Y.; LI, S.; HU, N.; HE, Y.; PONG, R.; LIN, D.; LU, L.; LAW, M. Comparison of Next-Generation Sequencing Systems. *J Biomed Biotechnol*. 2012;251364.
- LUO, M.; FADEEV, E.; GROVES, J. Mycobactin-mediated iron acquisition within macrophages. *Nat Chem Biol*. 2005;1(3):p.149-53.
- MACMANES, M.; EISEN, M. Improving transcriptome assembly through error correction of high-throughput sequence reads. *PeerJ*. 2013;1,e113.
- MAKAROVA, K.; GRISHIN, N.; SHABALINA, S.; WOLF, Y.; KOONIN, E. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct*. 2006;1-7.
- MAKAROVA, K.; HAFT, D.; BARRANGOU, R.; BROUNS, S.; CHARPENTIER, E.; HORVATH, P.; MOINEAU, S.; MOJICA, F.; WOLF, Y.; YAKUNIN, A.; VAN DER, J.; KOONIN, E. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol*. 2011;9(6):p.467-77.

- MARCAIS, G.; KINGSFORD, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 2011;27(6):764-70.
- MEDVEDEV, P.; SCOTT, E.; KAKARADOV, B.; PEVZNER, P. Error correction of high-throughput sequencing datasets with non-uniform coverage. *Bioinformatics*. 2011;27(13):i137-41.
- MERKER, M.; KOHL, T.; ROETZER, A.; TRUEBE, L.; RICHTER, E.; RÜSCH, S.; NIEMANN, S. Whole genome sequencing reveals complex evolution patterns of multidrug-resistant *Mycobacterium tuberculosis* Beijing strains in patients. *PLoS One*. 2013;8(12):e82551.
- MIDDLEBROOK, G.; REGGIARDO, Z.; TIGERTT, W. Automatable radiometric detection of growth of *Mycobacterium tuberculosis* in selective media. *Am. Rev. Respir. Dis.* 1977;115:1066–1069.
- MINSA: Análisis de la Situación Epidemiológica de la tuberculosis en el Perú, 2015. Ministerio de Salud, Dirección General de Epidemiología, 2016. Edición electrónica, Febrero 2016.
- MOKROUSOV, I. Genetic geography of *Mycobacterium tuberculosis* Beijing genotype: a multifacet mirror of human history?. *Infect Genet Evol.* 2008;8(6):777-785.
- MOLINA, C.; MORENO, E.; OCAMPO, J.; RENDON, A.; BLACKWOOD, K.; KREMER, K.; RASTOGI, N.; WELSH, O.; VERA, L. *Mycobacterium tuberculosis* Spoligotypes in Monterrey, México. *J. Clin. Microbiol.* 2010;48:448-455.
- MONTESERIN, J. Genotipos de *Mycobacterium tuberculosis* prevalentes en Argentina y su relación con mutaciones de resistencia a isoniácida (Tesis de posgrado). Universidad Nacional de Quilmes, Bernal, Argentina, 2016.

- MOORE, D.; EVANS, C.; GILMAN, R.; CAVIEDES, L.; CORONEL, J.; VIVAR, A.; SANCHEZ, E.; PIÑEDO, Y.; SARAVIA, J.; SALAZAR, C.; OBERHELMAN, R.; HOLLM-DELGADO, M.; LACHIRA, D.; ESCOMBE, A.; FRIEDLAND, J. Microscopic-observation drug-susceptibility assay for the diagnosis of TB. *N Engl J Med.* 2006;355(15):1539-50.
- MOSTRÖM, P.; GORDON, M.; SOLA, C.; RIDELL, M.; RASTOGI, N. Methods used in the molecular epidemiology of tuberculosis. *Clinical Microbiology and Infection.* 2002;8:694–704.
- MUKHOPADHYAY, S.; BALAJI, K. The PE and PPE proteins of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb).* 2011;91(5):p.441-7.
- MURRAY, P.; ROSENTHAL, K.; KOBAYASHI, G.; PFALLER, M. *Medical Microbiology.* En: 4th ed. Mosby, Inc., St. Louise, Missouri USA. 2002;366–368.
- NAKAMURA, Y. Molecular analyses of the serotype of *Cryptococcus neoformans*. *Nippon Ishinkin Gakkai Zasshi.* 2001;42:69–74.
- NICOL, M.; WILKINSON, R. The clinical consequences of strain diversity in *Mycobacterium tuberculosis*. *Trans R Soc Trop Med Hyg.* 2008;102(10):p.955-65.
- ODONE, A.; TILLMANN, T.; SANDGREN, A.; WILLIAMS, G.; RECHEL, B.; INGLEBY, D.; NOORI, T.; MLADOVSKY, P.; MCKEE, M. Tuberculosis among migrant populations in the European Union and the European Economic Area. *Eur J Public Health.* 2014.
- ORGANIZACIÓN PANAMERICANA DE LA SALUD. PAHO. Plan regional de Tuberculosis 2006-2015 Washington, D.C: OPS 2006 ISBN 92 75 32673 8.

- ORGANIZACIÓN PANAMERICANA DE LA SALUD. La tuberculosis en la Región de las Américas: Informe Regional. Epidemiología, control y financiamiento. Washington, DC: OPS, 2013.
- PARWATI, I.; VAN CREVEL, R.; VAN SOOLINGEN, D. Possible underlying mechanisms for successful emergence of the *Mycobacterium tuberculosis* Beijing genotype strains. *Lancet Infect Dis.* 2010;10(2):p.103-11.
- PASRICHA, R.; CHANDOLIA, A.; PONNAN, P.; SAINI, N.; SHARMA, S.; CHOPRA, M.; BASIL, M.; BRAHMACHARI, V.; BOSE, M. Single nucleotide polymorphism in the genes of *mce1* and *mce4* operons of *Mycobacterium tuberculosis*: analysis of clinical isolates and standard reference strains. *BMC Microbiol.* 2011;11:p.41.
- PEARSON, T.; BUSCH, J.; RAVEL, J.; READ, T.; RHOTON, S.; U'REN, J.; SIMONSON, T.; KACHUR, S.; LEADEM, R.; CARDON, M.; VAN ERT, M.; HUYNH, L.; FRASER, C.; KEIM, P. Phylogenetic S. Gagneux discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from wholegenome sequencing. *Proc Natl Acad Sci USA.* 2004;101:13536–13541.
- PETER, F.; BARNES, M.; DONALD, C. Molecular epidemiology of tuberculosis. *N Engl J Med.* 2003;349:1149-56.
- POURCEL, C.; SALVIGNOL, G.; VERGNAUD, G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology.* 2005;151(Pt 3): 653-663.

- RAGHEB, M.; FORD, C.; CHASE, M.; LIN, P.; FLYNN, J.; FORTUNE, S. The mutation rate of mycobacterial repetitive unit *loci* in strains of *M. tuberculosis* from cynomolgus macaque infection. BMC Genomics. 2013;14:145.
- RAMAZANZADEH, R.; SAYHEMIRI, K. Prevalence and occurrence rate of *Mycobacterium tuberculosis* Haarlem family multi-drug resistant in the worldwide population: A systematic review and meta-analysis. J Res Med Sci. 2015;20(1):p.78-88.
- RATH, D.; AMLINGER, L.; RATH, A.; LUNDGREN, M. The CRISPR-Cas immune system: biology, mechanisms and applications. Biochimie. 2015;117:p.119-28.
- RAVIGLIONE, M.; O'BRIEN, R. Mycobacterial diseases. Harrison's Principles of Internal Medicine. 2012:1340-59.
- REYES, F.; TIRADO, Y.; PUIG, A.; BORRERO, R.; REYES, G.; FERNÁNDEZ, S.; PÉREZ, J.; KADIR, R.; ZAYAS, C.; NORAZMI, M.; SARMIENTO, M.; ACOSTA, A. Immunogenicity and cross-reactivity against *Mycobacterium tuberculosis* of proteoliposomes derived from *Mycobacterium bovis* BCG. BMC Immunol. 2013;14(Suppl 1):S7.
- RINDI, L.; LARI, N.; CUCCU, B.; GARZELLI, C. Evolutionary pathway of the Beijing lineage of *Mycobacterium tuberculosis* based on genomic deletions and *mutT* genes polymorphisms. Infections, Genetics and Evolution. 2009;9:48-53.
- RODRIGUEZ, G. Control of iron metabolism in *Mycobacterium tuberculosis*. Trends Microbiol. 2006;14(7):320-7.
- RODRÍGUEZ, J.; RAMÍREZ, A.; SALAS, L.; HELGUERA, C.; GONZALEZ, J.; SOTO, C.; HERNÁNDEZ, R. Transcription of Genes Involved in Sulfolipid and



Polyacyltrehalose Biosynthesis of *Mycobacterium tuberculosis* in Experimental Latent Tuberculosis Infection. PLoS One. 2013;8(3):e58378.

- ROETZER, A.; DIEL, R.; KOHL, T.; RUCKERT, C.; NUBEL, U.; BLOM, J.; WIRTH, T.; JAENICKE, S.; SCHUBACK, S.; RUSCH, S.; SUPPLY, P.; KALINOWSKI, S.; NIEMANN, S. Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. PLoS Med. 2013;10(2):e1001387.
- ROSALES, S.; PIEDA, L.; GHEBREMICHAEL, S.; RASTOGI, N.; HOFFNER, S. Molecular diversity of *Mycobacterium tuberculosis* isolates from patients with tuberculosis in Honduras. BMC Microbiology. 2010;10:208.
- SANGER INSTITUTE. *Mycobacterium tuberculosis*. 2007-03-29. Retrieved 2008-11-16.
- SCHORK, N.; FALLIN, D.; LANCHBURY, J. Single nucleotide polymorphisms and the future of genetic epidemiology. Clin. Genet. 2000;58:250–264.
- SHABBEER, A.; OZCAGLAR, C.; YENER, B.; BENNETT, K. Web tools for molecular epidemiology of tuberculosis. Infect Genet Evol. 2012;12(4):767-81.
- SHARMA, J.; ROY, K.; PUSHPARAJ, M.; KUMAR, S.; MALHOTRA, N.; MITTAL, S. Laparoscopic findings in female genital tuberculosis. Arch Gynecol Obstet. 2008;278:359-364.
- SIMEONE, R.; BOTTAI, D.; FRIGUI, W.; MAJLESSI, L.; BROSCHE, R. ESX/type VII secretion systems of mycobacteria: Insights into evolution, pathogenicity and protection. Tuberculosis (Edinb). 2015.

- SMITH, I. *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. Clin Microbiol Rev. 2003;16(3):463-96.
- SOLA, C.; FILLIOL, I.; GUTIERREZ, M.; MOKROUSOV, I.; VINCENT, V.; RASTOGI, N. Spoligotype database of *Mycobacterium tuberculosis*: biogeographic distribution of shared types and epidemiologic and phylogenetic perspectives. Emerg Infect Dis. 2001;7(3):390-6.
- SOVIC´, I. *Approaches to DNA de novo Assembly*. (Tesis Doctoral). 2012. Faculty of Electrical Engineering and Computing: Ruder Boškovic Institute, Zagreb, Croacia.
- STREICHER, P.; SUFFYS, E.; TORTOLI, T.; TRACEVSKA, V.; VINCENT, T.; VICTOR, C.; WARREN, R.; YAP, S.; ZAMAN, K.; PORTAELS, F.; RASTOGI, N.; SOLA, C. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol. 2006;6:23.
- STREICHER, E.; VICTOR, T.; VAN DER SPUY, G.; SOLA, C.; RASTOGI, N.; VAN HELDEN, P.; WARREN, R. Spoligotype signatures in the *Mycobacterium tuberculosis* complex. J. Clin. Microbiol. 2007;45:237–240.
- STUCKI, D.; MALLA, B.; HOSTETTLER, S.; HUNA, T.; FELDMANN, J.; YEBOAH, D.; BORRELL, S.; FENNER, L.; COMAS, I.; COSCOLLA, M.; GAGNEUX, S. Two new rapid SNP-typing methods for classifying Genetic Diversity in *Mycobacterium tuberculosis* complex into the main phylogenetic lineages. PLoS ONE. 2012;7:e41253.
- SUPPLY, P.; MAGDALENA, J.; HIMPENS, S.; LOCHT, C. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. Molecular microbiology. 1997;26(5):991-1003.

- SUPPLY, P.; MAZARS, E.; LESJEAN, S.; VINCENT, V.; GICQUEL, B.; LOCHT, C. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. *Molecular microbiology*. 2000;36(3):762-71.
- SUPPLY, P.; LESJEAN, S.; SAVINE, E.; KREMER, K.; VAN SOOLINGEN, D.; LOCHT, C. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. *Journal of clinical microbiology*. 2001;39(10):3563-71.
- SUPPLY, P.; ALLIX, C.; LESJEAN, S.; CARDOSO, M.; RÜSCH, S.; WILLERY, E.; SAVINE, E.; DE HAAS, P.; VAN DEUTEKOM, H.; RORING, S.; BIFANI, P.; KUREPINA, N.; KREISWIRTH, B.; SOLA, C.; RASTOGI, N.; VATIN, V.; GUTIERREZ, M.; FAUVILLE, M.; NIEMANN, S.; SKUCE, R.; KREMER, K.; LOCHT, C.; VAN SOOLINGEN, D. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *Journal of clinical microbiology*. 2006;44(12):4498-510.
- TALARICO, S.; CAVE, M.; FOXMAN, B.; MARRS, C.; ZHANG, L.; BATES, J.; YANG, Z. Association of *Mycobacterium tuberculosis* PE PGRS33 polymorphism with clinical and epidemiological characteristics. *Tuberculosis (Edinb)*. 2008;87(4):338-46.
- TARAZONA, D.; BORDA, V.; GALARZA, M.; AGAPITO, J.; GUIO, H. Functional analysis using whole-genome sequencing of a drug-sensitive *Mycobacterium tuberculosis* strain from Peru. *Genome Announc*. 2014;00087-14.
- TATUSOV, R.; NATALE, D.; GARKAVTSEV, I.; TATUSOVA, T.; SHANKAVARAM, U.; RAO, B.; KIRYUTIN, B.; GALPERIN, M.; FEDOROVA, N.; KOONIN, E. The COG database: new developments in phylogenetic

classification of proteins from complete genomes. *Nucleic Acids Res.* 2001;29(1):22-28.

- TAYPE, C.; AGAPITO, J.; ACCINELLI, R.; ESPINOZA, J.; GODREUIL, S.; GOODMAN, S.; BAÑULS, A.; SHAW, M. Genetic diversity, population structure and drug resistance of *Mycobacterium tuberculosis* in Perú. *Infect Genet Evol.* 2012;12(3):577-85.
- TELENTI, A.; IMBODEN, P.; MARCHESI, F.; LOWRIE, D.; COLE, S.; COLSTON, M.; MATTER, L.; SCHOPFER, K.; BODMER, T. Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. *Lancet.* 1993;341(8846):647-50.
- THUMAMO, B.; ASUQUO, A.; ABIA, L.; LAWSON, L.; HILL, V.; ZOZIO, T.; EMENYONU, N.; EKO, F.; RASTOGI, N. Molecular epidemiology and genetic diversity of *Mycobacterium tuberculosis* complex in the Cross River State, Nigeria. *Infect. Gen. and Evol.* 2012.
- TIWARI, B.; KANNAN, N.; VEMU, L.; RAGHUNAND, T. The *Mycobacterium tuberculosis* PE proteins Rv0285 and Rv1386 modulate innate immunity and mediate bacillary survival in macrophages. *PLoS One.* 2012;7(12):e51686.
- TORTOLI, E.; BENEDETTI, M.; FONTANELLI, A.; SIMONETTI, M. Evaluation of automated BACTEC MGIT 960 system for testing susceptibility of *Mycobacterium tuberculosis* to four major antituberculous drugs: comparison with the radiometric BACTEC 460TB method and the agar plate method of proportion. *J Clin Microbiol.* 2002;40(2):607-10.
- TREANGEN, T.; ONDOV, B.; KOREN, S.; PHILLIPPY, A. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* 2014;15(11):524.

- VALCHEVA, V.; MOKROUSOV, I.; RASTOGI, N.; NARVSKAYA, O.; MARKOVA, N. Molecular characterization of *Mycobacterium tuberculosis* isolates from different regions of Bulgaria. J Clin Microbiol. 2008;46:1014-1018.
- VAN EMBDEN, J.; CAVE, M.; CRAWFORD, J.; DALE, J.; EISENACH, K.; GICQUEL, B.; HERMANS, P.; MARTIN, C.; MCADAM, R.; SHINNICK, T. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol. 1993;31:406-9.
- VAN EMBDEN, J.; VAN GORKOM, T.; KREMER, K.; JANSEN, R.; VAN DER ZEIJST, B.; SCHOOLS, L. Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria. J Bacteriol. 2000;182(9):2393-2401.
- VAN SOOLINGEN, D. Molecular epidemiology of tuberculosis and other mycobacterial infections: main methodologies and achievements. J Intern Med. 2001;249(1):1-26.
- VIEGAS, S.; MACHADO, A.; GROENHEIT, R.; GHEBREMICHAEL, S.; PENNHAG, A.; GUDO, P.; CUNA, Z.; MIOTTO, P.; HILL, V.; MARRUFO, T.; CIRILLO, D.; RASTOGI, N.; KÄLLENIUS, G.; KOIVULA, T. Molecular diversity of *Mycobacterium tuberculosis* isolates from patients with pulmonary tuberculosis in Mozambique. BMC Microbiology. 2010;10:195.
- WAMOTO, T.; GRANDJEAN, L.; ARIKAWA, K.; NAKANISHI, N.; CAVIEDES, L.; CORONEL, J.; SHEEN, P.; WADA, T.; TAYPE, C.; SHAW, M.; MOORE, D.; GILMAN, R. Genetic diversity and transmission characteristics of Beijing family strains of *Mycobacterium tuberculosis* in Peru. PLoS One. 2012;7(11):e49651.
- WARREN, R.; VICTOR, T.; STREICHER, E.; RICHARDSON, M.; BEYERS, N.; GEY VAN PITTIUS, N.; VAN HELDEN, P. Patients with active tuberculosis often

have different strains in the same sputum specimen. *Am J Respir Crit Care Med.* 2004;169(5):610-4.

- WENIGER, T.; KRAWCZYK, J.; SUPPLY, P.; NIEMANN, S.; HARMSSEN, D. MIRU-VNTRplus: a web tool for polyphasic genotyping of *Mycobacterium tuberculosis* complex bacteria. *Nucleic Acids Res.* 2010;38:326–331.
- WORLD HEALTH ORGANIZATION: Multidrug and extensively drug-resistant TB (M/XDR-TB) 2010 Global Report on Surveillance and Response.
- WORLD HEALTH ORGANIZATION: Global tuberculosis report. WHO Report 2014. Geneva, Switzerland.
- WORLD HEALTH ORGANIZATION. Global tuberculosis report 2015. Switzerland: World Health Organization, 2015.
- YOUNG, D.; PERKINS, M.; DUNCAN, K.; BARRY, C. Confronting the scientific obstacles to global control of tuberculosis. *J Clin Invest.* 2008;118:1255–1265
- ZAMMARCHI, L.; TORTOLI, E.; BORRONI, E; BARTALESI, F; STROHMEYER, M; BARETTI, S.; SIMONETTI, M.; LIENDO, C.; SANTINI, M.; ROSSOLINI, G.; GOTUZZO, E.; BARTOLONI, A. High prevalence of clustered tuberculosis cases in peruvian migrants in florence, Italy. *Infect Dis Rep.* 2014;6(4): 5646.
- ZHANG, H.; LI, D.; ZHAO, L.; FLEMING, J.; LIN, N.; WANG, T.; LIU, Z.; LI, C.; GALWEY, N.; DENG, J.; ZHOU, Y.; ZHU, Y.; GAO, Y.; WANG, T.; WANG, S.; HUANG, Y.; WANG, M.; ZHONG, Q.; ZHOU, L.; CHEN, T.; ZHOU, J.; YANG, R.; ZHU, G.; HANG, H.; ZHANG, J.; LI, F.; WAN, K.; WANG, J.; ZHANG, X.; BI, L. Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from China identifies genes and intergenic regions associated with drug resistance. *Nat Genet.* 2013;45(10):1255-60.

## **10. ANEXOS**

**ANEXO I:** 249 secuencias genómicas de *Mycobacterium tuberculosis* analizadas en este estudio

<b>Cepa</b>	<b>Número de accesión</b>	<b>País</b>	<b><i>Spoligotyping</i></b>	<b>Filogenia de Referencia</b>
210	NZ_ADAB01000000	USA	Beijing	-
293	SRR671865	China	Beijing	-
336	SRR671866	China	Beijing	-
1027	ERP001731	China	-	Linaje 2
1034	APHO01000001	China	-	Linaje 2
1336	ERP001731	China	-	Linaje 2
349404	ERP001731	Sierra Leone	-	Linaje 5
1047301	ERP001731	Ghana	-	Linaje 5
02-1987	NZ_DS985180	USA	Beijing	-
11051-03	SRR023466	Tailandia/Hamburgo	-	Linaje 1,3
11821-3	SRR022878	Alemania	-	-
12778-03	ERP001731	Vietnam/Hamburgo	-	Linaje 1,3
1558-08	ERP001731	Tanzania	-	Linaje 3
1571-08	ERP001731	NA	-	-
1571-99	ERP001731	Uganda	-	-
2173-99	ERP001731	Uganda	-	-
2197-99	SRP001127	Uganda	-	-
2318-06	ERP001731	Alemania	-	-
4141-04	SRR022880	Alemania	AFRI2	-
43-16836	ATNF00000000	Tailandia	-	Linaje 1,3
4783-04	SRR022877	Alemania	-	-
5444-04	SRR022879	Alemania	AFRI1	-
7199-99	HE663067	Alemania	Haarlem	-
94-M4241A	NZ_ABLL01000000	China	Beijing	-
98-R604	NZ_ABVM01000000	South África	-	Linaje 4
Beijing-NITR203	NC_021054.1	India	Beijing	-
BJ05010	SRR671858	China	Beijing	-
BJ05026	SRR671859	China	-	-
BJ05043	SRR671860	China	-	-
BT1	CP002883.1		-	-
BTB05-552	AEGC01000000	Suecia	-	-
BTB05-559	AEGD01000000	Suecia	-	-
BTBH-587	ERR1200614	Etiopía	-	-



<b>CAS-NITR204</b>	NC_021193	South India	CAS	-
<b>CCDC5079</b>	NC_021251	China	Beijing	-
<b>CCDC5180</b>	YP_005907833	China	Beijing	-
<b>CDC1551</b>	NC_002755	USA	X	-
<b>CDC-602</b>	SRR030017	USA	-	Linaje 4
<b>CDC-603</b>	SRR024094	USA	-	Linaje 4
<b>CDC-605</b>	SRR023971	USA	-	Linaje 2
<b>CDC-606</b>	SRR024092	USA	-	Linaje 2
<b>CDC-609</b>	SRR023957	USA	-	-
<b>CPHL-A</b>	NZ_GG663502	South África	-	Linaje 5
<b>CTRI-2</b>	NC_017524	Rusia	LAM	-
<b>CTRI-4</b>	AIIE01000000	Rusia	-	Linaje 2
<b>DY21</b>	ERP001731	Ghana	AFRI1	-
<b>DY28</b>	ERP001731	Ghana	-	Linaje 1,3
<b>EAI5</b>	CP006578.1	India	EAI	-
<b>EAI5-NITR206</b>	NC_021194	India	EAI	-
<b>EAI-OSDD271</b>	AQQC01000000	India	EAI	-
<b>EAS054</b>	ABOV01000000	USA	EAI	-
<b>ErdmanATCC35801</b>	NC_020559	USA	Haarlem	-
<b>F11</b>	NC_009565	South África	LAM	-
<b>FJ05141</b>	SRR671841	China	Beijing	-
<b>FJ05194</b>	ANBL00000000	China	-	Linaje 2
<b>FJ07070</b>	SRR671842	China	Beijing	-
<b>GM-0981</b>	SRR022876	Gambia	AFRI2	-
<b>GM-1503</b>	NZ_ABQG01000000	Gambia	LAM	-
<b>GQ366</b>	ERP001731	China	-	Linaje 2
<b>GQ762</b>	ERP001731	China	-	Linaje 2
<b>GuangZ0008</b>	SRR671863	China	Beijing	-
<b>GuangZ0016</b>	SRR671861	China	-	-
<b>GuangZ0017</b>	SRR671864	China	Beijing	-
<b>GuangZ0019</b>	SRR671864	China	-	-
<b>GuangZ0026</b>	SRR671862	China	-	-
<b>GZ077</b>	SRR671873	China	Beijing	-
<b>GZ10057</b>	SRR671835	China	-	-
<b>GZ10109</b>	SRR671844	China	Beijing	-
<b>GZ104</b>	SRR671845	China	Beijing	-
<b>GZ146</b>	SRR671847	China	Beijing	-
<b>H37Ra</b>	NC_009525	EEUU	-	-

<b>H37Rv</b>	NC_000962	EEUU	T	-
<b>HAARLEM</b>	NZ_AASN01000000	Irán	Haarlem	-
<b>Haarlem-NITR202</b>	CP004886	India	Haarlem	-
<b>HM-Uruguay</b>	AZHK01000001	Uruguay	-	Linaje 4
<b>HN878</b>	ADNF01000000	USA	Beijing	-
<b>INS1-SEN</b>	JAQH01000000	Perú	LAM	-
<b>INS2-MDR</b>	JAQI01000000	Perú	LAM	-
<b>INS3-XDR</b>	JANH01000000	Perú	LAM	-
<b>Italia-P1</b>	SRR023955	Italia	-	-
<b>Italia-P10</b>	SRR024055	Italia	-	-
<b>Italia-P2</b>	SRR024054	Italia	-	-
<b>Italia-P5</b>	SRR024056	Italia	-	-
<b>Italia-P6</b>	SRR023956	Italia	-	-
<b>Italia-P8</b>	SRR024053	Italia	-	-
<b>K100</b>	SRR022871		-	-
<b>K21</b>	SRR006916	Holanda	-	Linaje 1,3
<b>K37</b>	SRR006918	Holanda	-	-
<b>K49</b>	SRR006917	Holanda	-	Linaje 3
<b>K67</b>	SRR006919	Holanda	-	Linaje 1,3
<b>K85</b>	NZ_ACHQ01000000	West África	AFRI2	-
<b>K93</b>	SRR006920	Holanda	-	Linaje 1,3
<b>KZN-1435</b>	NC_012943	South África	LAM	-
<b>KZN-4207</b>	NC_016768	South África	LAM	-
<b>KZN-605</b>	NZ_ABGN02000000	South África	LAM	-
<b>KZN-R506</b>	NZ_ACVU02000000	South África	LAM	-
<b>KZN-V2475</b>	NZ_ACVT02000000	South África	LAM	-
<b>LN373</b>	SRR671854	China	-	Linaje 2
<b>Mex-X581</b>	-	México	-	Linaje 4
<b>Mex-X632</b>	SRR023455	México	-	Linaje 4
<b>Mex-X721</b>	SRR026444	México	-	Linaje 4
<b>MTB-476</b>	AZBA00000000	Kazajstán	-	Linaje 4
<b>MTB-489</b>	GCA_000512065.1	Kazajstán	-	-
<b>N0011</b>	ERP001731	México/San Francisco	-	Linaje 4
<b>N0020</b>	ERP001731	Mongolia/San Francisco	-	Linaje 2
<b>N0046</b>	ERP001731	México/San Francisco	-	Linaje 4
<b>N0051</b>	ERP001731	China/San Francisco	-	Linaje 2

<b>N0053</b>	ERP001731	China/San Francisco	-	Linaje 2
<b>N0079</b>	ERP001731	Vietnam/San Francisco	-	Linaje 1,3
<b>N0089</b>	ERP001731	Gambia	-	Linaje 6
<b>N0091</b>	ERP001731	Gambia	-	Linaje 6
<b>N0101</b>	ERP001731	Nicaragua/San Francisco	-	Linaje 4
<b>N0120</b>	ERP001731	Puerto Rico/San Francisco	-	-
<b>N0126</b>	ERP001731	Colombia/San Francisco	-	Linaje 4
<b>N0137</b>	ERP001731	México/San Francisco	-	Linaje 4
<b>N0185</b>	ERP001731	Reino Unido	-	-
<b>N1004</b>	ERP001731	Nepal	-	Linaje 1,3
<b>N1006</b>	ERP001731	Nepal	-	Linaje 1,3
<b>N1008</b>	ERP001731	Vietnam/San Francisco	-	Linaje 2
<b>N103</b>	ERP001731	Nicaragua/San Francisco	-	Linaje 4
<b>N107</b>	ERP001731	Guatemala/San Francisco	-	Linaje 4
<b>N109</b>	ERP001731	El Salvador/San Francisco	-	-
<b>N72</b>	ERP001731	India/San Francisco	-	Linaje 1,3
<b>NA-A0008</b>	ALYG01000000	India	-	Linaje 1,3
<b>NA-A0009</b>	ALYH01000000	India	-	Linaje 1,3
<b>NCGM2209</b>	BADQ01000000	Japón	Beijing	-
<b>NM11006</b>	SRR671879	China	Beijing	-
<b>OM-V02-005</b>		Japón	Beijing	-
<b>OSDD071</b>	AHHX01000000	India	CAS	-
<b>OSDD105</b>	PRJNA213252	India	-	Linaje 4
<b>OSDD493</b>	AVQJ01000000	India	URAL	-
<b>OSDD504</b>	AHHY01000000	India	CAS	-
<b>OSDD515</b>	PRJNA213253	India	-	-
<b>OSDD518</b>	AHHZ01000000	India	CAS	-
<b>PanR0201</b>	ANZG00000000	Panamá	LAM	-
<b>PanR0202</b>	ATAS00000000	Panamá	-	-
<b>PanR0203</b>	ATML00000000	Panamá	LAM	-
<b>PanR0205</b>	ATMM00000000	Panamá	LAM	-

<b>PanR0206</b>	ATMN00000000	Panamá	Haarlem	-
<b>PanR0207</b>	ATMQ00000000	Panamá	LAM	-
<b>PanR0208</b>	ATMO00000000	Panamá	-	-
<b>PanR0209</b>	ATMP00000000	Panamá	LAM	-
<b>PanR0301</b>	ATET00000000	Panamá	X	-
<b>PanR0304</b>	ATES00000000	Panamá	LAM	-
<b>PanR0305</b>	ATER00000000	Panamá	X	-
<b>PanR0306</b>	ATEQ00000000	Panamá	LAM	-
<b>PanR0307</b>	ATEP00000000	Panamá	LAM	-
<b>PanR0308</b>	ATEO00000000	Panamá	LAM	-
<b>PanR0309</b>	ATEN00000000	Panamá	-	-
<b>PanR0311</b>	ATEM00000000	Panamá	LAM	-
<b>PanR0313</b>	ATEL00000000	Panamá	LAM	-
<b>PanR0314</b>	ATEK00000000	Panamá	LAM	-
<b>PanR0315</b>	ATEJ00000000	Panamá	Haarlem	-
<b>PanR0316</b>	ATEI00000000	Panamá	X	-
<b>PanR0317</b>	ATEH00000000	Panamá	LAM	-
<b>PanR0401</b>	ATEG00000000	Panamá	LAM	-
<b>PanR0402</b>	ATEF00000000	Panamá	LAM	-
<b>PanR0403</b>	ATEE00000000	Panamá	LAM	-
<b>PanR0404</b>	ATED00000000	Panamá	LAM	-
<b>PanR0405</b>	ATEC00000000	Panamá	LAM	-
<b>PanR0407</b>	ATEB00000000	Panamá	LAM	-
<b>PanR0409</b>	ATEA00000000	Panamá	LAM	-
<b>PanR0410</b>	ATDZ00000000	Panamá	LAM	-
<b>PanR0411</b>	ATDY00000000	Panamá	LAM	-
<b>PanR0412</b>	ATDX00000000	Panamá	LAM	-
<b>PanR0501</b>	ATRP00000000	Panamá	LAM	-
<b>PanR0503</b>	ATRQ00000000	Panamá	LAM	-
<b>PanR0505</b>	ATRR00000000	Panamá	LAM	-
<b>PanR0601</b>	ATRS00000000	Panamá	LAM	-
<b>PanR0602</b>	ATRT00000000	Panamá	LAM	-
<b>PanR0603</b>	ATRU00000000	Panamá	LAM	-
<b>PanR0604</b>	ATRV00000000	Panamá	LAM	-
<b>PanR0605</b>	ATRW00000000	Panamá	Beijing	-
<b>PanR0606</b>	ATRX00000000	Panamá	Beijing	-
<b>PanR0607</b>	ATRY00000000	Panamá	LAM	-

<b>PanR0609</b>	ATRZ00000000	Panamá	LAM	-
<b>PanR0610</b>	ATSA00000000	Panamá	LAM	-
<b>PanR0611</b>	ATSB00000000;	Panamá	LAM	-
<b>PanR0702</b>	ATSC00000000	Panamá	LAM	-
<b>PanR0703</b>	ATSD00000000	Panamá	LAM	-
<b>PanR0704</b>	ANNN00000000	Panamá	LAM	-
<b>PanR0707</b>	ATSE00000000	Panamá	LAM	-
<b>PanR0708</b>	ATSF00000000	Panamá	LAM	-
<b>PanR0801</b>	ATSG00000000	Panamá	Haarlem	-
<b>PanR0802</b>	ANZH00000000	Panamá	LAM	-
<b>PanR0803</b>	ATSH00000000	Panamá	LAM	-
<b>PanR0804</b>	ATSI00000000	Panamá	LAM	-
<b>PanR0805</b>	ATSJ00000000	Panamá	LAM	-
<b>PanR0902</b>	ATSK00000000	Panamá	Haarlem	-
<b>PanR0903</b>	ATSL00000000	Panamá	LAM	-
<b>PanR0904</b>	ATSM00000000	Panamá	LAM	-
<b>PanR0906</b>	ATSN00000000	Panamá	LAM	-
<b>PanR0907</b>	ATSO00000000	Panamá	Haarlem	-
<b>PanR0908</b>	ATSP00000000	Panamá	-	-
<b>PanR0908</b>	ATSP00000000	Panamá	-	-
<b>PanR0909</b>	ATSQ00000000	Panamá	LAM	-
<b>PanR1005</b>	ANZI00000000	Panamá	LAM	-
<b>PanR1006</b>	ATST00000000	Panamá	LAM	-
<b>PanR1007</b>	CM002098	Panamá	LAM	-
<b>PanR1101</b>	ATSS00000000	Panamá	LAM	-
<b>Peru-02-R1848</b>	SRR023962	Perú	-	Linaje 4
<b>Peru-03-R0268</b>	SRR024072	Perú	-	Linaje 2
<b>Peru-03-R0303</b>	SRR023961	Perú	-	Linaje 4
<b>Peru-03-R0483</b>	SRR024071	Perú	-	Linaje 2
<b>Peru-03-R0985</b>	SRR024073	Perú	-	Linaje 4
<b>Peru-03-R1382</b>	SRR024074	Perú	-	Linaje 4
<b>Peru-04-R0292</b>	SRR024068	Perú	-	Linaje 4
<b>PR05</b>	AOMG00000000	Malasia	-	Linaje 1,3
<b>R1207</b>	ADNH00000000.1	South África	Beijing	-
<b>R179</b>	SRR023951	-	-	Linaje 4
<b>RGTB327</b>	CP003233	India	-	Linaje 4

<b>RGTB423</b>	CP003234	India	-	Linaje 1
<b>Russia-01-R0450</b>	SRR024069	Rusia	-	Linaje 4
<b>Russia-01-R0572</b>	SRR023964	Rusia	-	Linaje 4
<b>Russia-01-R0817</b>	SRR024067	Rusia	-	Linaje 4
<b>Russia-03-R0888</b>	SRR024070	Rusia	-	Linaje 4
<b>Russia-03-R1082</b>	SRR023963	Rusia	-	Linaje 4
<b>S96-129</b>	AEGB01000000	Suecia	-	-
<b>SP21</b>	PRJNA181180	Rusia	Beijing	-
<b>Strain-C</b>	NZ_AAKR01000000	USA	X	-
<b>SUMu001</b>	NZ_ADHQ01000000	Canadá	-	-
<b>SUMu002</b>	NZ_ADHR01000000	Canadá	-	-
<b>SUMu003</b>	NZ_ADHS01000000	Canadá	-	-
<b>SUMu004</b>	NZ_ADHT01000000	Canadá	-	-
<b>SUMu005</b>	NZ_ADHU01000000	Canadá	-	-
<b>SUMu006</b>	NZ_ADHV00000000	Canadá	-	-
<b>SUMu007</b>	NZ_ADHW01000000	Canadá	-	-
<b>SUMu008</b>	NZ_ADHX01000000	Canadá	-	-
<b>SUMu009</b>	NZ_ADHY01000000	Canadá	-	-
<b>SUMu010</b>	NZ_ADHZ01000000	Canadá	-	-
<b>SUMu011</b>	NZ_ADIA01000000	Canadá	-	-
<b>SUMu012</b>	PRJNA43625	Canadá	-	-
<b>T17</b>	NZ_DS986693	Filipinas	EAI	-
<b>T46</b>	NZ_GG663492	Filipinas	EAI	-
<b>T85</b>	NZ_DS985163	China	Beijing	-
<b>T92</b>	NZ_DS985146	Filipinas	EAI	-
<b>UM1072388579</b>	AMXW00000000	Malasia	-	Linaje 2
<b>UT205</b>	NC_016934	Colombia	LAM	-
<b>V119BJ</b>	ERP001731	Vietnam	-	Linaje 2
<b>V173IO</b>	ERP001731	Vietnam	-	Linaje 4
<b>V212BJ</b>	ERP001731	Vietnam	-	Linaje 2
<b>V232IO</b>	ERP001731	Vietnam	-	Linaje 1,3
<b>W-148</b>	ACSX01000000	Rusia	Beijing	-
<b>WX1-china</b>	AJGM01000000	China	-	Linaje 2
<b>WX3-china</b>	AJGL01000000	China	-	Linaje 2
<b>X122</b>	CM001044	South África	Beijing	-
<b>XDR1219-china</b>	AJGO01000000	China	-	Linaje 2
<b>XDR1221-china</b>	AJGN01000000	China	-	Linaje 2

<b>XZ06030</b>	SRR671838	China	-	-
<b>XZ06050</b>	SRR671839	China	Beijing	-

**Anexo II:** Base de datos de SNPs, que incluye SNPs correspondientes a genes y grupos COG por cada genotipo de MTB

SNPs	Genotipo	Cambio-aa	Gen	Grupo COG
1102468	Beijing	Gly	Probable adhesion component transport ATP-binding protein ABC transporter	Mecanismo de defensa
3830349	Beijing	Ala391Thr	Probable inosine-5'-monophosphate dehydrogenase GuaB2 (imp dehydrogenase) (inosinic acid dehydrogenase) (inosinate dehydrogenase) (imp oxidoreductase) (inosine-5'-monophosphate oxidoreductase) (IMPDH) (IMPD)	Transporte de nucleótido y metabolismo
4409954	CAS	Ala39Gly	Probable conserved transmembrane protein	Tráfico intracelular, secreción y transporte vesicular
1513189	CAS	Ala48Val	Iron-regulated transporter IrtA	Mecanismo de defensa
1540484	CAS	Leu	hypothetical protein	Mecanismo de defensa
3069566	CAS	Gly173Asp	Possible type I restriction/modification system DNA methylase HsdM (M protein) (DNA methyltransferase)	Mecanismo de defensa
3072285	CAS	Leu119Arg	Possible type I restriction/modification system specificity determinant HsdS (S protein)	Mecanismo de defensa
3273107	CAS	Ala	Daunorubicin-dim-transport ATP-binding protein ABC transporter DrrA	Mecanismo de defensa
3312620	CAS	Glu73Asp	Possible methyltransferase (methylase)	Mecanismo de defensa
3236442	CAS	His	Probable chromosome partition protein Smc	Control del ciclo celular, división celular, partición cromosómica



<b>3236497</b>	CAS	Val437Glu	Probable chromosome partition protein Smc	Control del ciclo celular, división celular, partición cromosómica
<b>3771009</b>	CAS	Ser	Antitoxin RelJ	Control del ciclo celular, división celular, partición cromosómica
<b>4372540</b>	CAS	Val47Gly	Probable conserved membrane protein	Control del ciclo celular, división celular, partición cromosómica
<b>1554216</b>	CAS	Pro10Leu	Probable dihydroorotase PyrC (DHOase)	Transporte de nucleótido y metabolismo
<b>1925136</b>	CAS	Val	Probable CTP synthase PyrG	Transporte de nucleótido y metabolismo
<b>2910852</b>	CAS	Ala17Pro	Adenine phosphoribosyltransferase Apt (APRT) (AMP diphosphorylase) (AMP pyrophosphorylase) (transphosphoribosidase)	Transporte de nucleótido y metabolismo
<b>2941179</b>	CAS	Arg	hypothetical protein	Transporte de nucleótido y metabolismo
<b>3413785</b>	CAS	Val160Ile	Ribonucleoside-diphosphate reductase (alpha chain) NrdE (ribonucleotide reductase small subunit) (R1F protein)	Transporte de nucleótido y metabolismo
<b>24007</b>	CAS	Arg	Conserved protein with FHA domain, FhaA	Mecanismos de transducción de señales
<b>123198</b>	CAS	Pro	hypothetical protein	Mecanismos de transducción de señales

123745	CAS	Gly477Arg	hypothetical protein	Mecanismos de transducción de señales
170671	CAS	Ala130Thr	Probable transcriptional regulatory protein (possibly TetR-family)	Transcripción
762434	CAS	Gly	DNA-directed RNA polymerase (beta chain) RpoB (transcriptase beta chain) (RNA polymerase beta subunit)	Transcripción
919551	CAS	Val	hypothetical protein	Transcripción
1308317	CAS	His159Arg	hypothetical protein	Transcripción
1827468	CAS	Val135Leu	Membrane-anchored adenyllyl cyclase Cya (ATP pyrophosphate-lyase) (adenylate cyclase)	Mecanismos de transducción de señales
2353385	CAS	Gln311Glu	Proteasome accessory factor C PafC	Transcripción
2477984	CAS	Ser	Adenyllyl cyclase (ATP pyrophosphate-lyase) (adenylate cyclase)	Mecanismos de transducción de señales
2516271	CAS	Met323Thr	hypothetical protein	Mecanismos de transducción de señales
2542543	CAS	Arg90Cys	hypothetical protein	Transcripción
2948524	CAS	Glu12Val	Possible Transcriptional regulatory protein	Transcripción
3219500	CAS	Lys121Met	Probable transcriptional regulatory protein (probably TetR-family)	Transcripción
3357464	CAS	Ala	hypothetical protein	Mecanismos de transducción de señales
3428897	CAS	Arg552Pro	Probable carbon starvation protein A homolog CstA	Mecanismos de transducción de señales

<b>3489665</b>	CAS	Pro62Ser	Transcriptional regulatory protein MoaR1	Mecanismos de transducción de señales
<b>3499497</b>	CAS	Ala140Gly	Two component transcriptional regulatory protein DevR (probably LuxR/UhpA-family)	Mecanismos de transducción de señales
<b>4195799</b>	CAS	Gly	Metal sensor transcriptional regulator (ArsR-SmtB family)	Transcripción
<b>4305243</b>	CAS	Pro148Leu	Transcriptional regulatory protein (probably TetR-family)	Transcripción
<b>3241244</b>	EAI	Thr239Ile	hypothetical protein	Control del ciclo celular, división celular, partición cromosómica
<b>865761</b>	EAI	His	Probable phosphoribosylamine-glycine ligase PurD (GARS) (glycinamide ribonucleotide synthetase) (phosphoribosylglycinamide synthetase) (5'-phosphoribosylglycinamide synthetase)	Transporte de nucleótidos y metabolismo
<b>3086261</b>	EAI	Gln157Arg	Possible transcriptional regulatory protein (probably Lrp/AsnC-family)	Transcripción
<b>1956930</b>	EAI	Thr39Pro	Possible penicillin-binding protein	Mecanismo de defensa
<b>4307344</b>	EAI	Leu160Val	Transcriptional regulatory protein (probably AraC-family)	Transcripción
<b>3672105</b>	EAI	His65Asp	Probable transcriptional regulatory protein LrpA (Lrp/AsnC-family)	Transcripción
<b>1521526</b>	EAI	Glu	hypothetical protein	Mecanismos de transducción de señales
<b>1923985</b>	EAI	Val53Ile	Probable CTP synthase PyrG	Transporte de nucleótidos y metabolismo

<b>2622508</b>	EAI	Leu	Probable deoxyguanosine triphosphate triphosphohydrolase Dgt (dGTPase) (dGTP triphosphohydrolase)	Transporte de nucleótidos y metabolismo
<b>3536008</b>	EAI	Pro17Gln	Probable transcriptional regulatory protein (probably TetR-family)	Transcripción
<b>7341116</b>	EAI	Met127Thr	Probable preprotein translocase SecE1	Tráfico intracelular, secreción y transporte vesicular
<b>74737</b>	EAI	Glu154Asp	Possible transcriptional regulatory protein (possibly TetR-family)	Transcripción
<b>2096430</b>	EAI	Leu57Arg	Transcriptional repressor Bial	Transcripción
<b>2731741</b>	EAI	Leu	Probable cyclase (adenylyl- or guanylyl-)(adenylate- or guanylate-)	Mecanismos de transducción de señales
<b>1826624</b>	EAI	Ala416Val	Membrane-anchored adenylyl cyclase Cya (ATP pyrophosphate-lyase) (adenylate cyclase)	Mecanismos de transducción de señales
<b>1417019</b>	EAI	Cys110Tyr	Probable transcriptional regulatory protein EmbR	Mecanismos de transducción de señales
<b>3274545</b>	EAI	Leu	Probable daunorubicin-dim-transport integral membrane protein ABC transporter DrrC	Mecanismo de defensa
<b>4187063</b>	EAI	Gly144Arg	Transcriptional regulatory protein (probably AraC/XylS-family)	Transcripción
<b>2092391</b>	EAI	Ile	Probable inosine-5'-monophosphate dehydrogenase GuaB1 (imp dehydrogenase) (IMPDH) (IMPD)	Transporte de nucleótidos y metabolismo
<b>2704291</b>	EAI	Ile	hypothetical protein	Mecanismos de transducción de señales

<b>765150</b>	Haarlem	Gly594Glu	DNA-directed RNA polymerase (beta' chain) RpoC (transcriptase beta' chain) (RNA polymerase beta' subunit)	Transcripción
<b>4063682</b>	Haarlem	Leu61Met	Hypoxanthine-guanine phosphoribosyltransferase Hpt (HGPRT) (HGPRTase) (hypoxanthine phosphoribosyltransferase) (imp pyrophosphorylase) (imp diphosphorylase) (transphosphoribosyltransferase) (guanine phosphoribosyltransferase)	Transporte de nucleótidos y metabolismo
<b>4306929</b>	Haarlem	Arg	Transcriptional regulatory protein (probably AraC-family)	Transcripción
<b>902413</b>	Haarlem	Val	Amidophosphoribosyltransferase PurF (glutamine phosphoribosylpyrophosphate amidotransferase) (ATASE) (gpatase)	Transporte de nucleótidos y metabolismo
<b>3236230</b>	LAM	Arg526Leu	Probable chromosome partition protein Smc	Control del ciclo celular, división celular, partición cromosómica
<b>3429202</b>	LAM	Tyr654Asp	Probable carbon starvation protein A homolog CstA	Mecanismos de transducción de señales
<b>1553633</b>	X	Gly	Probable aspartate carbamoyltransferase PyrB (ATCase) (aspartate transcarbamylase)	Transporte de nucleótidos y metabolismo
<b>2946570</b>	X	Gly63Asp	hypothetical protein	Transcripción
<b>3233940</b>	Ural	Ala67Gly	Probable cell division protein FtsY (SRP receptor) (signal recognition particle receptor)	Tráfico intracelular, secreción y transporte vesicular
<b>3866650</b>	Ural	Ala530Val	ESX conserved component EccC4 ESX-4 type VII secretion system protein Probable membrane protein	Control del ciclo celular, división celular, partición cromosómica

<b>3867404</b>	Ural	Glu279*	ESX conserved component EccC4 ESX-4 type VII secretion system protein	Control del ciclo celular, división celular, partición cromosómica
<b>2913440</b>	Ural	Ala191Val	Probable protein-export membrane protein SecF	Tráfico intracelular, secreción y transporte vesicular
<b>181326</b>	Ural	Val	Phosphotyrosine protein phosphatase PTPB (protein-tyrosine-phosphatase) (PTPase)	Mecanismos de transducción de señales
<b>3442970</b>	Ural	Ala1007Thr	Serine/threonine-protein kinase transcriptional regulatory protein PknK (protein kinase K) (STPK K)	Transcripción
<b>3471151</b>	Ural	Arg	Putative cell division ATP-binding protein FtsE (septation component-transport ATP-binding protein ABC transporter)	Control del ciclo celular, división celular, partición cromosómica
<b>3875230</b>	Ural	Thr158Ala	Probable conserved integral membrane protein	Transporte de nucleótidos y metabolismo
<b>3772811</b>	Ural	Arg19Gln	hypothetical protein	Mecanismos de transducción de señales
<b>423722</b>	Ural	Thr	Probable heat shock protein transcriptional repressor HspR (MerR family)	Transcripción
<b>4167910</b>	Ural	His73Gln	hypothetical protein	Transcripción
<b>1225462</b>	Ural	Asp228Gly	Probable membrane glycine and proline rich protein	Procesamiento y modificación de ARN
<b>2623012</b>	Ural	Ala	Probable deoxyguanosine triphosphate triphosphohydrolase Dgt (dGTPase) (dGTP triphosphohydrolase)	Transporte de nucleótidos y metabolismo
<b>3422104</b>	Ural	Val	Probable transcriptional regulatory protein (probably GntR-family)	Transcripción

<b>4406116</b>	Ural	Ala126Ser	Probable chromosome partitioning protein ParB	Transcripción
<b>753562</b>	Ural	Ile19Thr	Possible antitoxin VapB6	Transcripción
<b>3234710</b>	Ural	Leu	Probable chromosome partition protein Smc	Control del ciclo celular, división celular, partición cromosómica
<b>1151644</b>	Ural	Tyr15His	Probable sensor protein KdpD	Mecanismos de transducción de señales
<b>1422920</b>	Ural	Thr	Probable drugs-transport transmembrane ATP-binding protein ABC transporter	Mecanismo de defensa
<b>4376445</b>	Ural	Pro	ESX conserved component EccC2 ESX-2 type VII secretion system protein Possible membrane protein	Control del ciclo celular, división celular, partición cromosómica
<b>3447800</b>	Ural	Asp209Glu	Virulence-regulating transcriptional regulator VirS (AraC/XylS family)	Transcripción
<b>47843</b>	Ural	Asp	Probable transcriptional regulatory protein (probably GntR-family)	Transcripción
<b>407877</b>	Ural	Glu175Gly	Possible transcriptional regulatory protein	Mecanismos de transducción de señales
<b>3411778</b>	Ural	Gly	Probable transcriptional regulatory protein (probably AsnC-family)	Transcripción
<b>2478266</b>	Ural	Glu369Asp	Adenylyl cyclase (ATP pyrophosphate-lyase) (adenylate cyclase)	Mecanismos de transducción de señales
<b>4372205</b>	Ural	Asn159Tyr	Probable conserved membrane protein	Control del ciclo celular, división celular, partición cromosómica
<b>1279923</b>	Ural	Met90Thr	Probable transcriptional regulatory protein	Transcripción
<b>2262026</b>	Ural	Ala	hypothetical protein	Mecanismo de defensa
<b>919424</b>	Ural	Val44Leu	hypothetical protein	Transcripción

<b>3585577</b>	Ural	Cys192Arg	Probable transcriptional regulatory protein (probably TetR-family)	Transcripción
<b>4379186</b>	Ural	His423Tyr	ESX conserved component EccC2 ESX-2 type VII secretion system protein	Control del ciclo celular, división celular, partición cromosómica
<b>3550873</b>	Ural	Gln	Possible transcriptional regulatory protein	Transcripción
<b>695315</b>	Ural	Glu	Possible antitoxin VapB4	Control del ciclo celular, división celular, partición cromosómica
<b>3058293</b>	Ural	Arg80His	Transcriptional regulatory protein CigR	Transcripción
<b>2251179</b>	Ural	Pro	Universal stress protein family protein	Mecanismos de transducción de señales
<b>3422687</b>	Ural	Thr176Ile	Probable transcriptional regulatory protein (probably GntR-family)	Transcripción
<b>1605170</b>	Ural	Ala98Val	hypothetical protein	Mecanismos de transducción de señales
<b>2396770</b>	Ural	Ser	Possible conserved transmembrane protein	Mecanismo de defensa
<b>2151148</b>	Ural	Ile66Met	hypothetical protein	Mecanismos de transducción de señales
<b>2507412</b>	Ural	Leu	Phosphotyrosine protein phosphatase PtpA (protein-tyrosine-phosphatase) (LMPase) (LMW phosphatase)	Mecanismos de transducción de señales
<b>2176006</b>	AFRI1	Gly	Probable lipase LipD	Mecanismo de defensa
<b>921429</b>	AFRI1	Pro62Leu	Possible deaminase	Transporte de nucleótidos y metabolismo
<b>3875633</b>	AFRI1	Tyr292Cys	Probable conserved integral membrane protein	Transporte de nucleótidos y metabolismo



25386	AFRI1	Gly20Asp	Conserved protein with FHA domain, FhaA	Mecanismos de transducción de señales
477988	AFRI1	His190Pro	Possible conserved lipoprotein LpqK	Mecanismo de defensa
1482978	AFRI1	Ala370Thr	Possible adenylate cyclase (ATP pyrophosphatase) (adenylyl cyclase)	Mecanismos de transducción de señales
3445777	AFRI1	Arg71Pro	Serine/threonine-protein kinase transcriptional regulatory protein PknK (protein kinase K) (STPK K)	Transcripción
2910483	AFRI1	Leu140Ile	Adenine phosphoribosyltransferase Apt (APRT) (AMP diphosphorylase) (AMP pyrophosphorylase) (transphosphoribosidase)	Transporte de nucleótidos y metabolismo
3596354	AFRI1	Asp394Gly	Probable two component sensor kinase	Mecanismos de transducción de señales
227468	AFRI1	Val	Probable transmembrane multidrug efflux pump	Mecanismo de defensa
1846854	AFRI1	Tyr	hypothetical protein	Transcripción
3591082	AFRI1	Asn	Possible SOJ/para-related protein	Control del ciclo celular, división celular, partición cromosómica
3701552	AFRI1	Glu	Probable adenosine deaminase Add (adenosine aminohydrolase)	Transporte de nucleótidos y metabolismo
458116	AFRI1	Asp89Asn	Probable orotate phosphoribosyltransferase PyrE (OPRT) (oprtae)	Transporte de nucleótidos y metabolismo
141516	AFRI1	Ser106Phe	Oxidative stress response regulatory protein OxyS	Transcripción
229448	AFRI1	Leu857Phe	Probable transmembrane multidrug efflux pump	Mecanismo de defensa
408006	AFRI1	Pro132Gln	Possible transcriptional regulatory protein	Mecanismos de transducción de señales

<b>27947</b>	AFRI1	Ala118Val	Possible transcriptional regulatory protein	Transcripción
<b>3630235</b>	AFRI1	Arg	Possible transcriptional regulatory protein (probably TetR-family)	Transcripción
<b>1555432</b>	AFRI1	Thr	Probable dihydroorotase PyrC (DHOase)	Transporte de nucleótidos y metabolismo
<b>500223</b>	AFRI1	Val151Leu	Possible mutator protein MutT3 (7,8-dihydro-8-oxoguanine-triphosphatase) (8-oxo-dGTPase) (dGTP pyrophosphohydrolase)	Transporte de nucleótidos y metabolismo
<b>3416432</b>	AFRI1	Val118Ile	Possible transcriptional regulatory protein (probably TetR-family)	Transcripción
<b>3193575</b>	AFRI1	Ala	Probable transcriptional regulatory protein	Mecanismos de transducción de señales
<b>3648267</b>	AFRI1	Gly127Val	Conserved protein (CPSA-related protein)	Transcripción
<b>1403266</b>	AFRI1	Ala41Pro	Probable transcriptional regulatory protein	Transcripción
<b>904367</b>	AFRI1	Asp215His	Probable phosphoribosylformylglycinamide CYCLO-ligase PurM (AIRS) (phosphoribosyl-aminoimidazole synthetase) (air synthase)	Transporte de nucleótidos y metabolismo
<b>4347337</b>	AFRI1	Ala286Asp	ESX conserved component EccCa1 ESX-1 type VII secretion system protein Possible transmembrane protein	Control del ciclo celular, división celular, partición cromosómica
<b>3106231</b>	AFRI1	Ala	hypothetical protein	Control del ciclo celular, división celular, partición cromosómica
<b>3068778</b>	AFRI1	Glu436Lys	Possible type I restriction/modification system DNA methylase HsdM (M protein) (DNA methyltransferase)	Mecanismo de defensa

<b>47877</b>	AFRI1	Asp75Ala	Probable transcriptional regulatory protein (probably GntR-family)	Transcripción
<b>230197</b>	AFRI1	Thr1107Ile	Probable transmembrane multidrug efflux pump	Mecanismo de defensa
<b>3864816</b>	AFRI1	Ser	ESX conserved component EccC4 ESX-4 type VII secretion system protein Probable membrane protein	Control del ciclo celular, división celular, partición cromosómica
<b>4117097</b>	AFRI1	Glu207Ala	Transcriptional regulatory protein Crp (Crp/Fnr-family)	Mecanismos de transducción de señales
<b>1507920</b>	AFRI1	Val	hypothetical protein	Transporte de nucleótidos y metabolismo
<b>2136642</b>	AFRI1	Leu129Phe	Hypothetical protein	Tráfico intracelular, secreción y transporte vesicular
<b>883072</b>	AFRI1	Arg105Cys	Probable phosphoribosylformylglycinamide synthase I PURG (FGAM synthase I)	Transporte de nucleótidos y metabolismo
<b>1605016</b>	AFRI1	Leu	hypothetical protein	Mecanismos de transducción de señales
<b>1951438</b>	AFRI1	Pro105Leu	hypothetical protein	Transcripción
<b>3867027</b>	AFRI1	Thr	ESX conserved component EccC4 ESX-4 type VII secretion system protein Probable membrane protein	Control del ciclo celular, división celular, partición cromosómica
<b>2752132</b>	AFRI1	Leu	Probable resuscitation-promoting factor RpfE	Procesamiento y modificación de ARN
<b>2408524</b>	AFRI1	Ser334Leu	Cell division protein FtsZ	Control del ciclo celular, división celular, partición cromosómica
<b>534205</b>	AFRI1	Arg	Alternative RNA polymerase sigma factor SigK	Transcripción
<b>2477562</b>	AFRI1	Leu	Adenylyl cyclase (ATP pyrophosphate-lyase) (adenylate cyclase)	Mecanismos de transducción de señales

1686737	AFRI1	Pro	Possible toxin MazF4	Mecanismos de transducción de señales
3658226	AFRI1	Ala138Val	Probable phosphoribosylaminoimidazole carboxylase catalytic subunit PurE (air carboxylase) (AIRC)	Transporte de nucleótidos y metabolismo
3148511	AFRI1	Glu306Asp	Probable N utilization substance protein A NusA	Transcripción
1097633	AFRI1	Ala	Two component sensor kinase MprB	Mecanismos de transducción de señales
2868769	AFRI1	Gly60Asp	Possible antitoxin VapB19	Transcripción
1422079	AFRI1	Arg59Pro	Probable drugs-transport transmembrane ATP-binding protein ABC transporter	Mecanismo de defensa
3865398	AFRI1	Val	ESX conserved component EccC4 ESX-4 type VII secretion system protein Probable membrane protein	Control del ciclo celular, división celular, partición cromosómica
2530101	AFRI1	Asp241Val	hypothetical protein	Mecanismo de defensa
3858894	AFRI1	Leu266Phe	Conserved hypothetical alanine and proline rich protein	Procesamiento y modificación de ARN
1605569	AFRI1	Val231Ala	hypothetical protein	Mecanismos de transducción de señales
3499247	AFRI1	Val7Ile	Two component sensor histidine kinase DevS	Mecanismos de transducción de señales
1939763	AFRI1	Pro	Cytidylate kinase Cmk (CMP kinase) (cytidine monophosphate kinase) (ck)	Transporte de nucleótidos y metabolismo
919007	AFRI1	Cys183Gly	hypothetical protein	Transcripción
1911301	AFRI1	Ala	hypothetical protein	Transcripción

<b>331309</b>	AFRI1	Leu120Pro	Possible transcriptional regulatory protein (possibly TetR-family)	Transcripción
<b>2730711</b>	AFRI1	Ala680Val	Probable cyclase (adenylyl-or guanylyl-)(adenylate-or guanylate-)	Mecanismos de transducción de señales
<b>163148</b>	AFRI1	Arg	Possible transcriptional regulatory protein	Transcripción
<b>3866953</b>	AFRI1	Val429Ala	ESX conserved component EccC4 ESX-4 type VII secretion system protein Probable membrane protein	Control del ciclo celular, división celular, partición cromosómica
<b>3096576</b>	AFRI1	Arg489Gln	Conserved hypothetical alanine rich protein	Control del ciclo celular, división celular, partición cromosómica
<b>3416630</b>	AFRI1	Gly184Ser	Possible transcriptional regulatory protein (probably TetR-family)	Transcripción
<b>4080434</b>	AFRI1	Asp43Asn	Possible cell filamentation protein Fic	Control del ciclo celular, división celular, partición cromosómica
<b>86587</b>	AFRI1	Ile	Probable transcriptional regulatory protein	Transcripción
<b>1224174</b>	AFRI2	Ala393Val	Probable PHOH-like protein PhoH2 (phosphate starvation-inducible protein PSIH)	Mecanismos de transducción de señales
<b>1515003</b>	AFRI2	Ala653Thr	Iron-regulated transporter IrtA	Mecanismo de defensa
<b>760969</b>	AFRI2	Ser388Leu	DNA-directed RNA polymerase (beta chain) RpoB (transcriptase beta chain) (RNA polymerase beta subunit)	Transcripción
<b>2772741</b>	AFRI2	Ala99Ser	hypothetical protein	Mecanismo de defensa
<b>1537926</b>	AFRI2	Glu82Lys	Anti-anti-sigma factor RsfA (anti-sigma factor antagonist) (regulator of sigma F A)	Mecanismos de transducción de señales
<b>2478180</b>	AFRI2	Leu341Ile	Adenylyl cyclase (ATP pyrophosphate-lyase) (adenylate cyclase)	Mecanismos de transducción de señales

<b>3702543</b>	AFRI2	Arg309Trp	Probable thymidine phosphorylase DeoA (tdrpase) (pyrimidine phosphorylase)	Transporte de nucleótidos y metabolismo
<b>4098514</b>	AFRI2	Asp219Gly	hypothetical protein	Control del ciclo celular, división celular, partición cromosómica
<b>2592510</b>	AFRI2	Arg73Cys	Universal stress protein family protein	Mecanismos de transducción de señales
<b>4219820</b>	AFRI2	Ala46Pro	Probable lipase LipE	Mecanismo de defensa
<b>754754</b>	AFRI2	Arg	Toxin MazF2	Mecanismos de transducción de señales
<b>1828389</b>	AFRI2	Arg	Probable two-component system transcriptional regulator	Mecanismos de transducción de señales
<b>3829664</b>	AFRI2	His83Tyr	Probable inosine-5'-monophosphate dehydrogenase GuaB3 (imp dehydrogenase) (inosinic acid dehydrogenase) (inosinate dehydrogenase) (imp oxidoreductase) (inosine-5'-monophosphate oxidoreductase) (IMPDH) (IMPD)	Transporte de nucleótidos y metabolismo
<b>3137026</b>	AFRI2	Ser71Ala	Possible antitoxin VapB22	Control del ciclo celular, división celular, partición cromosómica
<b>1536183</b>	AFRI2	Arg488Cys	Possible sigma factor regulatory protein	Mecanismos de transducción de señales
<b>3236716</b>	AFRI2	Leu364Arg	Probable chromosome partition protein Smc	Control del ciclo celular, división celular, partición cromosómica
<b>1372002</b>	AFRI2	Leu	Probable Mrp-related protein Mrp	Control del ciclo celular, división celular, partición cromosómica
<b>4035242</b>	AFRI2	Phe283Leu	Probable conserved lipoprotein LpoF	Mecanismo de defensa

<b>1226021</b>	AFRI2	Gln42Lys	Probable membrane glycine and proline rich protein	Procesamiento y modificación de ARN
<b>4393838</b>	AFRI2	Leu	Possible mutator protein MutT4	Transporte de nucleótidos y metabolismo
<b>1912024</b>	AFRI2	Val20Phe	Probable conserved integral membrane protein ABC transporter	Mecanismo de defensa
<b>886178</b>	AFRI2	Leu157Ile	Probable transcriptional regulatory protein (probably GntR-family)	Transcripción
<b>4130604</b>	AFRI2	Ser42Asn	Probable conserved transmembrane protein	Procesamiento y modificación de ARN
<b>1828506</b>	AFRI2	Met112Ile	Probable two-component system transcriptional regulator	Mecanismos de transducción de señales
<b>1150490</b>	AFRI2	Ser	Probable sensor protein KdpD	Mecanismos de transducción de señales
<b>1646982</b>	AFRI2	Ala242Val	Probable transcriptional regulatory protein	Transcripción
<b>3596407</b>	AFRI2	Val	Probable two component sensor kinase	Mecanismos de transducción de señales
<b>1279184</b>	AFRI2	Arg145Gln	Transcriptional regulatory protein	Transcripción
<b>1150803</b>	AFRI2	Gly295Asp	Probable sensor protein KdpD	Mecanismos de transducción de señales
<b>3213255</b>	AFRI2	Lys	Probable signal peptidase I LepB (SPASE I) (leader peptidase I)	Tráfico intracelular, secreción y transporte vesicular
<b>477634</b>	AFRI2	Glu308Gly	Possible conserved lipoprotein LpqK	Mecanismo de defensa

<b>4346843</b>	AFRI2	Thr	ESX conserved component EccCa1 ESX-1 type VII secretion system protein	Control del ciclo celular, división celular, partición cromosómica
<b>2075089</b>	AFRI2	Tyr	hypothetical protein	Transcripción
<b>962133</b>	AFRI2	Asp160Glu	hypothetical protein	Transcripción
<b>3775639</b>	AFRI2	Gly	hypothetical protein	Mecanismos de transducción de señales
<b>4186864</b>	AFRI2	Ile	Transcriptional regulatory protein (probably AraC/XylS-family)	Transcripción
<b>4220427</b>	AFRI2	Pro248Leu	Probable lipase LipE	Mecanismo de defensa
<b>761723</b>	AFRI2	Glu639Asp	DNA-directed RNA polymerase (beta chain) RpoB (transcriptase beta chain) (RNA polymerase beta subunit)	Transcripción
<b>2623603</b>	AFRI2	Ala	Probable deoxyguanosine triphosphate triphosphohydrolase Dgt (dGTPase) (dGTP triphosphohydrolase)	Transporte de nucleótidos y metabolismo
<b>2274463</b>	AFRI2	Leu16Val	Two component sensor histidine kinase DosT	Mecanismos de transducción de señales
<b>781075</b>	AFRI2	Ala141Pro	Probable transcriptional regulatory protein (possibly TetR-family)	Transcripción
<b>3057375</b>	AFRI2	Arg230Gln	Conserved 35 kDa alanine rich protein	Transcripción
<b>24780</b>	AFRI2	Asp222Gly	Conserved protein with FHA domain, FhaA	Mecanismos de transducción de señales
<b>1912582</b>	AFRI2	Ala	Probable conserved ATP-binding protein ABC transporter	Mecanismo de defensa



<b>941054</b>	AFRI2	Pro18Leu	Possible nitrate/nitrite response transcriptional regulatory protein NarL	Mecanismos de transducción de señales
<b>3068710</b>	AFRI2	Lys458Asn	Possible type I restriction/modification system DNA methylase HsdM (M protein) (DNA methyltransferase)	Mecanismo de defensa
<b>2427828</b>	AFRI2	Gly	Probable conserved proline rich membrane protein	Procesamiento y modificación de ARN
<b>1454811</b>	AFRI2	Asn	Probable transcription termination factor Rho homolog	Transcripción
<b>1156704</b>	AFRI2	Thr418Ala	Two component sensor histidine kinase TrcS	Mecanismos de transducción de señales
<b>3591488</b>	AFRI2	Thr2Ile	Possible SOJ/para-related protein	Control del ciclo celular, división celular, partición cromosómica
<b>1069146</b>	AFRI2	Gly	Probable bifunctional purine biosynthesis protein PurH: phosphoribosylaminoimidazolecarboxamide formyltransferase (AICAR transferase) (5'-phosphoribosyl-5-aminoimidazole-4-carboxamide formyltransferase) + inosinemonophosphate cyclohydrolase (imp cyclohydrolase) (inosinase) (imp synthetase) (ATIC)	Transporte de nucleótidos y metabolismo
<b>3086728</b>	AFRI2	Met1Ile	Possible transcriptional regulatory protein (probably Lrp/AsnC-family)	Transcripción
<b>162948</b>	AFRI2	Thr101Ile	Possible transcriptional regulatory protein	Transcripción
<b>2772760</b>	AFRI2	Ser	hypothetical protein	Mecanismo de defensa

<b>1912617</b>	AFRI2	Asp102Asn	Probable conserved ATP-binding protein ABC transporter	Mecanismo de defensa
<b>2225175</b>	AFRI2	Leu5Arg	Ribonucleoside-diphosphate reductase (beta chain) NrdF1 (ribonucleotide reductase small subunit) (R2F protein)	Transporte de nucleótidos y metabolismo
<b>1508682</b>	AFRI2	Glu81Ala	Probable conserved lipoprotein LprD	Transcripción
<b>141261</b>	AFRI2	Ala21Val	Oxidative stress response regulatory protein OxyS	Transcripción
<b>3057309</b>	AFRI2	Pro252Leu	Conserved 35 kDa alanine rich protein	Transcripción
<b>1662719</b>	AFRI2	Ala	Probable transcriptional regulatory protein	Transcripción
<b>3599099</b>	AFRI2	Glu190Asp	Alternative RNA polymerase sigma-E factor (sigma-24) SigH (RPOE)	Transcripción
<b>3776265</b>	AFRI2	Arg	hypothetical protein	Mecanismos de transducción de señales
<b>2772954</b>	AFRI2	Ser28Gly	hypothetical protein	Mecanismo de defensa
<b>3014016</b>	AFRI2	Leu	Probable deoxyuridine 5'-triphosphate nucleotidohydrolase Dut (dUTPase) (dUTP pyrophosphatase) (deoxyuridine 5'-triphosphatase) (dUTP diphosphatase) (deoxyuridine-triphosphatase)	Transporte de nucleótidos y metabolismo
<b>4099060</b>	AFRI2	Pro37Leu	hypothetical protein	Control del ciclo celular, división celular, partición cromosómica
<b>4212174</b>	AFRI2	Phe6Ser	Probable two component transcriptional regulatory protein TcrX	Mecanismos de transducción de señales
<b>1975960</b>	AFRI2	Ser	Probable conserved transmembrane ATP-binding protein ABC transporter	Mecanismo de defensa

## 11. PRODUCTIVIDAD CIENTÍFICA

**Publicación de un artículo científico en la revista *Bioinformation* generada durante la ejecución de la Maestría en Biología Molecular**

TARAZONA, D.; JARAMILLO, L.; BORDA, V.; LEVANO, K.; GALARZA, M.; GUIO, H. A Genomic Signature for Genotyping *Mycobacterium tuberculosis*. *Bioinformation*. 2017;13(7):224-230.