



UNIVERSIDAD NACIONAL DE COLOMBIA

# **Variabilidad en secuencias proteicas de los sistemas Toxina-Antitoxina en *Mycobacterium tuberculosis***

**Juan Sebastian Solano Gutierrez**

Universidad Nacional de Colombia  
Facultad de Ciencias  
Departamento de Biociencias  
Medellín, Colombia  
2017



# Variabilidad en secuencias proteicas de los sistemas Toxina-Antitoxina en *Mycobacterium tuberculosis*

**Juan Sebastian Solano Gutierrez**

Tesis presentada como requisito parcial para optar al título de:  
**Magíster en Ciencias-Biotecnología**

Director:  
Jaime Robledo Ph.D  
Codirector:  
Rafael Arango Ph.D

Línea de Investigación:  
Genómica de *M. tuberculosis* y relación hospedero microorganismo  
Grupo de Investigación:  
Unidad de Bacteriología y Micobacterias, CIB-UPB

Universidad Nacional de Colombia  
Facultad de Ciencias  
Departamento de Biociencias  
Medellín, Colombia  
2017



## *Dedicatoria*

*A mis padres, hermanos y abuela. Por su constante apoyo, por el amor que me brindan y la fe que han puesto en mí.*

*Siempre serán el motivo para asumir retos, para seguir creciendo, para vivir.*



## **Agradecimientos**

A Jaime Robledo por su apoyo y guía durante la maestría.

A Camilo Pino por sus aportes en el diseño de la metodología.

A la unidad de bacteriología, la experiencia y conocimiento de cada integrante me han permitido crecer como investigador.

A todo el personal de la CIB.

A Colciencias por la financiación del proyecto.

En especial a mi familia, a mis padres, mis dos hermanos y mi abuela. La presencia de cada uno en mi vida, será la razón para seguir adelante.





## Resumen

El agente causal de la tuberculosis es *Mycobacterium tuberculosis*, un patógeno dividido en 7 linajes asociados a zonas geográficas específicas y con diferencias en su comportamiento. Una característica notable en *M. tuberculosis* es su alto número de sistemas toxina-antitoxina, operones involucrados en la formación de células persistentes bajo situaciones de estrés. Se analizó la variabilidad en la secuencia de las proteínas de los sistemas toxina-antitoxina entre los linajes de *M. tuberculosis*. La predicción se hizo utilizando perfiles de modelos ocultos de Markov, seguido de un agrupamiento de genes ortólogos y posteriormente el alineamiento múltiple para la búsqueda de mutaciones. Se encontraron seis posibles antitoxinas y 63 proteínas completamente conservadas: 22 toxinas, 36 antitoxinas y 5 proteínas sin caracterizar. 41 toxinas y 18 antitoxinas presentaron mutaciones. Se encontraron mutaciones asociadas a cada linaje. Estos resultados sugieren que en *M. tuberculosis* se favorece el surgimiento de nuevas antitoxinas y que las toxinas son más propensas a cambios que las antitoxinas.

**Palabras clave:** *Mycobacterium tuberculosis*, toxina-antitoxina, operón, linaje

## Abstract

Tuberculosis is caused by *Mycobacterium tuberculosis*. It is a pathogen divided into 7 lineages associated with specific geographical areas and with differences in their behavior. The high numbers of toxin- antitoxin systems are a notable feature in *M. tuberculosis*, which are operons involved in the formation of persistent cells under stress situations. The sequence variability of the toxin-antitoxin systems among the *M. tuberculosis* lineages was analyzed. Profiles of hidden Markov models, followed by a clustering of orthologous genes, and subsequently multiple alignments for mutations search were used to predict protein sequences. Six possible antitoxins and 63 fully conserved proteins were found: 22 toxins, 36 antitoxins and 5 uncharacterized proteins. 41 toxins and 18 antitoxins showed mutations. Mutations associated with each lineage were found. The results suggest that *M. tuberculosis* favors the emergence of new antitoxins and that toxins are more likely to change than antitoxins.

**Keywords:** *Mycobacterium tuberculosis*, toxin-antitoxin, operon, lineage

# Contenido

	Pág.
<b>Resumen</b> .....	<b>IX</b>
<b>Abstract</b> .....	<b>X</b>
<b>Lista de figuras</b> .....	<b>XIII</b>
<b>Lista de tablas</b> .....	<b>XIV</b>
<b>Introducción</b> .....	<b>15</b>
<b>1. Análisis de sistemas toxina-antitoxina</b> .....	<b>17</b>
1.1 Marco teórico.....	17
1.1.1 La Tuberculosis.....	17
1.1.2 Patógeno.....	18
1.1.3 Sistemas toxina-antitoxina .....	20
1.1.4 Blancos de Toxinas.....	23
1.1.5 Sistemas toxina-antitoxina en <i>Mycobacterium tuberculosis</i> .....	25
1.1.6 Antecedentes .....	28
1.2 Metodología.....	30
1.2.1 Selección de aislamientos.....	30
1.2.2 Extracción de ADN y secuenciación de genomas .....	30
1.2.3 Corrección de lecturas y ensamblaje de genomas .....	30
1.2.4 Enriquecimiento de base de datos y construcción de perfiles de modelos ocultos de Markov (HMM) .....	30
1.2.5 Predicción de sistemas toxina-antitoxina.....	32
1.2.6 Agrupamiento y alineamiento múltiple.....	33
1.3 Resultados.....	35
1.4 Discusión .....	43
1.4.1 Tipos de toxina-antitoxina y nuevas proteínas.....	43
1.4.2 Core de sistemas toxina-antitoxina.....	44
1.4.3 Mutaciones de proteínas toxina-antitoxina .....	46
<b>2. Conclusiones y perspectivas</b> .....	<b>49</b>
2.1 Conclusiones .....	49
2.2 Perspectivas .....	49
<b>A. Toxinas de referencia</b> .....	<b>51</b>
<b>B. Antitoxinas de referencia</b> .....	<b>55</b>

<b>C. Sistemas TA hipotéticos en H37Rv .....</b>	<b>58</b>
<b>D. Reguladores .....</b>	<b>59</b>
<b>Bibliografía .....</b>	<b>61</b>

## Lista de figuras

	<b>Pág.</b>
<b>Figura 1:</b> Distribución geográficas de linajes.....	20
<b>Figura 2:</b> Esquema de la respuesta a estrés de sistemas TA.....	21
<b>Figura 3:</b> Mecanismos de inhibición de sistemas tipo I.....	21
<b>Figura 4:</b> Mecanismos de inhibición de sistemas tipo II.....	22
<b>Figura 5:</b> Mecanismos de inhibición de sistemas tipo III.....	22
<b>Figura 6:</b> Mecanismos de inhibición de sistemas tipo IV.....	23
<b>Figura 7:</b> Mecanismos de inhibición de sistemas tipo V.....	23
<b>Figura 8:</b> Sistemas TA reportados en H37Rv.....	26
<b>Figura 9:</b> Flujo de trabajo para la predicción y análisis de sistemas TA.....	32
<b>Figura 10:</b> Número de mutaciones identificadas de acuerdo a los genomas estudiados de cada linaje.....	37
<b>Figura 11:</b> Comparación de variabilidad en secuencias de toxinas y antitoxinas.....	37
<b>Figura 12:</b> Principales resultados obtenidos durante el análisis de sistemas TA.....	41

## Lista de tablas

	<b>Pág.</b>
<b>Tabla 1:</b> Genomas utilizados en el estudio.....	31
<b>Tabla 2:</b> Posibles proteínas TA no caracterizadas en MTB.....	34
<b>Tabla 3:</b> Nombre y familia de las proteínas TA completamente conservadas en los genomas de MTB estudiados.....	35
<b>Tabla 4:</b> Mutaciones asociadas a cada linaje.....	38
<b>Tabla 5:</b> Continuación de la tabla 4.....	39
<b>Tabla 6:</b> Mutaciones encontradas en solo una cepa y solo un linaje.....	40

## Introducción

La Tuberculosis (TB), es uno de los principales problemas de salud mundial. En 2014, según estadísticas de la Organización Mundial de la Salud (OMS) se infectaron 9,6 millones de personas y se registraron 1,5 millones de muertes (WHO 2016). En Colombia, para el 2014 se registraron 12824 casos de TB en todas las formas, de los cuales 2134 presentaron infección con VIH (Lucía Martínez et al. 2016).

El control efectivo de la enfermedad se ve impedido por distintos factores, entre estos, la latencia. Durante este estado, la bacteria se encuentra dentro de granulomas donde puede evadir la respuesta inmune del hospedero y vive en una forma inactiva (Brunner et al. 2016). En la mayoría de los individuos infectados, la bacteria permanece en este estado, sirviendo como reservorio hasta la reactivación, la cual se produce entre 2 a 23% de los casos (Alnimr 2015; Brunner et al. 2016). Una segunda amenaza al control de la TB es el surgimiento de tuberculosis multidrogo-resistentes (MDR-TB), definida como la resistencia a isoniazida y rifampicina, principales medicamentos de primera línea (Ali et al. 2016). Adicionalmente, se ha reportado tuberculosis con resistencia extendida (XDR-TB) definida como MDR-TB mas resistencia a cualquier fluoroquinolona y a un medicamento de segunda línea inyectable (kanamicina, amikacina o capreomicina) (Singhal et al. 2016).

El agente causal de la TB es *Mycobacterium tuberculosis* (MTB) (Peters et al. 2016), una bacteria en co-evolución con el humano (Brites and Gagneux 2015). Durante el proceso evolutivo, MTB desarrolló estrategias para sobreponerse a las condiciones hostiles producidas por el sistema inmune del hospedero y poder propagarse de un individuo a otro (Brites and Gagneux 2015). El genoma de MTB codifica varios factores de virulencia que la ayudan a sobrevivir en microambientes agresivos como el producido por la liberación de especies reactivas de oxígeno al interior del macrófago (Forrellad et al. 2013).

MTB es representada en siete linajes, los cuales están asociados con comunidades y zonas geográficas específicas (Jackson 2014; Yimer et al. 2015). Por ejemplo, el linaje Euro-Americano es el más frecuente en Europa y América, mientras que el linaje Beijing se encuentra predominantemente en Asia (Realpe et al. 2014). La relación entre estos linajes y las poblaciones humanas son soportados por la tendencia de cada linaje a causar enfermedad a poblaciones específicas (Realpe et al. 2014). Además de las diferencias encontradas sobre las poblaciones y territorios asociados, los siete linajes presentan variabilidad en la secuencia nucleotídica, reflejada en polimorfismos de nucleótido simple (SNP) (Coscolla and Gagneux 2014).

Sumado a las diferencias señaladas en relación a la distribución geográfica, los genotipos de MTB muestran diferencias en el tipo de TB causada (Caws et al. 2008), tasa de crecimiento en pulmones (Krishnan et al. 2011), cantidad inducida de IL-1 $\beta$  y

TNF- $\alpha$  (Krishnan et al. 2011) y en su capacidad de adquirir resistencia (Nguyen et al. 2016; Yuen et al. 2013). Estos resultados indican que los linajes de MTB se comportan de manera distinta en interacción con el hospedero y bajo presión de medicamentos.

Los sistemas toxina-antitoxina (TA), son operones compuestos de dos genes, uno que codifica una toxina y otro que codifica una antitoxina (Zhu et al. 2010). La toxina tiene un efecto bacteriostático o bactericida que es neutralizado por la antitoxina (Goeders et al. 2016). Bajo distintas condiciones de estrés, la antitoxina es degradada, dejando a la toxina libre para que inhiba un proceso celular esencial, disminuyendo el crecimiento celular o incluso deteniendo el ciclo celular (Lee and Lee 2016; Sala, Bordes, and Genevoux 2014). Los procesos que son blanco de los sistemas TA incluyen la síntesis de proteínas, replicación del ADN, división celular y síntesis de pared celular (Lobato-Márquez, Díaz-Orejas, and García-Del Portillo 2016). Gracias a este mecanismo de acción, los sistemas TA se asocian a la formación de células persistentes (Gelens et al. 2013) favoreciendo el desarrollo de resistencia a medicamentos y el surgimiento de latencia (Gelens et al. 2013; Sala et al. 2014). Por ejemplo, en MTB los sistemas vapBC3, vapBC31 y higBA2 son sobre expresado en presencia de D-cicloserina (Keren et al. 2011).

La función de sistemas TA ha sido ligada a condiciones de estrés oxidativo y nitrosativo, pH ácido y falta de nutrientes. Debido a que estas condiciones son similares a las generadas por la respuesta del hospedero, los sistemas TA pueden ser expresados en microambientes encontrados dentro de diferentes tipos de células durante la interacción con el hospedero (Lobato-Márquez et al. 2016).

MTB tiene un amplia gama de sistemas TA en comparación con mycobacterias fuera del complejo *Mycobacterium tuberculosis* (MTBC), como *M. smegmatis*, *M. leprae*, *M. ulcerans* y *M. avium* (Albrethsen et al. 2013; Ramage, Connolly, and Cox 2009). Estos hallazgos conducen a la hipótesis de que la regulación del metabolismo por sistemas TA ha beneficiado a MTB en su coevolución con el humano (Albrethsen et al. 2013). Evidencia que soporta esta hipótesis está dada por la sobre expresión de distintos sistemas TA bajos condiciones de estrés como hipoxia, estrés oxidativo, estrés nutricional, interacción con macrófagos y presión de medicamentos (Albrethsen et al. 2013; Sala et al. 2014). Indicando que los sistemas TA pueden estar implicados en el surgimiento de células persistentes en MTB.

A pesar de los estudios dirigidos a conocer la función de los sistemas TA en MTB, que muestran su abundancia en la bacteria y su asociación con la persistencia bajo situaciones de estrés como las provocadas por el sistema inmune y los antibióticos, no se ha realizado un análisis de la variabilidad de los sistemas en los linajes que la componen. Nuestro objetivo fue evaluar la variabilidad en secuencias proteicas de los sistemas TA en distintos linajes de MTB utilizando modelos ocultos de Markov para la predicción por homología de sistemas TA seguido de un alineamiento múltiple para la búsqueda de mutaciones.



# 1. Análisis de sistemas toxina-antitoxina

## 1.1 Marco teórico

### 1.1.1 La Tuberculosis

La TB es una enfermedad infecciosa que ha acompañado a la humanidad durante su historia (Lee 2016). La TB afecta principalmente los pulmones, sin embargo puede desarrollarse de forma extra pulmonar afectando cualquier órgano (Lawn and Zumla 2011). Su ocurrencia depende del estado inmune del paciente, genotipo de la cepa, edad, raza del paciente y se tiene un 10% de posibilidad de desarrollar la enfermedad, después de la infección (Caws et al. 2008). El riesgo de transmisión es alto en los primeros años después de la infección pero disminuye con el paso del tiempo. En la mayoría de individuos inmunocompetentes la bacteria es eliminada o permanece en forma latente (Lawn and Zumla 2011). Los síntomas y signos incluyen aquellos asociados con sitios específicos de la enfermedad al igual que fiebre y pérdida de peso (Lawn and Zumla 2011), sin embargo, en los estados tempranos de la enfermedad el paciente puede ser asintomático (Lawn and Zumla 2011).

El bacilo puede ser transmitido en gotas que permanecen en el aire por varias horas y pueden ser inhaladas y localizarse en los alveolos. Las secreciones que rodean al bacilo cuando está en el ambiente, lo protegen de la deshidratación, radiación y otros tipos de estrés medioambientales (Lee 2016). La respuesta mediada por células inmunes se inicia de 2 a 10 semanas después de la infección con MTB y se confirma con la prueba de tuberculina. La mayoría de las personas infectadas no desarrollan la enfermedad o desarrollan formas autolimitadas de esta (Lee 2016).

La interacción de MTB con el hospedero desencadena la respuesta inmune mediada por células T y macrófagos, los cuales tienen el principal rol en la protección contra MTB. En individuos inmunocompetentes la acción de las células termina en la formación del granuloma, el cual es altamente efectivo en contener, mas no en eliminar la infección. Dentro del granuloma la bacteria es capaz de protegerse de las condiciones que genera el sistema inmune para atacarla (Lawn and Zumla 2011).

### 1.1.2 Patógeno

MTB es parte del MTBC, grupo de micobacterias que se caracterizan por poseer más del 90% de similitud a nivel de nucleótidos y una secuencia de RNAr 16S idéntica, pero varían ampliamente en términos del hospedero, fenotipos y patogenicidad (Brosch et al. 2002). Se especula que el origen del MTBC está asociado al este de África donde pudo afectar a los primeros homínidos. La expansión del complejo al resto del mundo pudo haber coincidido con la migración de los humanos fuera de África y posteriormente un regreso a África desde Asia (Brudey et al. 2006).

Distintos estudios enfocados en la variabilidad de la secuencias nucleotídica entre bacterias del MTBC han mostrado que, aunque la variabilidad genómica global es baja comparada a otras bacterias, la mutaciones acumuladas en linajes filogenéticos dentro del complejo pueden explicar las diferencias fenotípicas (Coscolla and Gagneux 2014). Un amplio rango de mutaciones ocurren en MTBC, incluyendo SNPs, deleciones, duplicaciones y elementos móviles (Coscolla and Gagneux 2014).

Como tal, MTB es un patógeno intracelular obligatorio que infecta principalmente a la raza humana. Es un bacilo aerobio que no forma esporas ni presenta cápsula. Su crecimiento es más exitoso en tejidos con alto contenido de oxígeno como los pulmones (Lawn and Zumla 2011). Su división se realiza cada 15 a 20 horas y presenta la capacidad de permanecer dentro del hospedero en forma latente (Lawn and Zumla 2011) y evadir el estrés oxidativo producido por las células del hospedero que buscan destruir los microorganismos patógenos (Forrellad et al. 2013). La envoltura celular de MTB consiste de tres segmentos: la membrana plasmática, la pared celular y la capa más externa. El núcleo de la pared celular está constituida por peptidoglucanos en unión covalente con arabinogalactanos (Kaur et al. 2009). La capa más externa consiste de una mezcla de lípidos, polisacaridos y lipoglicanos unidos de forma no covalente (Kaur et al. 2009).

MTB está dividido en 7 linajes, adaptados a poblaciones humanas específicas (Figura 1) (Coscolla and Gagneux 2014). Gracias a la secuenciación de genomas completos (WGS) se ha logrado dividir los siete linajes en dos grupos, un grupo de linajes modernos, constituido por los linajes 2, 3 y 4, y un grupo de linajes antiguos constituido por los linajes 1, 5 y 6 (Figura 1). El linaje 2, también conocido como Beijing, se encuentra principalmente en el este de Asia, pero igualmente está presente en Asia central, Rusia y África del sur.

El linaje 4 o Euro-América es frecuente en poblaciones de África, Europa y América. Por su parte, los linajes 1 y 3 muestran distribución limitada al este de África, centro y sur este de Asia (Figura 1). Los linajes pueden estar asociados con diferentes manifestaciones de la enfermedad, por ejemplo Click et al. 2012 encontraron que los linajes Indo-Oceánico y Este de África provocan con más frecuencia TB extra pulmonar mientras que el linaje 4 está asociado a TB pulmonar (Coscolla and Gagneux 2010).

Aunque siguen habiendo muchas preguntas acerca de los orígenes de MTB, avances en genómica están ahora proporcionando evidencia de que la cantidad de variaciones en la secuencia del genoma de MTB ha sido subestimada y que parte de esta diversidad genética, tiene importantes consecuencias fenotípicas (Lawn and Zumla 2011). La cepa H37Rv derivada de H37 es una de las cepas de laboratorio mejor caracterizadas (Steenken, Oatway, and Petroff 1934). Su genoma, fue publicado en 1998 (Cole et al. 1998) permitiendo identificar secuencias repetidas, elementos de inserción y 4066 genes

que codifican familias de proteínas involucradas en la biosíntesis de ácidos grasos y policétidos, regulación, bombas de flujo para medicamentos (Ioerger et al. 2010) y las familias de genes PE/PPE, genes propios del MTB (McEvoy et al. 2012). El genoma de MTB tiene un alto contenido GC (Sreenu et al. 2007) que hace difícil su manejo a la hora de secuenciar y ensamblar, dando como resultado genomas más fragmentados (Chen et al. 2013).

El genoma de la cepa H37Rv fue el primero en ser secuenciado. Desde su publicación el número de genomas disponibles se ha incrementado, a la fecha hay 4745 genomas de MTB ensamblados, anotados y disponibles en la plataforma del NCBI (<https://www.ncbi.nlm.nih.gov>). Del total de genomas, solo 54 tienen carácter completo, es decir, la secuencia de su genoma ha sido dilucidada enteramente. El primer genoma secuenciado de un aislamiento colombiano fue publicado en 2012 (Isaza et al. 2012), el aislamiento denominado UT205 fue aislado de un paciente de 33 años de edad. Posteriormente, se ha publicado el genoma de un aislamiento XDR (Alvarez et al. 2016), un aislamiento del linaje Africa2 (Hurtado et al. 2016) y un aislamiento Beijing (Rodríguez et al. 2015).

La secuenciación de genomas y la disponibilidad de secuencias en bases de datos, permiten la comparación de distintos genotipos de MTB con el fin de dar luces en su evolución y comportamiento. La comparación de genomas de *M. africanum*, *M. bovis*, *M. canettii*, *M. caprae*, *M. orygis*, y MTB permitió establecer el core de genes entre estas bacterias y el impacto de las mutaciones en las proteínas que codifican dichos genes. Los resultados arrojados mostraron que los genes implicados en metabolismo de lípidos y respiración han acumulado la mayoría de las mutaciones. Los cambios encontrados están asociados con variaciones en flexibilidad/ductilidad que pueden explicar características fenotípicas (Yruela et al. 2017). En otro estudio se analizó los datos de secuenciación de 1082 aislamientos de MTB del linaje Beijing logrando identificar 44 SNPs y dos deleciones que separan aislamientos Beijing modernos de aislamientos Beijing antiguos (Liu et al. 2016).

Uno de los principales aportes de la secuenciación de genomas, es la determinación de resistencia en MTB. El genoma de aislamientos tomados de pacientes con TB puede ser comparado con bases de datos con el fin de conocer si posee mutaciones asociadas a resistencia y de esta forma guiar el diagnóstico y tratamiento de un paciente (Lee and Behr 2016). En un estudio publicado en 2015 se encontró que la WGS permite identificar la resistencia a Isoniacida y Rifampicina con una sensibilidad de 92.8% y 96.2% respectivamente (Coll et al. 2015). Sin embargo, la sensibilidad para detectar Pirazinamida fue de solo 70,9% (Coll et al. 2015).

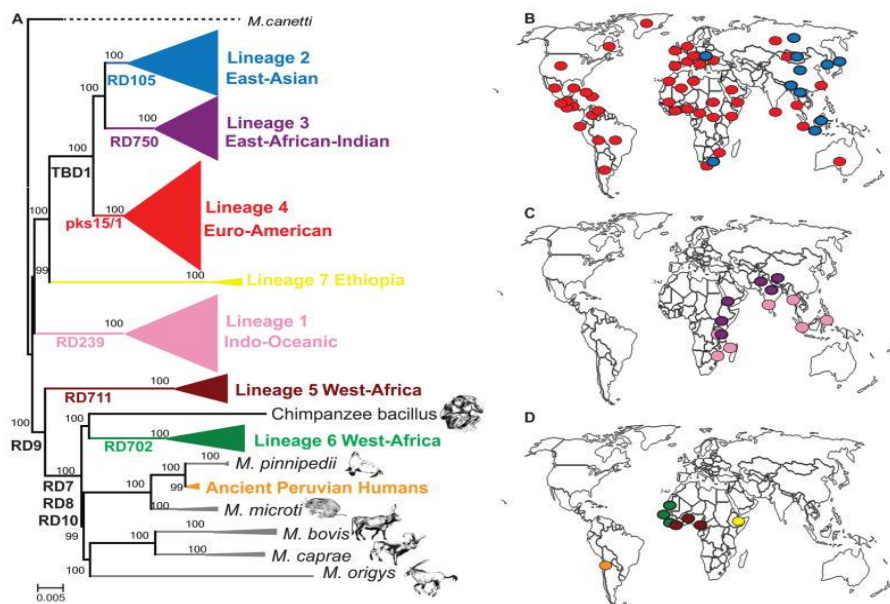
Otra de las ventajas de la WGS es su aplicación en estudios de salud pública. Por ejemplo, en un estudio realizado en Canadá sobre un brote de TB, se secuenciaron 32 aislamientos de MTB con el fin de establecer la dinámica del brote. Los datos mostraron la existencia de dos linajes de MTB y pusieron en evidencia la limitada resolución de MIRU-VNTR (Gardy et al. 2011). En otra investigación se utilizó WGS como herramienta en un estudio retrospectivo. La secuencia del genoma de 390 aislamientos permitió el análisis de patrones de mutaciones e inferir la dirección de la transmisión del brote (Walker et al. 2013).

### 1.1.3 Sistemas toxina-antitoxina

En bacterias, la regulación del crecimiento celular y de la muerte celular es importante bajo condiciones de estrés variadas. Casi todas las bacterias, contienen un gen que codifica una pequeña toxina de aproximadamente 100 aminoácidos. Esta toxina inhibe el crecimiento bacteriano al atacar un proceso celular esencial. Entre los procesos que inhiben esta la replicación del ADN, traducción, biosíntesis de la pared celular y síntesis de ATP. Estas toxinas son transcritas con una antitoxina en forma de operón llamado toxina-antitoxina (TA) (Y Yamaguchi, Park, and Inouye 2011).

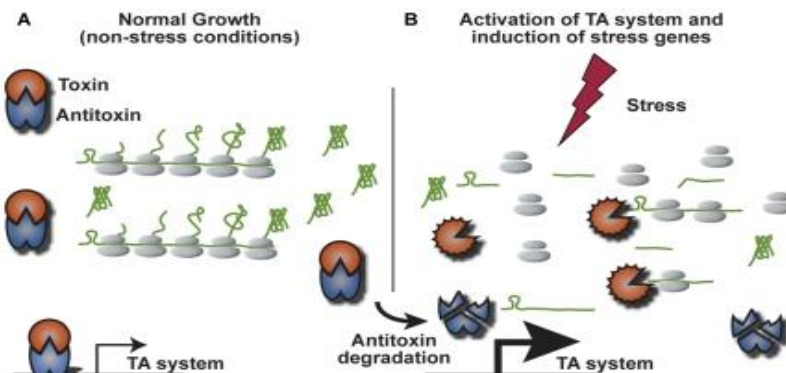
En condiciones normales de crecimiento, la toxina y su antitoxina forman un estable complejo que impide que la toxina ejerza su toxicidad. Bajo condiciones de estrés, la antitoxina es degradada dejando a la toxina libre para que ejerza su efecto toxico, lo que conduce a la detención del crecimiento (Figura 2) (Page and Peti 2016). Los sistemas TA no son esenciales para el crecimiento normal, sin embargo, ellos son ampliamente distribuidos en bacterias y arqueas (Y Yamaguchi et al. 2011) donde aseguran la supervivencia bajo condiciones de estrés (Page and Peti 2016).

**Figura 1:** Distribución geográficas de linajes (Coscolla and Gagneux 2014).



El primer sistema TA identificado fue encontrado en un plásmido y se observó que tenía un rol en la prevalencia del plásmido. Cuando dicho plásmido es perdido en una célula, la toxina es liberada del complejo TA, dado que la antitoxina es más inestable que la toxina. Esto conlleva a la inhibición del crecimiento y eventualmente a la muerte celular (Y Yamaguchi et al. 2011).

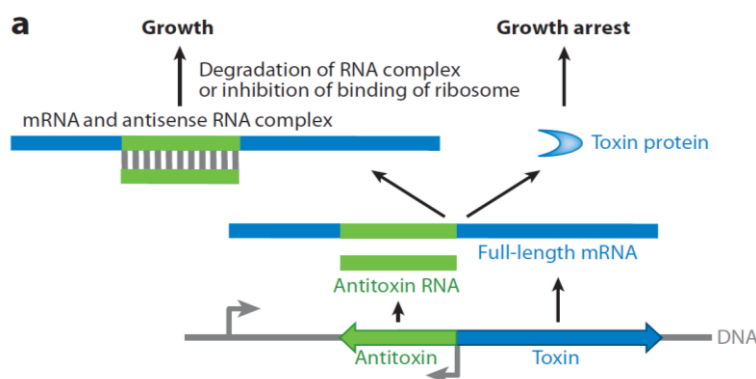
**Figura 2:** Esquema de la respuesta a estrés de sistemas TA (Ramage et al. 2009). A) Bajo condiciones normales de crecimiento, la toxina y la antitoxina forman un complejo que controla su transcripción, B) en condiciones de estrés la antitoxina es degradada permitiendo que la toxina se libere para inhibir un proceso esencial en la célula, en este caso la traducción.



Se han identificado 5 principales tipos de sistemas TA de acuerdo al mecanismo de inhibición que utiliza la antitoxina (Sala et al. 2014).

En el sistemas tipo I la antitoxina es un RNA antisentido que específicamente inhibe la síntesis de la toxina al interactuar con el RNAm de la toxina (Figura 3). Ejemplos de estos sistemas son RdlD-LdrA, SymR-SymE y IstR-TisB. El RNA antisentido se une al RNAm de la toxina formando una molécula de doble hebra. Este RNA de doble hebra estimula la degradación del RNAm de la toxina. Todas las toxinas del tipo I son pequeños péptidos hidrofóbicos de 19 a 38 aminoácidos (Y Yamaguchi et al. 2011).

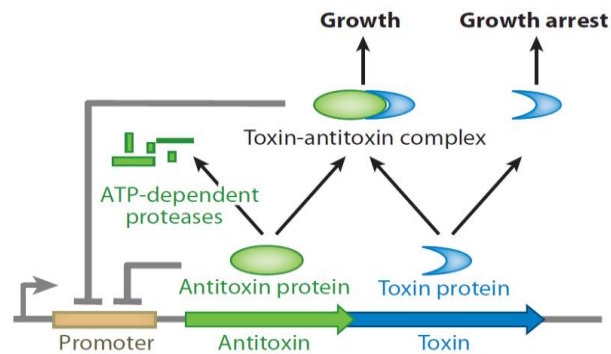
**Figura 3:** Mecanismos de inhibición de sistemas tipo I (Y Yamaguchi et al. 2011).



En el sistema tipo II la antitoxina es una proteína que se une con la toxina formando un complejo (Sala et al. 2014) (Figura 4). Entre los sistemas TA, los tipos II son los más ampliamente estudiado. mazE-mazF, relE-relB, YefM-YoeB y MqsR-MqsA son ejemplos de esos sistemas. La mayoría de los sistemas tipo II comparten un número de características comunes: a) los genes solapa en el operón por unas pocas bases o son

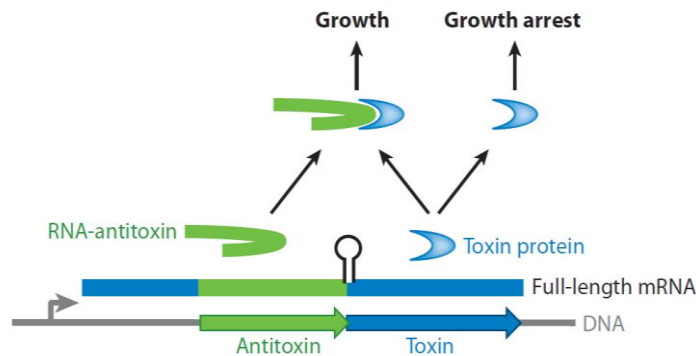
separados por pocas bases; b) la antitoxina es menos estable que la toxina; c) El complejo TA funciona como un represor transcripcional del operón TA; d) en la mayoría de los sistemas, la antitoxina se ubica corriente arriba de la toxina, de tal manera que la antitoxina parece tener una ventaja en su producción sobre la toxina (Y Yamaguchi et al. 2011).

**Figura 4:** Mecanismos de inhibición de sistemas tipo II (Y Yamaguchi et al. 2011)



En el sistema tipo III la antitoxina es un RNA codificado por una repetición en tándem corriente arriba de la toxina que interactúa directamente con la toxina formando un complejo proteína-RNA (Figura 5). El sistema ToxI-ToxN fue el primer sistema tipo III para el cual se detalló su mecanismo molecular, fue descubierto en el patógeno *Pectobacterium atrosepticum*. La toxina ToxN es una ribonucleasa que comparte similitud con la toxina MazF del sistema MazEF tipo II (Wen, Behiels, and Devreese 2014).

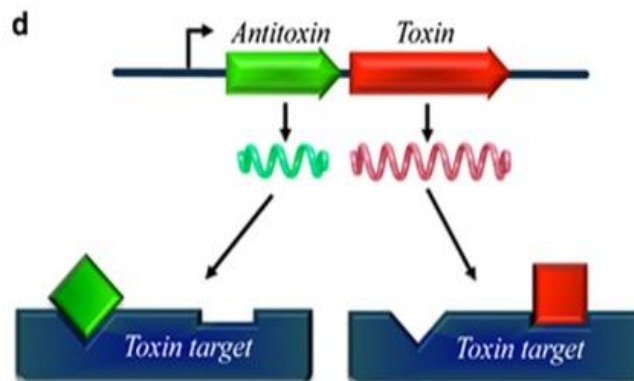
**Figura 5:** Mecanismos de inhibición de sistemas tipo III (Y Yamaguchi et al. 2011).



El sistema tipo IV se caracteriza porque las toxinas son proteínas que tienen el mismo blanco que la toxina, pero no interactúan entre ellos (Sala et al. 2014) (Figura 6). Ejemplo de este sistema es el módulo YeeU/YeeV (CbeA/CbtA), involucrado en la regulación de proteínas del citoesqueleto. La toxina CbtA inhibe la polimerización de las proteínas MreB y FtsZ. La antitoxina YeeU causa el efecto contrario al interactuar con dos proteínas que promueven el ensamble de los filamentos. Bajo condiciones de estrés, YeeU es

degradada, permitiendo a CbtA unirse a su blanco resultando en un citoesqueleto alterado (Allocati et al. 2015).

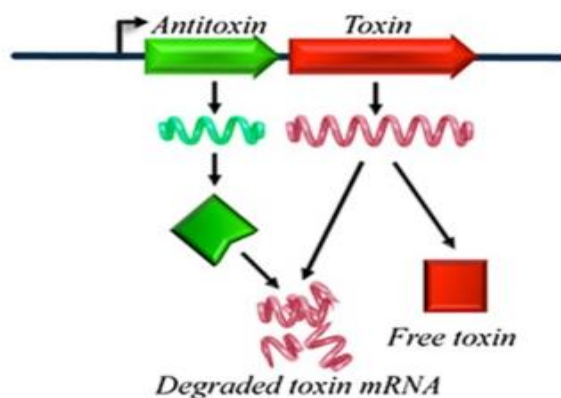
**Figura 6:** Mecanismos de inhibición de sistemas tipo IV (Allocati et al. 2015).



En los sistemas tipo V, la antitoxina es una ribonucleasa que degrada específicamente el RNAm de la toxina (Sala et al. 2014) (Figura 7).

El sistema GhoS/GhoT es representante de este tipo de sistemas y está presente en *E. coli* (Allocati et al. 2015).

**Figura 7:** Mecanismos de inhibición de sistemas tipo V (Allocati et al. 2015).



### 1.1.4 Blancos de Toxinas

Los blancos celulares de las toxinas codificadas por los sistemas TA son ampliamente variables y la toxicidad de cada una es activada bajo diferentes condiciones. Conocer los blancos de las toxinas son fundamentales para el entendimiento de los sistemas TA (Yamaguchi and Inouye 2016).

## ADN

En el 2014 se reportó el sistema tipo I RalR/RalA. La antitoxina RalA codifica un RNA pequeño con 16 nucleótidos complementarios al RNAm de la toxina. La toxina RalR funciona como una endonucleasa que corta ADN metilado y no metilado (Guo et al. 2014). Se ha demostrado que la toxina ralR es capaz de cortar el ADN de *E.coli* e inhibir el crecimiento de esta bacteria (Yamaguchi and Inouye 2016).

### Replicación del ADN

Las toxinas parE y CcdB bloquean la replicación del ADN al inhibir la acción de la ADN girasa la cual es esencial en el crecimiento. CcdB impide la conformación funcional de GyrA al unirse al dominio de dimerización de la proteína. Las toxinas parE parecen actuar de manera diferente a CcdB ya que una ADN girasa resistente a la acción de CcdB es inhibida por parE en *E.coli* (Yamaguchi and Inouye 2016). Otra diferencia, se muestra en *Vibrio cholerae*, en este organismo se encontró que a diferencia de CcdB, parE2 es dependiente de ATP (Yamaguchi and Inouye 2016).

## RNA

Entre los blancos celulares de los sistemas TA, el RNAm es el más frecuente de todos (Yamaguchi and Inouye 2016). Las familias higBA, relBE y YefM/YoeB codifican reibonucleasa (toxinas) que son dependientes de unión al ribosoma para realizar su tarea (Sala et al. 2014). Las toxinas relE y YoeB de *E. coli* cortan el RNAm en el sitio A del ribosoma. relE se une a la subunidad 30S del ribosoma e inhibe la elongación mientras que YoeB se une a la subunidad 50S e inhibe la iniciación de la traducción. La toxina higB del plasmido Rts1 corta el RNAm generalmente en secuencias AAA y al igual que YoeB, higB se une a la subunidad 50S (Sala et al. 2014).

Las familias vapBC y mazEF codifican ribonucleasas que no necesitan de unión al ribosoma. Las toxinas vapC1, vapC2, vapC5, vapC11 y vapC20 han mostrado actividad ribonucleasa in vitro. Las toxinas VapC1 y 19 de MTB cortan regiones ricas en GC. Además de dirigirse al RNAm, en *Salmonella entérica* se encontraron toxinas vapBC que degradan el RNAt (Sala et al. 2014). La toxina mazF, descubiertas en *E. coli*, se dirige al RNAm cortando secuencias ACA. Hasta el momento, las toxinas mazF identificadas cortan el RNAm en secuencias de 3, 5 y 7 nucleótidos pero no se conoce el porqué de dirigirse a secuencias impares (Yamaguchi and Inouye 2016).

Se ha evidenciado que la toxinas mazF-mt6, una de las toxinas homologas en MTB es capaz de cortar el RNAr 23S en la secuencia UUCCU lo que inhibe la traducción a proteínas (Yamaguchi and Inouye 2016).

## Ribosomas

El sistema Phd-Doc es derivado del bacteriófago P1. La expresión de Doc inhibe el crecimiento celular al interferir en la traducción. Doc no actúa sobre el RNAm como otras toxinas, en su lugar, la toxina funciona de manera similar a la del aminoglicosido



Higromicina B, inhibiendo la elongación de la traducción por unión al ribosoma (Yamaguchi and Inouye 2016).

### 1.1.5 Sistemas toxina-antitoxina en *Mycobacterium tuberculosis*

Las mycobacterias, proporcionan una mirada única de la adquisición, evolución y roles biológicos de loci toxina-antitoxina en procariontes. MTB tiene un genoma densamente poblado de sistemas TA (Gerdes 2013). Hay debate acerca del número exacto de sistemas que tiene el genoma de MTB (Gerdes 2013), aunque en la cepa de referencia H37Rv se han identificado 79 sistemas (Sala et al. 2014) (Figura 8). En ella 50 sistemas son de la familia vapBC, 10 pertenecen a la familia mazEF, los sistemas relBE, higBA y parDE tienen cada uno dos representantes, hay un sistema YefM/YoeB, un sistema tipo TAC, tres sistemas tipo IV y 8 sistemas hipotéticos (Sala et al. 2014).

#### Familia vapBC (TA Tipo II)

La familia vapBC, del inglés “Virulence associated protein” fue descubierta como un sistema de adicción en *Salmonella dublin*. Esta familia es la más abundante en MTB (Figura 8), los operones se encuentran constituidos por una toxina vapC, la cual tiene un dominio PIN encontrados en proteínas de todos los reinos y asociado a ribonucleasas. Las secuencias específicas de corte de las toxinas vapC1 y vapC19 de MTB han sido estudiadas. Ambas toxinas cortan regiones de 4mers ricas en GC.

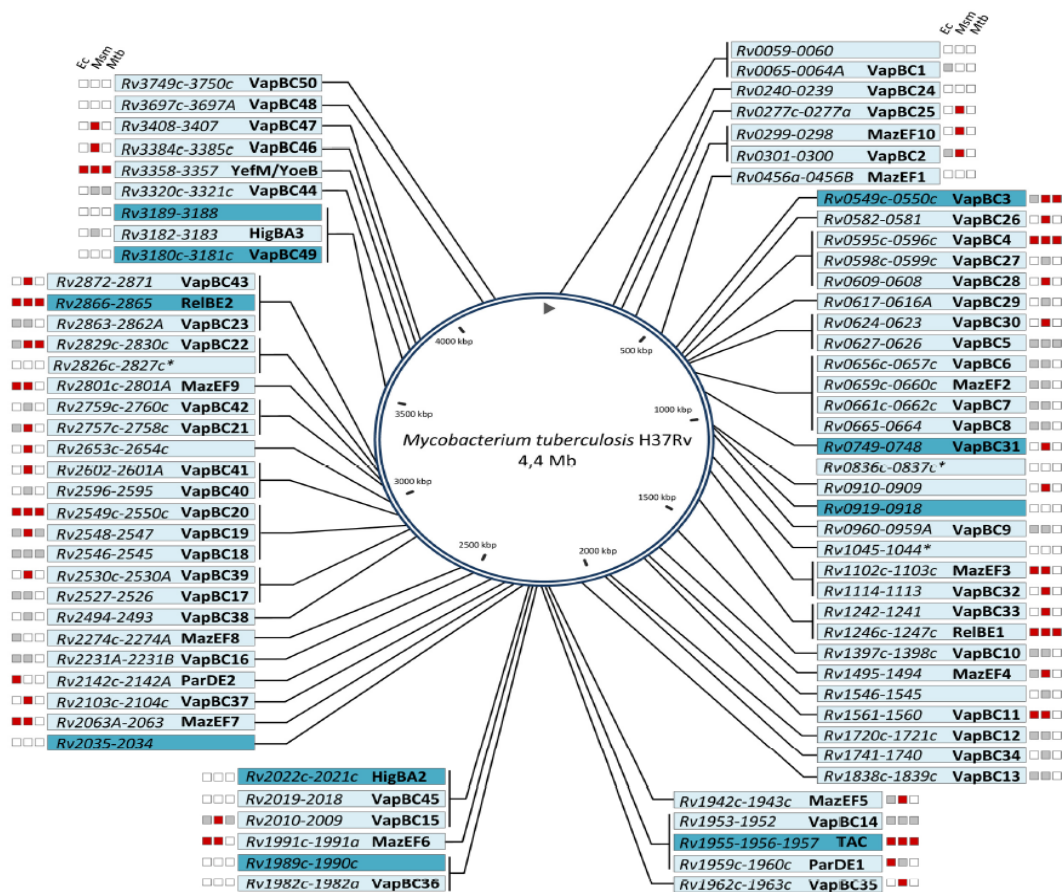
El cristal del complejo vapBC3 reveló un hetero-octámero formado por el ensamble de dos hetero-tetrámeros formados por dos toxinas y dos antitoxinas similar a la estructura del complejo vapBC en *Shigella flexneri*.

El sitio catalítico de las toxinas vapC está formado por residuos ácidos del dominio PIN que coordinan el ion  $Mg^{2+}$ . El mecanismo de inhibición de las antitoxinas vapB, se basa en impedir la unión del  $Mg^{2+}$  al sitio activo de la toxina. Las antitoxinas vapB de MTB son generalmente relacionadas a reguladores transcripcionales y a dominios de unión a ADN (Sala et al. 2014).

#### Familia mazEF (TA Tipo II)

En *E. coli*, la sobre expresión de las toxinas mazF disminuyen la tasa de crecimiento mediante la degradación del RNAm. El complejo mazEF en *E. coli* forma un hetero-hexámero compuesto de un dímero de la antitoxina mazE unido a dos dímeros de la toxina mazF. El análisis sobre las secuencias que cortan las toxinas en MTB reveló significativas diferencias. Mientras mazF9 corta secuencias UAC, mazF3 corta regiones ricas en uracilo, mazF6 escinde secuencias UUCU y mazF4 lo hace en UCGU. Tales diferencias en el reconocimiento de sustrato, sugieren que las toxinas mazF pueden desencadenar respuestas diversas (Sala et al. 2014).

**Figura 8:** Sistemas TA reportados en H37Rv (Sala et al. 2014).



Se ha evidenciado que la toxina mazF9 es neutralizada por las toxinas mazE6, vapB27 y vapB40. Además, este solapamiento en funciones está de acuerdo con la baja conservación entre distintas antitoxinas mazE y con el hecho de que vapB27 y vapB40 son cercanamente relacionados a la antitoxina mazE de *E.coli*. Este tipo de reacciones son esperadas debido a la gran cantidad de sistemas TA en MTB (Sala et al. 2014).

**Familias relBE y YefM/YoeB (TA Tipo II)**

Las toxinas relE y YoeB en *E.coli* son ribonucleasas dependientes de ribosoma que cortan el RNAm en el sitio A del ribosoma. Sin embargo, a pesar de la similitud de las dos toxinas, relE y YoeB parecen actuar por distintos mecanismos: relE se une a la subunidad 30S e inhibe la elongación de la traducción, mientras que YoeB se une a la subunidad 50S e inhibe el inicio de la traducción.

MTB codifica dos sistemas relBE y uno YefM/YoeB. Análisis sobre el complejo YefM/YoeB sugirieron que este podría formar un complejo hetero-trimétrico, como es el caso de su homólogo en *E. coli*. Estudios estructurales de relBE2 revelaron un hetero-tetrámero compuesto de dos hetero-dímeros de una toxina unida a una antitoxina. Las

toxinas relE y YoeB muestran características de plegamiento similares a las toxinas relE, pero sus antitoxinas (relB2 y YefM) son relacionadas a las antitoxinas YefM/Phd y no a la relB de *E. coli*.

#### Familia higBA (TA Tipo II)

El primer locus higBA (del inglés host inhibition of growth) se identificó en el plásmido Rts1 de *Proteus vulgaris*. Se mostró que la toxina higB del plásmido Rts1 actúa como una ribonucleasa dependiente de ribosoma con un mecanismo distinto a relE y YoeB. La toxina higB se une a la subunidad 50S y corta secuencias AAA del RNAm que está siendo procesado. MTB H37Rv posee tres sistemas TA homólogos a higBA, dos son sistemas de dos componentes y el restante es un sistema atípico de tres componentes llamado TAC. No se conoce nada de la participación de sistemas higBA en la patogénesis de MTB.

#### Familia parDE (TA Tipo II)

MTB H37Rv posee dos sistemas parDE. El sistema fue originalmente descubierto en el plásmido RK2. La toxina parE, a diferencia de las anteriores, no tiene por blanco el RNAm, en su lugar, inhibe la ADN girasa bloqueando la replicación del ADN. El cristal del complejo parDE1 de *Caulobacter crescentus* reveló un hetero-tetrámero compuesto de dos homo-dímeros de toxina y antitoxina. La antitoxina parD1 de MTB tiene un dominio RHH de unión al ADN mientras que parD2 no parece tener dominios conservados (Sala et al. 2014).

#### Sistema de tres componentes (TA Tipo II)

El sistema toxina-antitoxina-chaperona (TAC) de MTB es codificado por tres genes organizados en un operón, los genes *Rv1955-Rv1956-Rv1957*, codifican el sistema higBA1 y la chaperona *Rv1957*. Se ha demostrado que el sistema higBA1 es controlado de manera específica por la chaperona a través de una interacción directa con la antitoxina, permitiendo el plegamiento de la antitoxina para protegerse de la degradación (Sala et al. 2014). En *E. coli* y en *Mycobacterium marinum* se mostró que la toxicidad de la toxina fue inhibida cuando se expresó en conjunto *Rv1957* y la antitoxina higA (Bordes et al. 2011).

La transcripción del operón TAC es inducido por condiciones variadas de estrés, incluyendo daño de ADN, hipoxia, medicamentos e interacción con fagocitos. El operón está bajo el control de dos promotores: el promotor *higB1P1* localizado a 51 nucleótidos corriente arriba del codón de inicio de la toxina, por lo tanto controla la expresión de todo el operón TAC. El segundo promotor es *higB1P2* localizado a 29 nucleótidos corriente arriba del gen *Rv1954a* (Sala et al. 2014).

### 1.1.6 Antecedentes

Los estudios sobre los sistemas Toxina-Antitoxina, se han enfocado principalmente en el papel de expresión de genes particulares sobre el metabolismo de MTB. No obstante, se han realizado trabajos utilizando las secuencias de los sistemas TA. Ramage, Connolly, and Cox 2009 realizaron una predicción de genes TA en el genoma de MTB mediante un enfoque de homología por BLAST y búsqueda de motivos conservados, encontrando 88 sistemas putativos TA. Un segundo análisis en este estudio se realizó al comparar los módulos TA entre varios genomas del genero *Mycobacterium*, los resultados mostraron al operón Rv0909-0910 como único sistema conservado entre todos los genomas. Otros resultados sobre este mismo análisis, revelaron que especies pertenecientes al complejo MTBC como, *M. bovis*, *M. africanum* y *M. canetti* presentan diferencias en la cantidad y composición de TA presentes en el genoma.

En otro estudio se analizaron 142 genes pertenecientes a las familias VapBC, MazEF, HigAB, RelBE y ParDE sobre 173 genomas secuenciados de MTB. El análisis se centró en la búsqueda de polimorfismos de nucleótido simple (SNP) utilizando BLAST. De los 142 genes evaluados, 106 tienen polimorfismos, 42,2 % fueron mutaciones no sinónimas y 28,8% tiene inserciones y deleciones que cambian el marco de lectura. Con las mutaciones encontradas, los autores. Con el conjunto de resultados, los autores mostraron que MTB puede ser dividido en 9 principales genotipos: Beijing, CAS, EAI, Haarlem, S, LAM, T, Ural y X utilizando SNPs de genes TA. Además, proponen un grupo de 13 polimorfismos que pueden ser utilizados en la genotipificación de MTB. La genotipificación utilizando estos polimorfismos coincide con la hecha en base a SNPs de genes housekeeping (Zaychikova et al. 2015).

El sistema toxina-antitoxina-chaperona (TAC) de MTB fue analizado para encontrar SNPs asociados a los genes que lo codifican (Xiao et al. 2016). La secuencia de los genes *higA*, *higB* y *Rv1957* fueron obtenidas por secuenciación de 183 aislamientos clínicos de MTB. Se encontraron 6 SNPs no sinónimos y 2 sinónimos. Las mutaciones sinónimas están en *higA* y las no sinónimas se reparten en los tres genes. Las mutaciones sinónimas y 4 de las no sinónimas fueron asociadas al linaje Beijing (Xiao et al. 2016).

Leplae et al. 2011a Realizaron una búsqueda de genes TA sobre 2181 genomas procariontas con el fin de conocer la relación de estos sistemas con distintas especies. La búsqueda se realizó sobre cada una de las secuencias de referencia (proteínas TA) sobre una base de datos conteniendo las proteínas de los 2181 organismos. La búsqueda se hizo con PSI-BLAST con cuatro iteraciones y un corte de E-value de 0.001. Hits con menos del 50% de sus residuos alineados fueron excluidos. En la construcción de familias se realizó una comparación de todos contra todos con BLASTP, los pares seleccionados fueron la entrada para un agrupamiento de Markov.

Los autores encontraron 20 aislamientos con el más alto número de sistemas TA. Cinco fueron aislamientos patógenos del complejo *Mycobacterium tuberculosis*. Según estos resultados, el 1.1% del genoma de los 5 aislamientos codifica para sistemas TA. Los autores encontraron gran diferencia entre el número de sistemas encontrados en aislamientos del complejo (entre 43 y 45) y las predichas en *Mycobacterium leprae* (un solo sistema).

Un estudio realizado por Fozo et al. 2010a se enfocó en el análisis de sistemas tipo I. La búsqueda la realizaron con PSI-BLAST y TBLASTN sobre 774 genomas bacterianos para identificar sistemas homólogos a sistemas tipo I de referencia. La predicción además se

---

basó sobre características de los sistemas tipo uno como la longitud no más larga a 70 aminoácidos ni más pequeña que 16 residuos. Los autores no encontraron sistemas tipo I en el genoma de MTB.

Blower et al. 2012 realizaron una búsqueda sobre proteomas de bacterias y arqueas descargados de la base de datos integr8 (Kersey et al. 2004). La predicción se hizo basado en características extraídas del cristal del sistema ToxIN (PDB: 2XDD). Las secuencias que cumplían con las características fueron agrupadas utilizando BLASTp. No se encontraron proteínas homologas a sistemas tipo I en MTB.

Rose et al. 2013 realizaron una comparación del genoma y transcriptoma de cepas pertenecientes a los linajes 1 y 2, encontrando polimorfismos de nucleótido simple (SNP) no sinónimos en los genes que codifican para *vapC10* en el linaje 2 y *mazF7* en el linaje 1.

Otro acercamiento en la utilización de las secuencias de TA se hizo con los genes RelEB. Los investigadores construyeron un árbol filogenético utilizando un alineamiento múltiple de las secuencias proteicas. Las secuencias mostraron dos grupos principales, el primero formado por las toxinas de MTB y la proteína YoeB de *E.coli* y el segundo por las toxinas de arqueas y la toxina de *E.coli*. Los autores mencionan que la distancia entre las toxinas de MTB dentro del primer cluster podría indicar una posible divergencia en mecanismos catalíticos (Miallau et al. 2013).

Nuestro laboratorio, previamente ha encontrado mutaciones relacionadas con los sistemas TA; la comparación del genoma de seis aislamientos clínicos con fármaco resistencia múltiple, reveló la delección de los genes *Rv2274c-Rv2274a*, secuencias que codifican para el sistema *mazF8*. Basados en el fenotipo de los aislamientos, concluimos que la delección del sistema no afecta la capacidad de la bacteria de adquirir resistencia y tampoco altera drásticamente el desarrollo de la enfermedad (datos no publicados).

## 1.2 Metodología

### 1.2.1 Selección de aislamientos

Los aislamientos fueron tomados de la colección de la unidad de Bacteriología y Micobacterias de la Corporación para Investigaciones Biológicas (CIB). Se tomaron 18 aislamientos pertenecientes a los linajes Beijing, Euro-América, EAI y Africa-2 (Tabla 1). La colección no cuenta con aislamientos de los linajes CAS, Africa1 y Etiopia. Los genomas representativos de estos linajes fueron tomados de bases de datos (subtitulo 1.2.5).

### 1.2.2 Extracción de ADN y secuenciación de genomas

Los aislamientos conservados a  $-70^{\circ}\text{C}$  se reconstituyeron en Middlebrook 7H9 (Becton Dickinson GmbH, Heidelberg/Germany) y se incubaron a  $37^{\circ}\text{C}$  durante una semana. Luego se subcultivaron en Agar Middlebrook 7H11 (Becton Dickinson GmbH, Heidelberg/Germany) durante tres semanas. Una vez obtenido el crecimiento, se procedió a la extracción del ADN utilizando el método cetyltrimethylammonium bromide (CTAB)/NaCl (van Soolingen et al. 1994). La secuenciación se realizó en el instituto BioFrontiers (Universidad de Colorado-Boulder-USA) utilizando la plataforma Illumina HiSeq 2000 con lecturas pareadas de 100 pb y alcanzando una cobertura de entre 78X y 906X. Todos los procedimientos hasta la extracción de ADN fueron realizados en un laboratorio nivel 3 de bioseguridad.

### 1.2.3 Corrección de lecturas y ensamblaje de genomas

La calidad de las lecturas fue visualizada utilizando FastQC (disponible en: <http://www.bioinformatics.babraham.ac.uk/>). Posteriormente fueron corregidas con el programa Trimmomatic versión 0.33 (Bolger, Lohse, and Usadel 2014). Una vez corregidas las lecturas se procedió a ensamblarlas con SOAPdenovo2 versión 2.04 (Luo et al. 2012) utilizando tamaños k-mer de 27 hasta 69 y se escogió el mejor ensamblaje basado en los parámetros N50, número de scaffolds y numero de gaps. Por último, se hizo el cierre de gaps con el programa GapCloser (disponibles en: <http://soap.genomics.org.cn>).

### 1.2.4 Enriquecimiento de base de datos y construcción de perfiles de modelos ocultos de Markov (HMM)

Se realizó una búsqueda bibliográfica de los sistemas reportados para cada tipo de TA. Se estudiaron un total de 225 proteínas, 111 toxinas, 89 antitoxinas, 3 reguladores de sistemas de tres componentes y 22 secuencias de sistemas putativos de TA hallados en H37Rv (Sala et al. 2014). Las secuencias de las proteínas de referencia fueron tomadas de la base de datos de Uniprot (<http://www.uniprot.org/>). En los anexos A, B, C y D se especifica el nombre de cada proteína para cada tipo de TA y el código de acceso.

**Tabla 1:** Genomas utilizados en el estudio.

Aislamientos	Linaje	ID	Origen	
UT1	Euro-América	Secuenciación	Antioquia-Colombia	
UT2		Secuenciación		
UT95		Secuenciación		
UT107		Secuenciación		
UT179		Secuenciación		
UT211		Secuenciación	Cali-Colombia	
UT213		Secuenciación	Antioquia-Colombia	
UT219		Secuenciación		
UT245		Secuenciación		
UT294		Secuenciación		
UT441		Secuenciación		
UT477		Secuenciación	Cali-Colombia	
UT486		Secuenciación	Antioquia-Colombia	
Haarlem		NC_022350.1/NCBI	Haarlem-Holanda	
F11		NC_009565.1/NCBI	Sin datos	
CDC1551		NC_002755.2/NCBI	Kentucky-Tennessee	
Mt256		Etiopia	ERS182920/ENA	Cuerno de África
MTB_BTBH0444			ERS158320/ENA	Etiopia
MTB_BTBH0127	ERS158319/ENA		Etiopia	
NZ3MDR1	CAS	CCSJ01000001.1/NCBI	Nueva Zelanda	
NZ3DS1		JZDR01000001.1/NCBI	Auckland-Nueva Zelanda	
CAS/NITR204		CP005386.1/NCBI	Tamil Nadu-India	
MAL010074	Africa1	KK338837.1/NCBI	Mali	
MAL010118		KK338934.1/NCBI	Mali	
EAI5	EAI	CP006578.1/NCBI	Mumbai-India.	
EAI5/NITR206		CP005387.1/NCBI	Tamil Nadu-India	
UT393		Secuenciación	Antioquia-Colombia	
GM041182	Africa2	NC_015758.1/NCBI	Gambia	
strain 25		CP010334.1/NCBI	Sin datos	
UT307		Secuenciación	Antioquia-Colombia	
Beijing/NITR203	Beijing	CP005082.1/NCBI	Tamil Nadu-India	
CCDC5079		CP001641.1/NCBI	Fujian-China	
CCDC5180		NC_017522.1/NCBI	Fujian-China	
033-46		Secuenciación	Colombia	
033-2293		Secuenciación		
033-2883		Secuenciación		
H37Rv	Referencia	NC_000962.3	Nueva York	

Con cada proteína se realizó una búsqueda con PSI-BLAST(Altschul et al. 1997) sobre la base de datos no redundante del NCBI. Se tomaron las proteínas con porcentaje de

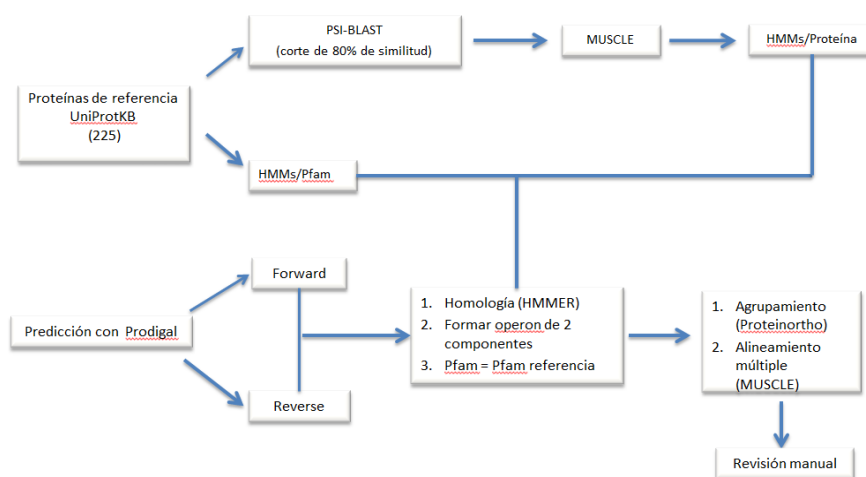
similitud mayor o igual a 80%. Con cada grupo de proteínas se corrió un alineamiento múltiple con MUSCLE (Edgar 2004). Posteriormente se utilizó la herramienta “hmmbuild” del paquete HMMER (Eddy 1998) para construir los perfiles de modelos ocultos de Márkov (PHMM) utilizando los alineamientos múltiples obtenidos.

Adicionalmente, se buscó homología de las proteínas de referencia sobre los HMMs de la base de datos Pfam (Finn et al. 2016). Cada resultado se guardó de forma individual.

Con el fin de confirmar la pérdida o inexistencia de los sistemas TA, se corrió un alineamiento local con tblastn utilizando los genomas como base de datos y las secuencias TA como secuencias objetivo. Los parámetros se corrieron por defecto.

En la Figura 9 se presenta esquemáticamente la metodología utilizada para la predicción y análisis de sistemas TA.

**Figura 9:** Flujo de trabajo para la predicción y análisis de sistemas TA.



### 1.2.5 Predicción de sistemas toxina-antitoxina

Se utilizaron las secuencias de genomas disponibles para completar los 7 linajes reportados de MTB. Se tomaron 3 Africa1, 3 Etiopia, 3 CAS, 2 EAI y las cepas de referencia H37Rv, CDC1551, F11 y Haarlem (Tabla 1). La predicción se realizó sobre 37 genomas en total.

Una vez obtenidos los genomas, se corrió Prodigal (Hyatt et al. 2010) con cada uno de ellos, obteniendo las proteínas predichas para cada genoma. Las proteínas se separaron en dos grupos, uno con las proteínas que fueron codificadas por genes que se encuentran en la hebra forward y otro con proteínas codificadas por genes de la hebra reverse (Figura 9). Lo anterior se hizo con el fin de encontrar proteínas codificadas por operones ya que los componentes de un sistema TA son codificados sobre la misma hebra.



Después de establecer los dos grupos de proteínas, se procedió a buscar homología de cada grupo de secuencias sobre los PHMMs construidos en el paso anterior, utilizando la herramienta “hmmscan” del paquete HMMER y precisando el formato de salida tabulado. Sobre el resultado de cada corrida se extrajeron las secuencias que formaron un operón de dos componentes con otra secuencia de la lista y el tipo de toxina o antitoxina de las proteínas pertenecientes al operón se asumió basado en el menor E-value. Por ejemplo, si dos secuencias A y B están de forma consecutiva en la predicción de Prodigal y además obtuvieron resultado contra los PHMMs, entonces estas proteínas son tomadas como un operón (AB) de un sistema TA. Las proteínas A o B son anotadas como una proteína x basado en el menor E-value obtenido, es decir, si el resultado con menor E-value de la secuencia A fue el PHMM de la toxina vapC1 entonces la secuencia A se interpreta como una toxina vapC1.

Las proteínas obtenidas con este procedimiento se corrieron contra la base de datos de Pfam. Si la proteína en cuestión obtuvo el mismo resultado que la proteína de referencia con la base de datos de Pfam, entonces esa proteína se consideró perteneciente a un sistema TA. Tomando el ejemplo anterior, si la proteína A obtuvo el mismo resultado contra la base de datos de Pfam que la toxina vapC1, entonces la proteína A se consideró una toxina vapC1.

En resumen, una secuencia fue considerada como perteneciente a un sistema TA si cumplió con tres condiciones: 1) Homología con algún PHMM 2) Formar operón con una secuencia que también obtuvo resultado con los PHMMs y 3) Obtener el mismo resultado contra la base de datos de Pfam (Figura 9).

Debido a que las antitoxinas de los sistemas TA 1 y 3 son RNAs no codificantes, las toxinas de estos sistemas no pueden cumplir con el segundo criterio, por lo tanto las secuencias que fueron consideradas toxinas tipo 1 o 3 estuvieron sujetas únicamente a las condiciones uno y tres. Las antitoxinas pertenecientes a estos sistemas fueron buscadas sobre un genoma si su toxina asociada fue identificada por esta metodología.

Para evitar omitir toxinas o antitoxinas reales pero que no forman un operón debido a procesos de delección, se tuvieron en cuenta las secuencias que cumplieron con el primer y segundo criterio, sin embargo no se utilizaron en análisis de mutaciones.

### **1.2.6 Agrupamiento y alineamiento múltiple**

Los sistemas TA predichos de todos los genomas fueron agrupados utilizando Proteinortho (Lechner et al. 2011), un algoritmo diseñado para agrupar secuencias ortólogas. Con cada grupo de ortólogos se realizó un alineamiento múltiple utilizando MUSCLE y se visualizaron los resultados con AliView (Larsson 2014). Los alineamientos fueron revisados manualmente para determinar mutaciones en los aislamientos (Figura 9). Las alteraciones encontradas fueron reportadas basadas en la posición que ocupan en H37Rv.

Si dos grupos de ortólogos presentaron secuencias pertenecientes a la misma proteína (toxina o antitoxina) se evaluaron las secuencias sobre la base de datos de Uniprot, las secuencias que daban como resultado una proteína no caracterizada se tomaron como posibles toxinas o antitoxinas.

### 1.3 Resultados

El análisis sobre los 37 genomas mostró en promedio 151 proteínas asociadas a sistemas TA. Las proteínas encontradas pertenecen exclusivamente a los sistemas II y IV reflejando la inexistencia de los otros tipos en los genomas de MTB estudiados. Los resultados obtenidos con tblastn confirmaron la inexistencia de los sistemas I, III y V en MTB. La proteína IdrB fue la única que obtuvo resultados contra los 37 genomas, sin embargo, el alineamiento cuenta con solo 18 aminoácidos idénticos de los 55 que hacen parte de la secuencia de referencia, las 18 identidades se extienden desde el aminoácido 17 hasta el 49 y presentan 4 deleciones. Por otro lado, las coordenadas del alineamiento con el genoma de H37Rv concuerdan con las del gen *cptI*. Por lo tanto se confirma la ausencia de los sistemas tipo I, III y V en MTB. Siete proteínas fueron establecidas como posibles secuencias pertenecientes a sistemas TA no previamente caracterizadas. Una de estas secuencias guarda homología con la proteína Rv0918 y las otras seis a las antitoxinas vapB11, vapB5, vapB45, vapB13, parD1 y RelB. En la Tabla 2 se presenta la secuencia y el código de la proteína homologa en Uniprot. Las siete proteínas están presentes en los siete linajes.

**Tabla 2:** Posibles proteínas TA no caracterizadas en MTB.

Proteína	Tamaño	Secuencia	ID Uniprot
Homologa vapB11	70	MLKRVEIEVDDDLIQKVIRRYRVKGAREAVNLALRTLL GEADTAEHGHDDEYDEFSDPNVAVVPRRSRDTG	A0A109SUE6
Homologa vapB5	63	MTVTPRPAQADPRSMPEVASRELNRNNTAGLLRRV QAGEDITITANGKPVALLTAGSPHGADG	A0A109T017
Homologa vapB45	238	MRADMSVTSMLDREYVYVAEVDKLIPLPAGTAKRWI NGYERGGKDHPPILRVTPGATPWVTWGEFVETRM AEYRDRRKVPIVRQRAAIEELRARFNLRYP LAHLRPF LSTHERDLTMGGEEIGLPDAEVTIRTG QALLGDARWL ASIATPGRDEVGEAVIVELP VDKAFPEIVINPSRYSGQ PTFVGRRVSPV TIAQMVDGGEEREDLAADYGLSLKQ IQDAIDY TKKYRLARLVAA	P9WLC4
Homologa vapB13	122	MTKKPRNPADYVIGDDVEVSDVDLKQEEVYVDGERL TDERVEQMASESLRLAREANLIPGGKSLSGGSAH SPAVQVVVSKATHAKLKLARSRKMSVSKLLRPVLD EFVQRETGRILPRR	A0A109ST13
Homologa parD	66	MTATIGFRPTEKDEQIINAAMRSGERKSDVIRRALQLL EREVWIKQARTDAERLRDEDVSTEPDAW	V5QPS4
Homologa RelB	85	MTSSLVPISEARAHL SRLVRESADDDVLMNHGRPA AILISAERYESLMEELEDLRDRLSVHEREHVTMPLDK LGAELGVDIGRV	A0A109STC0
Homologa Rv0918	155	MTCPSLVGLRTEAAELSYSDQPDALGVAMRERREQ QNLVRRPPRRNASRRINTDQTSTKYVYITYMPETLTGR LNFRLSPEQEQLRHAALTGQSLSGFVLSAAVDHA HDLLARANRIELSEAAFRRFVAALDEPDEAAPELVRL ARRKSRIPPH	R4MJV5

Un total de 63 proteínas se encontraron conservadas en todos los genomas estudiados (Tabla 3). Cinco de estas proteínas pertenecen a sistemas hipotéticos, las restantes, pertenecen a familias del sistema tipo II a excepción de la antitoxina AbiEi (tipo IV). Este grupo de secuencias está constituido principalmente por antitoxinas (36 secuencias, Figura 12) y la familia más abundante es vapBC con 38 representantes.

**Tabla 3:** Nombre y familia de las proteínas TA completamente conservadas en los genomas de MTB estudiados.

Proteína	Familia	Proteína	Familia
Antitoxina-vapB29	vapBC	Toxina-vapC14	vapBC
Antitoxina-vapB28		Toxina-vapC21	
Antitoxina-vapB4		Toxina-vapC35	
Antitoxina-vapB3		Toxina-vapC50	
Antitoxina-vapB11		Toxina-vapC8	
Antitoxina-vapB9		Toxina-vapC30	
Antitoxina-vapB37		Antitoxina-mazE9	mazEF
Antitoxina-vapB41		Antitoxina-mazE10	
Antitoxina-vapB40		Antitoxina-mazE6	
Antitoxina-vapB20		Toxina-mazF4	
Antitoxina-vapB51		Toxina-mazF9	
Antitoxina-vapB26		Toxina-mazF5	parDE
Antitoxina-vapB44		Antitoxina-parD2	
Antitoxina-vapB12		Antitoxina-parD1	
Antitoxina-vapB1		Toxina-parE2	relBE
Antitoxina-vapB13		Toxina-parE1	
Antitoxina-vapB2		Antitoxina-RelB	
Antitoxina-vapB36		Antitoxina-RelB2	yefM/yoeB
Antitoxina-vapB35		Toxina-RelE2	
Antitoxina-vapB15		Antitoxina-YefM	higBA
Antitoxina-vapB14		Toxina-YoeB	
Antitoxina-vapB45		Antitoxina-HigA2	
Antitoxina-vapB7		Toxina-HigB1	pemK/pemI
Antitoxina-vapB6		Toxina-HigB3	
Antitoxina-vapB30		Toxina-PemK	AbiEii/AbiEi
Antitoxina-vapB21		Antitoxina-AbiEi	
Toxina-vapC33		Rv0910	Desconocida
Toxina-vapC41		Rv3189	
Toxina-vapC39		Rv3188	
Toxina-vapC17		ArsR	
Toxina-vapC42		Rv1545	
Toxina-vapC22			

Por otro lado, los sistemas Rv3188-Rv3189, yefM/yoeB, RelBE2, parDE1, parDE2, mazEF9, vapBC14, vapBC21, vapC30, vapBC35 y vapBC41 son los sistemas completos (operón) conservados.

El alineamiento múltiple reveló mutaciones que estuvieron presentes en todos los genomas de un linaje. Dichas mutaciones se encontraron en 38 proteínas, de las cuales 20 son toxinas, 11 son antitoxinas y 7 son proteínas no caracterizadas. En los 7 grupos, los linajes Africa1, Africa2 y Etiopia son los que contienen mayor número de mutaciones (Tabla 4).

En este grupo de proteínas mutadas se encontró la delección completa del sistema Rv2653c-Rv2654c (sistema no caracterizado) en los linajes Africa2 y Etiopia. Esta misma delección se presencié en las cepas UT179 y UT393. Otras delecciones se encontraron en las proteínas vapC3, vapB24 y Rv1045 con pérdida de 35, 24 y 121 aminoácidos. Por otra parte, se registraron inserciones en Rv0909 de 6 aminoácidos en el extremo N-terminal y en Rv0836c de 23 aminoácidos en el extremo C-terminal.

También se observaron cambios puntuales (cambio de un solo aminoácido) en este grupo de 38 proteínas (Tabla 4 y 5). Se registraron 45 cambios de un solo aminoácido, de los cuales 17 cambian un aminoácido polar por uno no polar, 8 son de la forma inversa a la anterior, tres son cambios de aminoácidos polar positivos por uno polar no cargado, dos son cambios de polar no cargado por polar negativo, 14 son entre aminoácidos no polares y solo hay un caso de cambio entre aminoácidos polares no cargados. No se registraron cambios entre aminoácidos positivos por negativos y viceversa. Los linajes Africa2 y Etiopia tienen la mayor cantidad de mutaciones encontradas con 19 y 17 respectivamente, mientras que los linajes CAS y Beijing cuentan con siete mutaciones cada uno (Figura 11). La mayor cantidad de intercambios entre aminoácidos polares y no polares la presentan los linajes Africa1 y Africa2 con 11 y 12 respectivamente.

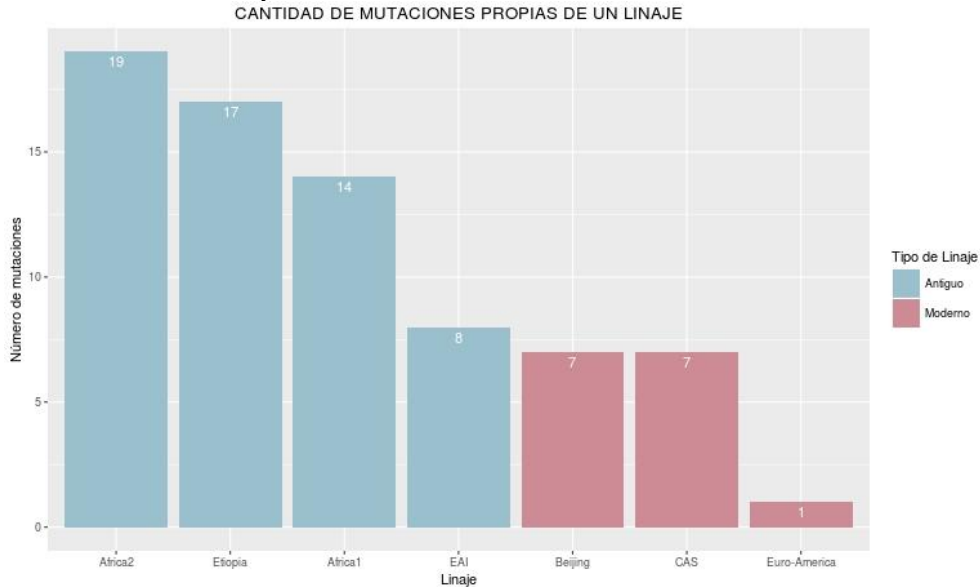
Cinco de las antitoxinas muestran cambios entre aminoácidos polares y no polares mientras que 14 toxinas tienen este tipo de alteraciones. Las proteínas vapC51 y Rv0918 son las únicas proteínas que presentan dos cambios puntuales diferentes en dos linajes y son junto a Rv0909 y la toxina vapC25 las proteínas que mostraron más de una mutación puntual.

En la Tabla 6, se lista un grupo de mutaciones puntuales que son encontradas solo en un linaje pero no hacen parte de todos los genomas que lo representan, es decir, mutaciones únicas de un solo genoma y de un solo linaje. Son 48 mutaciones encontradas en 21 toxinas, 7 antitoxinas, 7 proteínas no caracterizadas y un regulador de un sistema de tres componentes (SecB-like chaperone Rv1957). En la Figura 12 se presenta el número global de toxinas y antitoxinas mutadas y conservadas en los linajes.

Las proteínas vapC25, vapC43, HigB2, AbiEii, mazE4 y Rv0918 son las secuencias que se repiten en este grupo. Los principales cambios se presentaron entre aminoácidos polares por no polares (13 mutaciones), no polares por no polares (11 mutaciones) y no

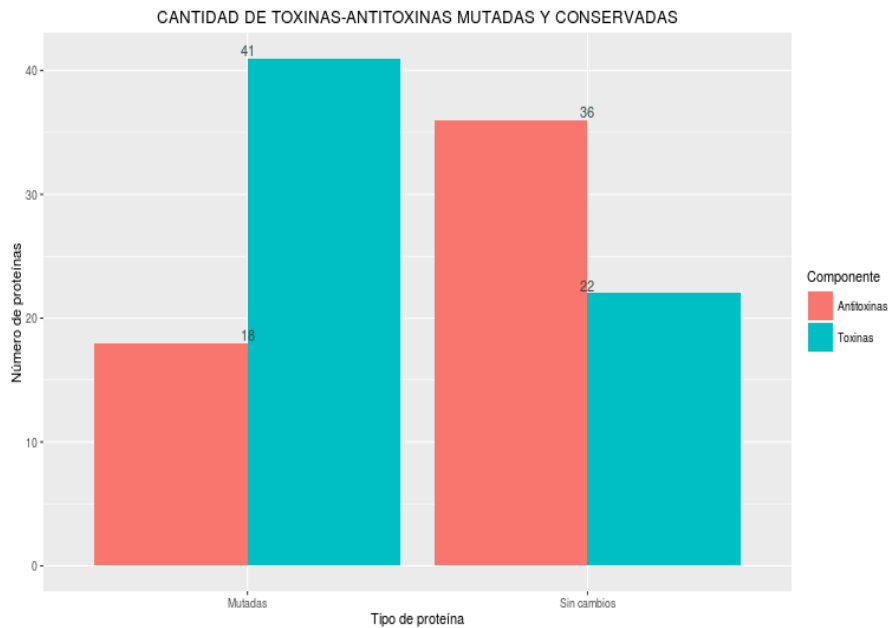
polares por polares (10 mutaciones). 13 de las mutaciones entre aminoácidos polares y apolares están en toxinas, mientras que 4 en antitoxinas y las restantes en proteínas no caracterizadas.

**Figura 10:** Número de mutaciones identificadas de acuerdo a los genomas estudiados de cada linaje.



Las toxinas también tienen mayor cantidad de mutaciones entre aminoácidos no polares (8 mutaciones) en comparación a las antitoxinas (1 mutación). A diferencia de las mutaciones de las Tablas 4 y 5, en este grupo de secuencias sí se presentaron cambios puntuales entre aminoácidos con carga positiva y negativa (4 mutaciones).

**Figura 11:** Comparación de variabilidad en secuencias de toxinas y antitoxinas.



La cepa CAS/NITR204 (CP005386.1), con 10 mutaciones, es la cepa con mayor número de alteraciones. Además de mostrar estas mutaciones, 15 proteínas en CAS/NITR204 cambian su longitud con respecto a la referencia. Las proteínas son: vapC26, vapC12, HigA3, vapC9, Rv0837c, mazF7, vapB17, vapC32, mazE2, vapC5, vapC31, vapBC29, Rv2826c, vapC15 y vapB46. Un análisis sobre la secuencia nucleotídica del genoma, reveló inserciones y deleciones de un número de nucleótidos no múltiplo de 3 en todos estos genes, llevando a cambios en la secuencia de la proteína.

**Tabla 4:** Mutaciones asociadas a cada linaje. Inserciones (INS) y Deleciones (DEL) tienen entre paréntesis el número de aminoácidos involucrados. En las mutaciones puntuales se muestra el cambio de aminoácido y la posición donde ocurre. Se indican las proteínas sin mutaciones (SM) en cada linaje.

Proteína	Linaje						
	Africa2	EAI	Africa 1	Etiopia	Euro América	CAS	Beijing
Toxina-vapC3	DEL (35) a partir del aminoácido 93	SM	SM	SM	SM	SM	SM
Toxina-vapC10	D-G(103)	D-G(103)	D-G(103)	D-G(103)	SM	SM	SM
Toxina-RelE	SM	SM	SM	L-I(99)	SM	SM	SM
Antitoxina-vapB34	SM	T-A(21)	SM	SM	SM	SM	SM
GCN5-like	SM	SM	L-P(151)	SM	SM	SM	SM
Rv0836c	INS (23) en extremo C-terminal	INS (23) en extremo C-terminal	INS (23) en extremo C-terminal	INS (23) en extremo C-terminal	SM	INS (23) en extremo C-terminal	INS (23) en extremo C-terminal
Rv0909	INS (6) en extremo N-terminal, I-N(3), L-P(4), D-G(5), K-Q(6), V-G(7), N-D(9), H-Q(45)	H-Q(45)	H-Q(45)	H-Q(45)	SM	SM	SM
Toxina-mazF10	SM	SM	SM	SM	SM	A-T(30)	SM
Toxina-vapC24	A-S(31)	SM	A-S(31)	SM	SM	SM	SM
Antitoxina-vapB24	SM	SM	SM	DEL (24) en extremo N-terminal	SM	SM	SM
Toxina-vapC40	C-R(77)	C-R(77)	C-R(77)	C-R(77)	SM	C-R(77)	C-R(77)
Antitoxina-vapB19	SM	SM	D-G(60)	SM	SM	SM	SM
Toxina-mazF3	T-I(65)	T-I(65)	T-I(65)	T-I(65)	SM	T-I(65)	T-I(65)
Toxina-vapC38	SM	SM	SM	SM	SM	SM	A-V(48)
Antitoxina-vapB5	SM	SM	SM	N-D(69)	SM	SM	SM
Toxina-vapC23	SM	SM	SM	L-V(102)	SM	SM	SM
Antitoxina-vapB22	S-A(71)	SM	SM	SM	SM	SM	SM
Toxina-vapC13	SM	SM	I-T(67)	SM	SM	SM	SM
Toxina-vapC6	SM	SM	SM	SM	SM	W-L(65)	SM

**Tabla 5:** Continuación de la tabla 4.

Proteína	Linaje						
	Africa2	EAI	Africa1	Etiopia	Euro América	CAS	Beijing
Antitoxina-vapB31	V-G(50)	SM	SM	SM	SM	SM	SM
Rv1045	SM	SM	SM	SM	SM	SM	DEL de 121 aminoácidos iniciales
Toxina-vapC9	SM	R-W(58)	SM	SM	SM	SM	SM
Antitoxina-mazE4	R-W(90)	SM	SM	SM	SM	SM	SM
Antitoxina-vapB39	A-V(15)	SM	SM	SM	SM	SM	SM
Rv2654c	DEL (completa)	SM	SM	DEL (completa)	SM	SM	SM
Rv2653c	DEL (completa)	SM	H-Q(98)	DEL (completa)	SM	SM	SM
Antitoxina-vapB47	SM	SM	SM	SM	SM	SM	R-C(84)
Toxina-vapC37	SM	SM	SM	SM	SM	SM	T-A(16)
Toxina-vapC32	SM	SM	R-C(14)	SM	SM	SM	SM
Antitoxina-vapB50	SM	SM	SM	L-F(8)	SM	SM	SM
Antitoxina-vapB33	SM	SM	SM	L-C(8)	SM	SM	SM
Toxina-vapC31	SM	M-V(90)	SM	SM	SM	SM	SM
Toxina-vapC25	SM	SM	SM	S-N(83), V-M(90), G-A(112), E-G(117), R-W(131)	SM	SM	SM
Toxina-vapC51	SM	SM	S-P(46)	SM	SM	R-L(86)	SM
Toxina-vapC45	Q-P(2)	SM	SM	SM	SM	SM	SM
Toxina-vapC48	G-R(34)	SM	G-R(34)	SM	SM	SM	SM
Rv0918	S-G(39)	SM	S-G(39)	S-G(39), P-R(61)	SM	S-G(39)	S-G(39)
Toxina-vapC47	S-L(46)	S-L(46)	S-L(46)	S-L(46)	S-L(46)	S-L(46)	S-L(46)

La cepa Strain 25 (CP010334.1) muestra mutaciones similares, los genes que codifican para vapC34, Rv0060, vapC45, vapB39, Rv2827c, vapC37, Rv1990c, vapC25 y vapC49 sufren un cambio abrupto por delección o inserción de un número de nucleótidos no múltiplo de 3. Otras proteínas que sufren modificaciones en sus secuencias por este tipo de mutaciones son Rv0837c y mazE5 en EAI5 (CP006578.1), vapC45 y vapC7 en EAI5/NITR206 (CP005387.1), vapB31, vapB13, vapC25 y vapB49 en CDC5079 (CP001641.1), vapB43 y mazE3 en CDC1551 y vapC4 en UT294.

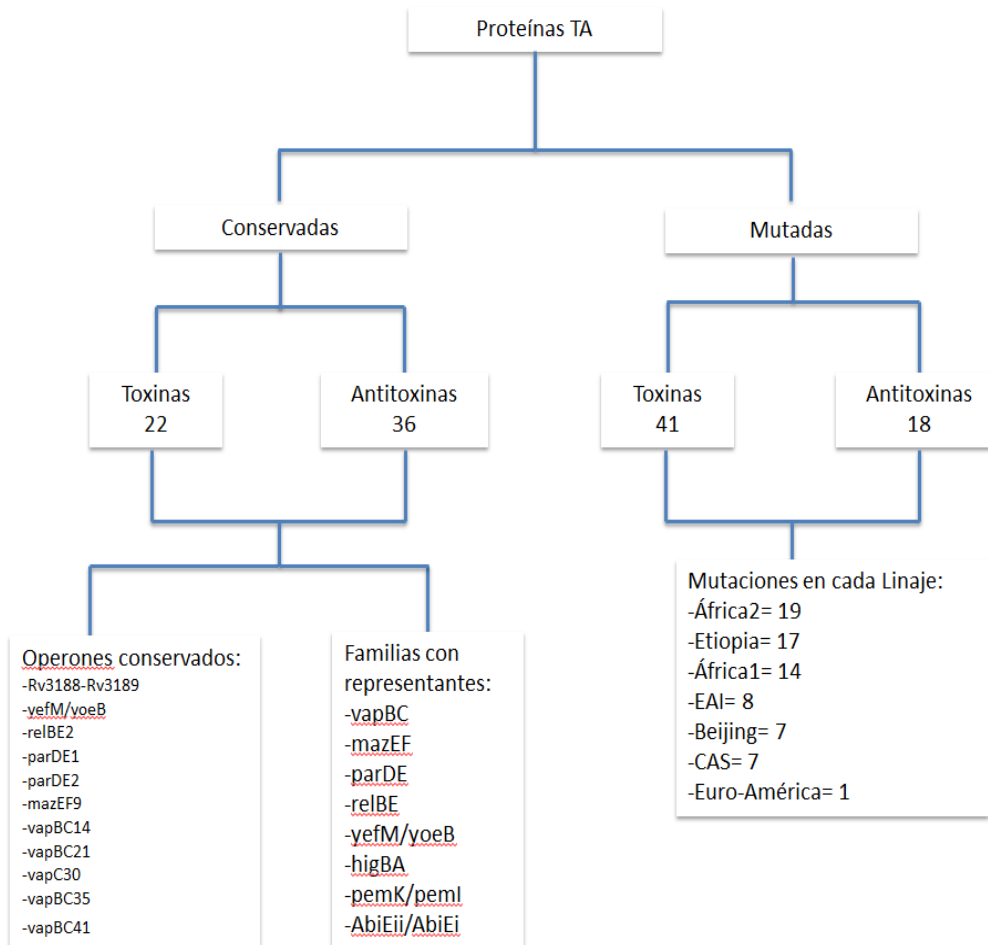


**Tabla 6:** Mutaciones encontradas en un solo aislamiento y un solo linaje.

Linaje	Cepa	Proteína	Mutación
Africa2	GM041182 (NC_015758.1)	Toxina-vapC2	R-H(93)
	Strain 25 (CP010334.1)	Antitoxina-vapB32	P-L(58)
	UT307	Toxina-mazF2	R-H(55)
	UT307 y GM041182 (NC_015758.1)	Toxina-vapC34	K-T(67)
	UT307	Rv0837c	S-N(229)
	Strain 25 (CP010334.1)	Toxina-HigB2	E-K(131)
EAI	UT393	Toxina-vapC11	R-W(124)
	EAI5/NITR206 (CP005387.1) y EAI5 (CP006578.1)	Toxina-vapC20	V-L(87)
	EAI5/NITR206 (CP005387.1)	Toxina-vapC43	N-T(53)
	EAI5/NITR206 (CP005387.1)	Antitoxina-vapB8	R-G(41)
	EAI5/NITR206 (CP005387.1)	Rv0918	Q-P(35)
	UT393	Toxina-mazF7	R-P(69)
	EAI5 (CP006578.1)	Toxina-vapC25	R-Q(49)
Etiopia	MTB_BTBH0127 (ERS158319/ENA)	Toxina-vapC19	L-I(13)
	Mt256 (ERS182920)	Rv0918	R-W(29)
	Mt256 (ERS182920)	Rv2035	E-A(51)
Euro/América	UT294	Antitoxina-vapB10	K-R(2)
	UT107	Toxina-vapC44	T-I(129)
	UT1-UT213-UT245	Rv0059	A-P(83)
	UT211	Toxina-vapC18	E-K (69)
	UT211	Toxina-mazF6	G-S(59)
	UT179	Rv1989c	C-R(25)
	UT219-UT1-UT441-UT245-UT213-UT486-Haarlem y CDC1551	Toxina-vapC36	L-H(126)
	UT441-UT1-UT245-UT213-UT486-UT219-Haarlem y CDC1551	SecB-like chaperone Rv1957	D-N(84)
	UT211	Antitoxina-vapB48	R-C(49)
	UT219-UT441	Rv0918	T-A(53)
	UT107	Antitoxina-mazE4	R-G(22)
	CDC1551	Antitoxina-mazE4	R-H(15)
	UT486	Toxina-vapC37	A-T(109)
	UT211	Antitoxina-HigA1	A-S(149)
	UT107-UT179-UT294-UT2-UT477-UT95	Toxina-HigB2	V-A(118)
	UT294	Toxina-HigB2	P-L(57)
	UT294	Toxina-vapC7	G-D(86)
	UT219 y UT441	Antitoxina-mazE2	D-N(45)
	UT486	Toxina-vapC51	A-T(62)
	CAS	NZ3DS1 (JZDR01000001.1)	Toxina-vapC29
CAS/NITR204 (CP005386.1)		Rv1546	D-A(72)
CAS/NITR204 (CP005386.1)		Toxina-AbiEii	V-I(55), V-C(105), I-H(106)
CAS/NITR204 (CP005386.1)		Toxina-vapC43	V-A(99)
CAS/NITR204 (CP005386.1)		Rv2827c	A-V(71)
NZ3DS1 (JZDR01000001.1) NZ3MDR1(CCSJ01000001.1)		Toxina-vapC46	A-T(42)
CAS/NITR204 (CP005386.1)		Toxina-vapC25	R-P(139)
CAS/NITR204 (CP005386.1)		Toxina-vapC49	H-D(91), L-I(92), H-D(96)
Beijing	Beijing/NITR203 (CP005082.1)	Toxina-AbiEii	P-S(222)

En la figura 10 se representa los principales datos obtenidos durante el análisis de variabilidad de sistemas TA en MTB. Se representa el número de toxinas y antitoxinas conservadas y mutadas, los operones completamente conservados, las familias de los sistemas TA sin variaciones y el número de mutaciones por linaje.

**Figura 12:** Principales resultados obtenidos durante el análisis de sistemas TA.



## 1.4 Discusión

### 1.4.1 Tipos de toxina-antitoxina y nuevas proteínas

Los sistemas TA regulan el metabolismo de una bacteria, sus toxinas, bajo situaciones de estrés, dirigen su acción sobre distintos procesos metabólicos, siendo entonces parte de la maquinaria de regulación epigenética (Wen et al. 2014; Y Yamaguchi et al. 2011). Son sistemas ampliamente difundidos en arqueas y bacterias (Y Yamaguchi et al. 2011) y se dividen en 5 grupos, dependiendo de la forma en la que la antitoxina inhibe a la toxina (Wen et al. 2014). Nuestros datos muestran la inexistencia de sistemas TA tipo I, III y V en todos los linajes de MTB estudiados, resultados que concuerdan con estudios previos desarrollados en la búsqueda de sistemas tipo I y III (Blower et al. 2012; Fozo et al. 2010a). Los sistemas tipo I se han encontrado en Firmicutes y Gammaproteobacterias y se cree que se expanden por duplicación en líneas evolutivas específicas y no por transferencia horizontal (Fozo et al. 2010b), por otra parte, los sistemas tipo III son encontrados principalmente en Fusobacterias y Firmicutes. A diferencia de los sistemas tipo I, se ha sugerido que los sistemas tipo III presentan transferencia horizontal entre bacterias no relacionadas y se ha descrito homología con las toxinas MazF y RelE (tipo II) (Blower et al. 2012). A pesar de esto, no se encontró homología en ningún linaje de MTB. El sistema GhoT/GhoS fue el primer sistema tipo V reportado (Wang et al. 2012), es encontrado en *E. coli* y puede no estar presente en líneas evolutivas de MTB. Sin embargo, solo tener un representante de este tipo de sistema TA, limita la búsqueda en las bases de datos y por tanto el diseño de PHMM. Se necesitaría una mayor cantidad de sistemas tipo V para hacer una búsqueda adecuada en MTB.

Los sistemas tipo II (la antitoxina es una proteína que interactúa directamente con la toxina) son encontrados en gran variedad de bacterias (Leplae et al. 2011b) y son ampliamente distribuidos en MTB en comparación a especies cercanas como *Mycobacterium marinum* (Ramage et al. 2009). La expansión de los sistemas TA parece haber ocurrido después de haber divergido MTB y *M. marinum*, adquiriéndose por transferencia horizontal, sumado a eventos de duplicación, que también pudieron tener un papel en el incremento de los sistemas (Ramage et al. 2009). Los sistemas TA se expandieron y mantuvieron en MTB, especialmente los pertenecientes a la familia vapBC que cuenta con 50 de los 79 sistemas reportados en H37Rv (Sala et al. 2014). Nuestros resultados, revelaron 7 posibles secuencias adicionales asociadas a sistemas TA (Tabla 2), de las cuales 4 son homólogas a antitoxinas vapB, mostrando relación con la expansión y conservación de sistemas TA, en especial de la familia vapBC.

De las 7 posibles proteínas TA, 6 son antitoxinas, resultados que sugieren que en MTB se favorece el surgimiento de nuevas antitoxinas. Este fenómeno podría obedecer a que las antitoxinas son más propensas a la degradación que las toxinas, lo que implica su constante recambio para evitar dejar libre las toxinas (Chan, Espinosa, and Yeo 2016) y a

la capacidad que tiene una antitoxina de interactuar con diferentes toxinas (Leplae et al. 2011b). La menor estabilidad de las antitoxinas favorecería el surgimiento de nuevas antitoxinas que puedan inhibir distintas toxinas y que aseguren una mayor oferta de antitoxinas para su constante recambio. Las nuevas antitoxinas podrían actuar de forma colaborativa para impedir la liberación de toxinas. Este tipo de redes de interacción han sido previamente reportadas (Zhu et al. 2010).

### 1.4.2 Core de sistemas toxina-antitoxina

El core genómico son genes que son conservados en todos los aislamientos de una especie bacteriana (Medini et al. 2005). En el pasado, se han realizado esfuerzos por determinar el core de genes en el MTBC, por ejemplo, sobre 96 genomas del MTBC se estableció un core con secuencias que formaban grupos de homólogos al tener 70% de similitud global y estar presentes en todos los genomas (Periwal et al. 2015), por otro lado Marmiesse et al. 2004 definieron un core en 5 especies de Micobacterias (incluida MTB) utilizando secuencias proteicas que tuvieran un porcentaje de identidad mayor al 40% en al menos 70% de la longitud total de la proteína. En este estudio se obtuvo un grupo de 63 proteínas de sistemas TA que comparten 100% de similitud global y que están presentes en todos los aislamientos estudiados (Tabla 3). Los genes que codifican las 63 proteínas podrían hacer parte de un core inalterado de MTB, no obstante, se necesita un número mayor de genomas para comprobarlo.

Los genes esenciales en una bacteria tienden a organizarse en forma de operón y tienen un alto grado de conservación (Grazziotin, Vidal, and Venancio 2015). A pesar de esto, los sistemas TA no son esenciales en bacterias y arqueas (Y Yamaguchi et al. 2011). La prevalencia de los 63 genes en MTB puede tener diferentes causas, por un lado, se ha propuesto que los sistemas TA se comportan como secuencias egoístas, manteniéndose en el cromosoma con un efecto similar al que ocurre con la adición de plásmidos, en donde las células que no adquieren una copia del plásmido, mueren por no poder producir la cantidad suficiente de antitoxina (Pandey and Gerdes 2005). También se han asociado en la protección contra ADN foráneo como el de plásmidos y fagos (Hazan and Engelberg-Kulka 2004; Van Melderren 2010) y parece, pueden prevenir la pérdida de ADN (Van Melderren 2010). Sin embargo, una de los principales roles biológicos que se les atribuye, es el de ser un mecanismo de la bacteria para afrontar situaciones de estrés (Van Melderren 2010; Ramage et al. 2009). Por ejemplo, la familia vapBC, la más representativa en el grupo de 63 proteínas, está involucrada en distintas situaciones de estrés. Los sistemas vapBC7, 15 y 25 son inducidos en condiciones de hipoxia e interacción con macrófagos, mientras que los sistemas vapBC3, 31 y 49 están sobre expresados en bacterias resistentes a medicamentos (Sala et al. 2014). Las toxinas mazF también han sido involucradas en la respuesta de MTB a condiciones de estrés: mutantes de las toxinas mazF3, 6 y 9 muestran reducción en la formación de células persistentes bajo estrés oxidativo, nutricional y presión de medicamentos, y se observó una menor carga bacteriana de estos mutantes en infección experimental en curies (Tiwari et al. 2015).

Los sistemas relBE, con un operón completamente conservado en los linajes (Tabla 3), son sobre expresados en condiciones de estrés oxidativo, y a diferencia de los sistemas vapBC y higBA los cuales son sobre expresado bajo hipoxia, los sistemas relBE son sub expresados en esta condición (Korch et al. 2015). Además, las toxinas relE1 y relE2 son sobre expresadas en tratamiento con antibióticos y en tejidos de pulmón de ratones infectados (Sala et al. 2014).

Por su parte, los sistemas parDE, completamente conservados en los 7 linajes, son expresados bajo estrés oxidativo (Gupta et al. 2016) y podrían reducir la pérdida de ADN cromosómico (Fiebig et al. 2010). En *E. coli*, las toxinas parE de MTB, logran inhibir su crecimiento (Sala et al. 2014).

La evidencia mencionada anteriormente de los sistemas vapBC, mazEF, parDE y relBE como mecanismo de la bacteria para confrontar condiciones de estrés, sugiere que los sistemas TA son activados bajo situaciones comunes, como estrés oxidativo y presión de medicamentos. Esta característica y su amplia distribución en MTB podrían asegurar una respuesta efectiva ante distintos estímulos, de allí la importancia de su conservación. Sin embargo, hay sistemas que tienen expresión contraria en una misma situación de estrés, por ejemplo, en condiciones de hipoxia, los sistemas vapBC son sobre expresados, mientras que los sistemas relBE son sub expresados, lo que indica que aunque los sistemas TA parecen tener la misma función (regular el metabolismo en situaciones de estrés), puede que estén expresados bajo distintos controles regulatorios, de tal forma que un subconjunto es activado en una condición particular de estrés (Ramage et al. 2009), de ahí, la necesidad de tener diferentes familias conservadas.

Aparte de ser activados bajo distintas situaciones de estrés, los sistemas TA tienen blancos distintos. Por ejemplo, las toxinas parE inhiben la acción de la girasa mientras que las toxinas mazF y vapC son ribonucleasas que impiden la traducción (Sala et al. 2014). Las diferencias también son observadas dentro de familias, es el caso de las toxinas mazF6 y vapC20, las cuales actúan sobre el RNAr 23S y no sobre el RNAm (Sala et al. 2014). La expresión de un grupo de sistemas TA bajo una misma condición pero con distintos blancos, haría más eficiente la regulación del metabolismo bacteriano. Esto también daría respuesta a la importancia de conservar distintos sistemas y familias TA en MTB.

El vínculo de los sistemas TA con la capacidad de MTB de entrar en latencia podría dar explicación a la prevalencia de algunos sistemas en los 7 linajes. Tal vínculo se observa sobre toxinas vapBC, las cuales alteran la morfología de *M. smegmatis* volviéndola ovoide. Las células con esta morfología presentan baja tasa de transcripción y disminuyen su consumo de oxígeno. Además, se muestra sobre expresión de toxinas vapC en la transición a células dormantes (Demidenok, Kaprelyants, and Goncharenko 2014). Sistema parDE suman evidencia al reducir la capacidad de *M. smegmatis* de formar colonias bajo estrés oxidativo, una condición mantenida en el interior del macrófago. A pesar de esto, las células siguen siendo viables (Gupta et al. 2016). Ya que las toxinas de los sistemas relBE y maszEF son activadas bajo situaciones similares a

vapBC y parDE, es posible que los sistemas conservados en los linajes sean necesarios para que la bacteria entre en latencia.

### 1.4.3 Mutaciones de proteínas toxina-antitoxina

En este estudio se encontraron mutaciones que por linaje, están en todos los genomas que los representan. Este grupo de mutaciones, junto con las halladas en un único aislamiento (Tabla 6), muestran un mayor número de toxinas mutadas con relación a las antitoxinas. Estos resultados se asocian con los grupos de secuencias nuevas y secuencias completamente conservadas, en los cuales la mayor cantidad de proteínas son antitoxinas. En conjunto, estos resultados sugieren que, en MTB, no solo se favorece el surgimiento de nuevas antitoxinas, sino que también se tiende a conservar más las antitoxinas que las toxinas (Figura 12). Cabe también señalar, que la mayor cantidad de cambios de aminoácidos polares por apolares o viceversa, se dio en las toxinas, reflejando un carácter menos restrictivo al cambio de grupos funcionales con características distintas. Esto sugiere una relación con la protección que da la antitoxina a la célula para no ser atacada por su propia toxina y a su papel como factores de regulación de su propio operón (Chan et al. 2016).

Las mutaciones encontradas pueden ser utilizadas como marcadores para la determinación de linaje. Un análisis de polimorfismos sobre 5 familias tipo II, logró determinar los genes y las mutaciones necesarias para establecer el genotipo y subtipos de cepas MTB (Zaychikova et al. 2015). Los autores proponen un grupo de 13 genes que pueden ser utilizados para la determinación de los linajes (Zaychikova et al. 2015). En este grupo, las mutaciones reportadas sobre los genes mazF3, vapC47, vapC6, vapC10 y vapC37 concuerdan con nuestras mutaciones. Adicionalmente, nosotros aportamos mutaciones que pueden ser utilizadas en la determinación del linaje Etiopia, Africa1 y 2, como fueron las mutaciones en GCN5-like, vapB22 y relE para determinar los linajes África1, África2 y Etiopia respectivamente (Tabla 4). Estos datos aportan información para una completa genotipificación de los linajes de MTB utilizando como base estas mutaciones.

Finalmente, los autores concluyen que mutaciones en genes TA pueden alterar la actividad de la proteína y que estas alteraciones podrían guardar relación con la eficiencia de adaptación de un genotipo a las condiciones del hospedero. Esta conclusión es interesante puesto que la mayor cantidad de nuestras mutaciones fueron encontradas en los linajes antiguos, con menos distribución geográfica con respecto a los modernos (Coll et al. 2014) (Figura 11).

Los linajes modernos parecen tener más éxito en su propagación que linajes antiguos (Coscolla and Gagneux 2014). Además, muestran diferencia en el desarrollo de la enfermedad; Los linajes 2 y 4 son más propensos a desarrollar enfermedad activa con respecto al linaje 6 y pacientes infectados con bacterias del linaje 4 presentan mayor conteo de neutrófilos y un índice corporal más bajo que pacientes infectados con bacterias del linaje 1 (Coscolla and Gagneux 2014). En otros estudios, se ha evidenciado

que el linaje Euro América muestra diferencias en consolidación pulmonar con respecto a EAI y Beijing, indicando respuestas inflamatorias distintas entre linaje (Thwaites et al. 2008). Caws et al. 2008 encontraron que pacientes infectados con cepas Euro América tienen baja mortalidad en TB meníngea comparado con los linajes EAI y Beijing.

La evidencia presentada anteriormente liga a los linajes comportamientos distintos en el desarrollo de la enfermedad e interacción con el hospedero. Los sistemas TA, al ser activados bajo situaciones de estrés como la presión de medicamento y el sistema inmune, pueden tener alguna relación con el desarrollo de la enfermedad, por lo tanto, las mutaciones encontradas en los sistemas TA de los diferentes linajes (Tabla 4, 5) podrían estar relacionadas con las características patobiológicas asociadas a cada linaje.

En Africa1 y Africa2 (linajes antiguos) encontramos mayor cantidad de variaciones entre aminoácidos polares y no polares que las encontradas en linajes modernos (Tabla 4 y 5). Dichas variaciones también implican un mayor número de cambios de aminoácidos inactivos catalíticamente como Glicina, Alanina, Leucina y Prolina, por aminoácidos como Arginina, Aspartato y Serina, involucrados en el intercambio de protones, activación de especies reactivas y estabilización de moléculas (Holliday, Mitchell, and Thornton 2009). Estos cambios indican que, en el tiempo, MTB ha sufrido cambios en las secuencias de toxinas y antitoxinas que podrían influir sobre sitios activos de las proteínas o en la interacción toxina-antitoxina, con un posible impacto en la función de los sistemas TA en aislamientos de linajes antiguos y modernos. Además, cambios entre aminoácidos de distinto tamaño y polaridad, como el de Glicina por una Arginina, podrían afectar la estructura de la proteína ya que los aminoácidos apolares tienden a estar en el interior de la molécula (Alberts 2013).

Los sistemas TA muestran variabilidad no solo a nivel de linajes sino también de aislamiento (Tabla 6), teniendo mutaciones vinculadas únicamente a un aislamiento. Algunos genes en estos aislamientos tienen inserciones y deleciones que cambian el marco de lectura, cambiando la estructura primaria y probablemente la secundaria y terciaria de la proteína. Si esto es así, los sitios activos de la proteína pueden cambiar y hacer a las proteínas inactivas para su función original. Este tipo de cambios se observaron en las cepas CAS/NITR204, Strain 25, EAI5 (CP006578.1), EAI5/NITR206 (CP005387.1), CCDC5079, CDC1551 y UT294. Sin embargo, el aislamiento CAS/NITR204 con 15 proteínas alteradas, es el aislamiento que tiene el mayor número de mutaciones. CAS/NITR204 es un aislamiento clínico de la India (Narayanan and Deshpande 2013), indicando que sus 15 proteínas con cambio en la estructura primaria, no estarían afectando drásticamente la capacidad de este aislamiento para desarrollar la enfermedad. Lo mismo podríamos concluir de los aislamientos clínicos EAI5/NITR206, CDC1551 y UT294. No obstante, la probable pérdida de la función de distintos genes TA, sugiere que estos eventos pudieran no estar ligados a un carácter prescindible en los procesos de infección y desarrollo de la enfermedad, sino más bien a la redundancia de funciones entre los diferentes sistemas TA. Al igual que en el surgimiento de nuevos genes, donde la duplicación de un gen permite a una copia mantener su función ancestral (usualmente importante) y a la otra adquirir mutaciones en el tiempo

(Kaessmann 2010), los sistemas TA en MTB pueden tener la misma libertad para mutar debido a que distintos sistemas responden a la misma situación de estrés, sin haber duplicación exacta de los genes, hay redundancia en la función.

La libertad de los sistemas TA de adquirir mutaciones explicaría la variabilidad que se percibe por aislamiento. Además, la redundancia en funciones podría ser clave para el surgimiento de nuevas toxinas, ya que se han identificado estructuras similares entre toxinas que no comparten similitud significativa en sus secuencias (Leplae et al. 2011b).



## 2. Conclusiones y perspectivas

### 2.1 Conclusiones

- Se encontraron mutaciones relacionadas con cada linaje que pueden ser utilizadas para la genotipificación de *Mycobacterium tuberculosis*.
- Los cambios que encontramos en la estructura primaria de proteínas de sistemas toxina-antitoxina en aislamientos clínicos de *Mycobacterium tuberculosis*, son evidencia de que la redundancia en la función de los sistemas toxina-antitoxina permitiría a las proteínas adquirir mutaciones sin afectar drásticamente el fenotipo de la bacteria y permitiría el surgimiento de nuevas proteínas asociadas a sistemas toxina-antitoxina.
- Las posibles nuevas antitoxinas encontradas en este estudio, indican que la inestabilidad de las antitoxinas y su capacidad de interactuar con distintas toxinas, favorece el surgimiento de nuevas antitoxinas que actúen en forma sinérgica con las demás.
- Se hallaron sistemas completamente conservados en los linajes de *Mycobacterium tuberculosis* que pueden ser parte del core genómico de la especie y además, asegurarían a la bacteria una respuesta efectiva bajo condiciones de estrés variada, ya que son activados bajo distintas situaciones de estrés y actúan sobre blancos diferentes.

### 2.2 Perspectivas

Las antitoxinas hipotéticas detectadas necesitan ser validadas experimentalmente, primero evaluando su expresión bajo diferentes condiciones de estrés y posteriormente analizando su interacción con las toxinas de la familia a la cual pertenece su antitoxina homóloga.

En los casos en que se disponga de un cristal, se podría evaluar el efecto de las mutaciones asociadas a cada linaje sobre la estructura de la proteína y se podrían adelantar análisis de dinámica molecular con el fin de evaluar el efecto sobre la interacción toxina-antitoxina, toxina-blanco y antitoxina-ADN.

Por último es necesario evaluar las mutaciones sobre un conjunto adicional de genomas de cada linaje para verificar su utilidad como marcadores en la genotipificación de *Mycobacterium tuberculosis*.

## A. Toxinas de referencia

Tipo TA	Gen	Proteina	ID
1	<i>bsrG</i>	Phage toxin type I toxin-antitoxin system	L8EAY0
	<i>fmlA</i>	Stable plasmid inheritance protein	P62670
	<i>hok</i>	Protein hok	P11895
	<i>ibsA</i>	Small toxic protein IbsA	C1P607
	<i>ibsB</i>	Small toxic protein IbsB	C1P608
	<i>ibsC</i>	Small toxic protein IbsC	C1P615
	<i>ibsD</i>	Small toxic protein IbsD	C1P616
	<i>ibsE</i>	Small toxic protein IbsE	C1P617
	<i>ldrA</i>	Small toxic polypeptide LdrA/LdrC	Q6BF86
	<i>ldrB</i>	Small toxic polypeptide LdrB	Q6BF87
	<i>ldrC</i>	Small toxic polypeptide LdrA/LdrC	T6B5I3
	<i>ldrD</i>	Small toxic polypeptide LdrD	Q6BF25
	<i>ldrE</i>	Small toxic polypeptide	D3H0Z5
	<i>PndA</i>	Protein PndA	P11902
	<i>ralR</i>	Endodeoxyribonuclease RalR	P33229
	<i>shoB</i>	Small toxic protein ShoB	C1P611
	<i>srnB</i>	Protein SrnB	P13970
	<i>symE</i>	Toxic protein SymE	P39394
	<i>txpA</i>	Toxic peptide TxpA	P54373
	<i>tisB</i>	Small toxic protein TisB	A5A627
2	<i>hicA1</i>	Probable mRNA interferase HicA 1	Q7N1I2
	<i>hicA2</i>	Probable mRNA interferase HicA 2	Q7MZD9
	<i>higB1</i>	Toxin HigB-1	Q9KMG5
	<i>higB2</i>	Toxin HigB-2	Q9KMA6
	<i>hipA</i>	inactivating GltX kinase facilitating persister	P23874

<i>mazF1</i>	Probable endoribonuclease MazF1	Q6MX40
<i>mazF10</i>	toxin mzF10	O07226
<i>mazF2</i>	Probable endoribonuclease MazF2	P9WII1
<i>mazF3</i>	Endoribonuclease MazF3	P9WIH9
<i>mazF4</i>	Endoribonuclease MazF4	P9WII5
<i>mazF5</i>	Probable endoribonuclease MazF5	P95272
<i>mazF6</i>	Endoribonuclease MazF6	P9WII3
<i>mazF7</i>	Probable endoribonuclease MazF7	P0CL62
<i>mazF8</i>	Putative toxin MazF8	P9WIH6
<i>mazF9</i>	Endoribonuclease MazF9	P71650
<i>toxin zeta</i>	toxin zeta	Q54944
<i>pemK</i>	mRNA interferase PemK	P13976
<i>parE1</i>	Toxin ParE1 OS=Caulobacter crescentus	Q9A9T8
<i>parE2</i>	Toxin ParE2	P9WHG5
<i>parE3</i>	Toxin ParE3	Q9A4S4
<i>parE4</i>	Toxin ParE4	Q9A459
<i>pasB</i>	Toxin-like protein	Q75ZL7
<i>doc</i>	Toxin Doc	A0QRY0
<i>RelE</i>	toxin RelE	O50461
<i>relG</i>	toxin RelG	O33348
<i>higB1</i>	Probable endoribonuclease HigB1	P9WJA5
<i>higB2</i>	Putative toxin HigB2	O53468
<i>higB3</i>	Putative toxin HigB3	O53332
<i>parE1</i>	Toxin ParE1	P9WHG7
<i>yoeB</i>	Toxin RelK	P9WF09
<i>VapC1</i>	ribonuclease VapC1	P9WFC1
<i>vapC10</i>	Ribonuclease VapC10	P9WFA7
<i>vapC11</i>	Ribonuclease VapC11	P9WFA5
<i>vapC12</i>	Ribonuclease VapC12	P9WFA3
<i>vapC13</i>	Ribonuclease VapC13	P9WFA1
<i>vapC14</i>	Ribonuclease VapC14	P9WF99
<i>vapC15</i>	Ribonuclease VapC15	P9WF97
<i>vapC16</i>	Ribonuclease VapC16	P0CV93
<i>vapC17</i>	Ribonuclease VapC17	P9WF95
<i>vapC18</i>	Ribonuclease VapC18	P95007
<i>vapC19</i>	Ribonuclease VapC19	P9WF93
<i>vapC2</i>	ribonuclease VapC2	P9WFB9
<i>vapC20</i>	23S rRNA-specific endonuclease VapC20	P95004
<i>vapC21</i>	Ribonuclease VapC21	P9WF91
<i>vapC22</i>	Ribonuclease VapC22	P71623
<i>vapC23</i>	Ribonuclease VapC23	P9WF89

	<i>vapC24</i>	Ribonuclease VapC24	P9WF87
	<i>vapC25</i>	Ribonuclease VapC25	P9WF85
	<i>vapC26</i>	Ribonuclease VapC26	O53779
	<i>vapC27</i>	Ribonuclease VapC27	P9WF83
	<i>vapC28</i>	Ribonuclease VapC28	P9WF81
	<i>vapC29</i>	Ribonuclease VapC29	P9WF79
	<i>vapC3</i>	ribonuclease VapC3	P9WFB7
	<i>vapC30</i>	Ribonuclease VapC30	P9WF77
	<i>vapC31</i>	Ribonuclease VapC31	P9WF75
	<i>vapC32</i>	Ribonuclease VapC32	P9WF73
	<i>vapC33</i>	Ribonuclease VapC33	P9WF69
	<i>vapC34</i>	Putative ribonuclease VapC34	P9WF71
	<i>vapC35</i>	Ribonuclease VapC35	P9WF67
	<i>vapC36</i>	Ribonuclease VapC36	P9WF65
	<i>vapC37</i>	Ribonuclease VapC37	O53501
	<i>vapC38</i>	Ribonuclease VapC38	O53219
	<i>vapC39</i>	Ribonuclease VapC39	P9WF63
	<i>vapC4</i>	Toxin VapC4	O07783
	<i>vapC40</i>	Ribonuclease VapC40	P9WF61
	<i>vapC41</i>	Ribonuclease VapC41	P9WF59
	<i>vapC42</i>	Ribonuclease VapC42	P9WF57
	<i>vapC43</i>	Ribonuclease VapC43	P9WF55
	<i>vapC44</i>	Ribonuclease VapC44	P9WF53
	<i>vapC45</i>	Ribonuclease VapC45	P9WF50
	<i>vapC46</i>	Ribonuclease VapC46	O50411
	<i>vapC47</i>	Ribonuclease VapC47	P9WF49
	<i>vapC48</i>	Ribonuclease VapC48	P9WF47
	<i>vapC49</i>	Ribonuclease VapC49	P9WF51
	<i>vapC5</i>	Ribonuclease VapC5	P96917
	<i>vapC50</i>	Putative ribonuclease VapC50	L0TGF0
	<i>vapC51</i>	Ribonuclease VapC51	L0T5V6
	<i>vapC6</i>	ribonuclease VapC6	P9WFB5
	<i>vapC7</i>	Ribonuclease VapC7	P9WFB3
	<i>vapC8</i>	Ribonuclease VapC8	P9WFB1
	<i>vapC9</i>	ribonuclease VapC9	P9WFA9
	<i>ccdB</i>	Toxin CcdB	P62554
	<i>mqsR</i>	mRNA interferase MqsR	Q46865
	<i>yhaV</i>	Toxin YhaV	P64594
	<i>ParE</i>	Toxin ParE	K4Y5N7
3	<i>abiQ</i>	Endoribonuclease AbiQ	Q9ZJ19
	<i>toxN</i>	Endoribonuclease ToxN	B8X8Z0

4	<i>abiEii</i>	AbiEii	Q48616
	<i>cbtA</i>	Cytoskeleton-binding toxin CbtA	P64524
	<i>cptA</i>	Toxin CptA	Q46824
5	<i>ghoT</i>	Toxin GhoT	P64646

## B. Antitoxinas de referencia

Tipo TA	Gen	Proteína	ID
2	<i>hicB1</i>	Antitoxin HicB 1	Q7N1I3
	<i>hicB2</i>	Antitoxin HicB 2	Q7MZD8
	<i>higA-1</i>	Antitoxin HigA-1	Q9KMG4
	<i>higA-2</i>	Antitoxin HigA-2	Q9KMA5
	<i>hipB</i>	Antitoxin HipB	P23873
	<i>mazE1</i>	Probable antitoxin MazE1	P0CL57
	<i>mazE10</i>	Antitoxin Rv0298	P9WJ09
	<i>mazE2</i>	Probable antitoxin MazE2	O06779
	<i>mazE3</i>	Antitoxin MazE3	O53451
	<i>mazE4</i>	Probable antitoxin MazE4	P9WJ91
	<i>mazE5</i>	Antitoxin MazE5	P9WJ89
	<i>mazE6</i>	Antitoxin MazE6	P9WJ87
	<i>mazE7</i>	Antitoxin MazE7	P9WJ85
	<i>mazE8</i>	Antitoxin MazE8	P0CL60
	<i>mazE9</i>	Antitoxin MazE9	P0CL61
	<i>Antitoxin epsilon</i>	Antitoxin epsilon	Q57231
	<i>pemI</i>	Antitoxin PemI	P13975
	<i>parD1</i>	Antitoxin ParD1	P58091
	<i>parD2</i>	Antitoxin ParD2	P9WJ75
	<i>parD3</i>	Antitoxin ParD3	P0CW74
	<i>parD4</i>	Antitoxin ParD4	Q9A458
	<i>pasA</i>	Anti-toxin-like protein	Q93TU2
	<i>phd</i>	Antitoxin Phd	A0QRX9
	<i>relB</i>	Antitoxin RelB	O50462
	<i>relF</i>	Antitoxin RelF	O33347
	<i>higA1</i>	Antitoxin HigA1	P9WJA7
	<i>higA2</i>	Putative antitoxin HigA2	O53467
	<i>higA3</i>	Putative antitoxin HigA3	O53333
	<i>parD1</i>	Antitoxin ParD1	P9WIJ7

<i>relJ</i>	Antitoxin RelJ	P9WF25
<i>vapB1</i>	Putative antitoxin VapB1	P0CW29
<i>vapB10</i>	Putative antitoxin VapB10	P9WLZ1
<i>vapB11</i>	Antitoxin VapB11	P9WLU3
<i>vapB12</i>	Putative antitoxin VapB12	P9WJ53
<i>vapB13</i>	Putative antitoxin VapB13	P9WJ51
<i>vapB14</i>	Putative antitoxin VapB14	P95262
<i>vapB15</i>	Antitoxin VapB15	P9WLM7
<i>vapB16</i>	Putative antitoxin VapB16	P0CW31
<i>vapB17</i>	Putative antitoxin VapB17	P9WJ49
<i>vapB18</i>	Putative antitoxin VapB18	P9WJ47
<i>vapB19</i>	Putative antitoxin VapB19	P95006
<i>vapB2</i>	Antitoxin VapB2	O07227
<i>vapB20</i>	Antitoxin VapB20	P9WJ44
<i>vapB21</i>	Antitoxin VapB21	P9WJ43
<i>vapB22</i>	Antitoxin VapB22	P71622
<i>vapB23</i>	Putative antitoxin VapB23	P0CW32
<i>vapB24</i>	Putative antitoxin VapB24	P9WJ41
<i>vapB25</i>	Antitoxin VapB25	P0CW33
<i>vapB26</i>	Antitoxin VapB26	O53778
<i>vapB27</i>	Antitoxin VapB27	O07779
<i>vapB28</i>	Antitoxin VapB28	P9WJ39
<i>vapB29</i>	Putative antitoxin VapB29	P9WJ37
<i>vapB3</i>	Antitoxin VapB3	P9WJ59
<i>vapB30</i>	Antitoxin VapB30	P9WJ35
<i>vapB31</i>	Antitoxin VapB31	O53811
<i>vapB32</i>	Antitoxin VapB32	P9WJ33
<i>vapB33</i>	Antitoxin VapB33	O50456
<i>vapB34</i>	Putative antitoxin VapB34	P9WJ31
<i>vapB35</i>	Antitoxin VapB35	P9WF17
<i>vapB36</i>	Putative antitoxin VapB36	P9WJ29
<i>vapB37</i>	Putative antitoxin VapB37	P9WJ27
<i>vapB38</i>	Putative antitoxin VapB38	P9WJ25
<i>vapB39</i>	Antitoxin VapB39	P9WJ23
<i>vapB4</i>	Antitoxin VapB4	P9WF21
<i>vapB40</i>	Antitoxin VapB40	P9WFC3
<i>vapB41</i>	Antitoxin VapB41	P9WJ21
<i>vapB42</i>	Putative antitoxin VapB42	P9WJ19
<i>vapB43</i>	Antitoxin VapB43	P9WL41
<i>vapB44</i>	Putative antitoxin VapB44	P9WJ17
<i>vapB45</i>	Putative antitoxin VapB45	O53464
<i>vapB46</i>	Antitoxin VapB46	P9WF13
<i>vapB47</i>	Antitoxin VapB47	P9WF23



	<i>vapB48</i>	Putative antitoxin VapB48	P9WJ15
	<i>vapB49</i>	Putative antitoxin VapB49	P9WF15
	<i>vapB5</i>	Putative antitoxin VapB5	P9WF19
	<i>vapB50</i>	Putative antitoxin VapB50	O69717
	<i>vapB51</i>	Putative antitoxin VapB51	I6WXS6
	<i>vapB6</i>	Putative antitoxin VapB6	P9WJ57
	<i>vapB7</i>	Putative antitoxin VapB7	O06777
	<i>vapB8</i>	Putative antitoxin VapB8	O06775
	<i>vapB9</i>	Putative antitoxin VapB9	P9WJ55
	<i>ccdA</i>	Antitoxin CcdA	P62552
	<i>mqsA</i>	Antitoxin MqsA	Q46864
	<i>prIF</i>	Antitoxin PrIF	P15373
	<i>PaaA</i>	Antitoxin PaaA	K4XDP0
4	<i>cptB</i>	Antitoxin CptB	P64559
	<i>abiEi</i>	AbiEi	Q48615
	<i>cbeA</i>	Cytoskeleton bundling-enhancing protein CbeA	P76364
5	<i>ghoS</i>	Endoribonuclease GhoS	P0AF61

## C. Sistemas TA hipotéticos en H37Rv

Putativos sistemas TA en H37Rv			
Tipo TA	Gen	Proteína	ID
4	<i>Rv0836c</i>	Uncharacterized protein	O53848
	<i>Rv0837c</i>	Uncharacterized protein	A0A0K2HU81
	<i>Rv1045</i>	Uncharacterized protein	A0A120J0C2
	<i>Rv1044</i>	Uncharacterized protein	R4M4G9
	<i>Rv2826c</i>	Uncharacterized protein	A0A120J2D6
	<i>Rv2827c</i>	Uncharacterized protein	M8CBB6
No caracterizada	<i>Rv0059</i>	Uncharacterized protein	R4M8S2
	<i>Rv0060</i>	Uncharacterized protein	R4M1L0
	<i>Rv0909</i>	Antitoxin Rv0909	P9WJ07
	<i>Rv0910</i>	Toxin Rv0910	P9WJ05
	<i>Rv0918</i>	Uncharacterized protein	A5U0V8
	<i>Rv0919</i>	GCN5-like N-acetyltransferase	A0A120J064
	<i>Rv1545</i>	Uncharacterized protein Rv1545	P9WLU9
	<i>Rv1546</i>	Uncharacterized protein Rv1546	P9WLU7
	<i>Rv1989c</i>	Uncharacterized protein Rv1989c	P9WLP9
	<i>Rv1990c</i>	Uncharacterized protein Rv1990c	P9WLP7
	<i>Rv2034</i>	HTH-type transcriptional regulator Rv2034	O53478
	<i>Rv2035</i>	Uncharacterized protein	A0A120J1H9
	<i>Rv2653c</i>	Toxin Rv2653c	P9WJ13
	<i>Rv2654c</i>	Antitoxin Rv2654c	P9WJ11
	<i>Rv3188</i>	Uncharacterized protein	A0A109T080
	<i>Rv3189</i>	Uncharacterized protein	A0A109T0A2

## D. Reguladores

Reguladores en sistema de tres componentes			
Sistema	Gen	Proteína	ID
HigBA1	<i>secBL</i>	SecB-like chaperone Rv1957	P95257
$\omega$ - $\epsilon$ - $\zeta$	<i>omega</i>	Omega transcriptional repressor	Q57468
paaR-paaA-parE	<i>ECs1069</i>	Putative regulatory protein	A0A0H3JH24



## Bibliografía

- Alberts, Bruce. 2013. *Molecular Biology of the Cell*.
- Albrethsen, Jakob et al. 2013. "Proteomic Profiling of Mycobacterium Tuberculosis Identifies Nutrient-Starvation-Responsive Toxin-Antitoxin Systems." *Molecular & cellular proteomics : MCP* 12(5):1180–91. Retrieved November 6, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/23345537>).
- Ali, Solomon et al. 2016. "Drug Resistance and Population Structure of M.tuberculosis Isolates from Prisons and Communities in Ethiopia." *BMC Infectious Diseases* 16(1):687. Retrieved November 29, 2016 (<http://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-016-2041-x>).
- Allocati, N., M. Masulli, C. Di Ilio, and V. De Laurenzi. 2015. "Die for the Community: An Overview of Programmed Cell Death in Bacteria." *Cell death & disease* 6(1):e1609. Retrieved November 2, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/25611384>).
- Alnimr, Amani M. 2015. "Dormancy Models for Mycobacterium Tuberculosis: A Minireview." *Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology]* 46(3):641–47. Retrieved November 29, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/26413043>).
- Altschul, S. F. et al. 1997. "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs." *Nucleic acids research* 25(17):3389–3402. Retrieved August 25, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/9254694>).
- Alvarez, N., D. Haft, U. A. Hurtado, J. Robledo, and F. Rouzaud. 2016. "Whole-Genome Sequencing of a Haarlem Extensively Drug-Resistant *Mycobacterium Tuberculosis* Clinical Isolate from Medellín, Colombia." *Genome Announcements* 4(3):e00566-16. Retrieved January 12, 2017 (<http://www.ncbi.nlm.nih.gov/pubmed/27313305>).
- Blower, Tim R. et al. 2012. "Identification and Classification of Bacterial Type III Toxin-Antitoxin Systems Encoded in Chromosomal and Plasmid Genomes." *Nucleic acids research* 40(13):6158–73. Retrieved September 30, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/22434880>).
- Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics (Oxford, England)* 30(15):2114–20. Retrieved July 9, 2014 (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4103590&tool=pmcentrez&rendertype=abstract>).
- Bordes, Patricia et al. 2011. "SecB-like Chaperone Controls a Toxin-Antitoxin Stress-Responsive System in Mycobacterium Tuberculosis." *Proceedings of the National Academy of Sciences of the United States of America* 108(20):8438–43. Retrieved December 12, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/21536872>).
- Brites, Daniela and Sebastien Gagneux. 2015. "Co-Evolution of Mycobacterium Tuberculosis and Homo Sapiens." *Immunological reviews* 264(1):6–24. Retrieved November 30, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/25703549>).
- Brosch, R. et al. 2002. "A New Evolutionary Scenario for the Mycobacterium Tuberculosis Complex." *Proceedings of the National Academy of Sciences of the United States of*

- America* 99(6):3684–89. Retrieved December 5, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/11891304>).
- Brudey, Karine et al. 2006. "Mycobacterium Tuberculosis Complex Genetic Diversity: Mining the Fourth International Spoligotyping Database (SpolDB4) for Classification, Population Genetics and Epidemiology." *BMC Microbiology* 6(1):23. Retrieved December 5, 2016 (<http://bmcmicrobiol.biomedcentral.com/articles/10.1186/1471-2180-6-23>).
- Brunner, Katharina et al. 2016. "Inhibitors of the Cysteine Synthase CysM with Antibacterial Potency against Dormant Mycobacterium Tuberculosis." *Journal of medicinal chemistry* 59(14):6848–59. Retrieved November 29, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/27379713>).
- Caws, Maxine et al. 2008. "The Influence of Host and Bacterial Genotype on the Development of Disseminated Disease with Mycobacterium Tuberculosis." *PLoS pathogens* 4(3):e1000034. Retrieved October 19, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/18369480>).
- Chan, Wai Ting, Manuel Espinosa, and Chew Chieng Yeo. 2016. "Keeping the Wolves at Bay: Antitoxins of Prokaryotic Type II Toxin-Antitoxin Systems." *Frontiers in molecular biosciences* 3:9. Retrieved October 3, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/27047942>).
- Chen, Yen-Chun, Tsunglin Liu, Chun-Hui Yu, Tzen-Yuh Chiang, and Chi-Chuan Hwang. 2013. "Effects of GC Bias in next-Generation-Sequencing Data on de Novo Genome Assembly." *PloS one* 8(4):e62856. Retrieved November 1, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/23638157>).
- Click, E. S., P. K. Moonan, C. A. Winston, L. S. Cowan, and J. E. Oeltmann. 2012. "Relationship Between Mycobacterium Tuberculosis Phylogenetic Lineage and Clinical Site of Tuberculosis." *Clinical Infectious Diseases* 54(2):211–19. Retrieved November 1, 2016 (<http://cid.oxfordjournals.org/lookup/doi/10.1093/cid/cir788>).
- Cole, S. T. et al. 1998. "Deciphering the Biology of Mycobacterium Tuberculosis from the Complete Genome Sequence." *Nature* 393(6685):537–44. Retrieved November 1, 2016 (<http://www.nature.com/doi/10.1038/31159>).
- Coll, Francesc et al. 2014. "A Robust SNP Barcode for Typing Mycobacterium Tuberculosis Complex Strains." *Nature communications* 5:4812. Retrieved July 8, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/25176035>).
- Coll, Francesc et al. 2015. "Rapid Determination of Anti-Tuberculosis Drug Resistance from Whole-Genome Sequences." *Genome Medicine* 7(1):51. Retrieved January 12, 2017 (<http://www.ncbi.nlm.nih.gov/pubmed/26019726>).
- Coscolla, Mireia and Sebastien Gagneux. 2014. "Consequences of Genomic Diversity in Mycobacterium Tuberculosis." *Seminars in immunology* 26(6):431–44. Retrieved June 20, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/25453224>).
- Coscolla, Mireia and Sebastien Gagneux. 2010. "Does M. Tuberculosis Genomic Diversity Explain Disease Diversity?" *Drug discovery today. Disease mechanisms* 7(1):e43–59. Retrieved December 5, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/21076640>).
- Demidenok, Oksana I., Arseny S. Kaprelyants, and Anna V Goncharenko. 2014. "Toxin-Antitoxin vapBC Locus Participates in Formation of the Dormant State in Mycobacterium Smegmatis." *FEMS microbiology letters* 352(1):69–77. Retrieved October 15, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/24417293>).
- Eddy, S. R. 1998. "Profile Hidden Markov Models." *Bioinformatics (Oxford, England)* 14(9):755–63. Retrieved August 25, 2016

- (<http://www.ncbi.nlm.nih.gov/pubmed/9918945>).
- Edgar, Robert C. 2004. "MUSCLE: A Multiple Sequence Alignment Method with Reduced Time and Space Complexity." *BMC bioinformatics* 5:113. Retrieved August 25, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/15318951>).
- Fiebig, Aretha, Cyd Marie Castro Rojas, Dan Siegal-Gaskins, and Sean Crosson. 2010. "Interaction Specificity, Toxicity and Regulation of a Paralogous Set of ParE/RelE-Family Toxin-Antitoxin Systems." *Molecular microbiology* 77(1):236–51. Retrieved October 7, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/20487277>).
- Finn, Robert D. et al. 2016. "The Pfam Protein Families Database: Towards a More Sustainable Future." *Nucleic Acids Research* 44(D1):D279–85. Retrieved August 26, 2016 (<http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkv1344>).
- Forrellad, Marina A. et al. 2013. "Virulence Factors of the *Mycobacterium Tuberculosis* Complex." *Virulence* 4(1):3–66. Retrieved November 1, 2016 (<http://www.tandfonline.com/doi/abs/10.4161/viru.22329>).
- Fozo, Elizabeth M. et al. 2010a. "Abundance of Type I Toxin-Antitoxin Systems in Bacteria: Searches for New Candidates and Discovery of Novel Families." *Nucleic acids research* 38(11):3743–59. Retrieved September 30, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/20156992>).
- Fozo, Elizabeth M. et al. 2010b. "Abundance of Type I Toxin-Antitoxin Systems in Bacteria: Searches for New Candidates and Discovery of Novel Families." *Nucleic acids research* 38(11):3743–59. Retrieved December 17, 2015 (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2887945&tool=pmcentrez&rendertype=abstract>).
- Gardy, Jennifer L. et al. 2011. "Whole-Genome Sequencing and Social-Network Analysis of a Tuberculosis Outbreak." *New England Journal of Medicine* 364(8):730–39. Retrieved January 13, 2017 (<http://www.ncbi.nlm.nih.gov/pubmed/21345102>).
- Gelens, Lendert, Lydia Hill, Alexandra Vandervelde, Jan Danckaert, and Remy Loris. 2013. "A General Model for Toxin-Antitoxin Module Dynamics Can Explain Persister Cell Formation in *E. Coli*." *PLoS computational biology* 9(8):e1003190. Retrieved November 6, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/24009490>).
- Gerdes, Kenn, ed. 2013. *Prokaryotic Toxin-Antitoxins*. Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved December 3, 2015 (<http://www.springerlink.com/index/10.1007/978-3-642-33253-1>).
- Goeders, Nathalie, Ray Chai, Bihe Chen, Andrew Day, and George P. C. Salmond. 2016. "Structure, Evolution, and Functions of Bacterial Type III Toxin-Antitoxin Systems." *Toxins* 8(10). Retrieved November 6, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/27690100>).
- Grazziotin, Ana Laura, Newton M. Vidal, and Thiago M. Venancio. 2015. "Uncovering Major Genomic Features of Essential Genes in Bacteria and a Methanogenic Archaea." *The FEBS journal* 282(17):3395–3411. Retrieved October 5, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/26084810>).
- Guo, Yunxue et al. 2014. "RalR (a DNase) and RaiA (a Small RNA) Form a Type I Toxin-Antitoxin System in *Escherichia Coli*." *Nucleic acids research* 42(10):6448–62. Retrieved December 9, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/24748661>).
- Gupta, Manish et al. 2016. "The Chromosomal parDE2 Toxin-Antitoxin System of *Mycobacterium Tuberculosis* H37Rv: Genetic and Functional Characterization." *Frontiers in microbiology* 7:886. Retrieved October 15, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/27379032>).
- Hazan, R. and H. Engelberg-Kulka. 2004. "Escherichia Coli mazEF-Mediated Cell Death

- as a Defense Mechanism That Inhibits the Spread of Phage P1." *Molecular Genetics and Genomics* 272(2):227–34.
- Holliday, Gemma L., John B. O. Mitchell, and Janet M. Thornton. 2009. "Understanding the Functional Roles of Amino Acid Residues in Enzyme Catalysis." *Journal of molecular biology* 390(3):560–77. Retrieved October 21, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/19447117>).
- Hurtado, U. A., J. S. Solano, A. Rodriguez, J. Robledo, and F. Rouzaud. 2016. "Draft Genome Sequence of a Mycobacterium Africanum Clinical Isolate from Antioquia, Colombia." *Genome announcements* 4(3). Retrieved January 12, 2017 (<http://www.ncbi.nlm.nih.gov/pubmed/27257203>).
- Hyatt, Doug et al. 2010. "Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification." *BMC Bioinformatics* 11(1):119. Retrieved August 25, 2016 (<http://www.biomedcentral.com/1471-2105/11/119>).
- Ioerger, T. R. et al. 2010. "Variation among Genome Sequences of H37Rv Strains of Mycobacterium Tuberculosis from Multiple Laboratories." *Journal of Bacteriology* 192(14):3645–53. Retrieved November 1, 2016 (<http://jb.asm.org/cgi/doi/10.1128/JB.00166-10>).
- Isaza, Juan Pablo et al. 2012. "Whole Genome Shotgun Sequencing of One Colombian Clinical Isolate of Mycobacterium Tuberculosis Reveals DosR Regulon Gene Deletions." *FEMS microbiology letters* 330(2):113–20. Retrieved February 2, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/22404577>).
- Jackson, Mary. 2014. "Diversity and Evolution of Mycobacterium."
- Kaessmann, Henrik. 2010. "Origins, Evolution, and Phenotypic Impact of New Genes." *Genome research* 20(10):1313–26. Retrieved October 23, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/20651121>).
- Kaur, Devinder, Marcelo E. Guerin, Henrieta Skovierová, Patrick J. Brennan, and Mary Jackson. 2009. "Chapter 2: Biogenesis of the Cell Wall and Other Glycoconjugates of Mycobacterium Tuberculosis." *Advances in applied microbiology* 69:23–78. Retrieved December 6, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/19729090>).
- Keren, I., S. Minami, E. Rubin, and K.C2-3119538 Lewis. 2011. "Characterization and Transcriptome Analysis of Mycobacterium Tuberculosis Persisters." *MBio* 2(3):e00100–11 ST–Characterization and transcriptome. Retrieved (<http://www.ncbi.nlm.nih.gov/pubmed/21673191?dopt=Citation>).
- Kersey, P. et al. 2004. "Integr8 and Genome Reviews: Integrated Views of Complete Genomes and Proteomes." *Nucleic Acids Research* 33(Database issue):D297–302. Retrieved April 4, 2017 (<http://www.ncbi.nlm.nih.gov/pubmed/15608201>).
- Korch, Shaleen B., Vandana Malhotra, Heidi Contreras, and Josephine E. Clark-Curtiss. 2015. "The Mycobacterium Tuberculosis relBE Toxin:antitoxin Genes Are Stress-Responsive Modules That Regulate Growth through Translation Inhibition." *Journal of microbiology (Seoul, Korea)* 53(11):783–95. Retrieved October 7, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/26502963>).
- Krishnan, Nitya et al. 2011. "Mycobacterium Tuberculosis Lineage Influences Innate Immune Response and Virulence and Is Associated with Distinct Cell Envelope Lipid Profiles." *PLoS one* 6(9):e23870. Retrieved November 4, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/21931620>).
- Larsson, A. 2014. "AliView: A Fast and Lightweight Alignment Viewer and Editor for Large Datasets." *Bioinformatics* 30(22):3276–78. Retrieved August 26, 2016 (<http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btu531>).
- Lawn, Stephen D. and Alimuddin I. Zumla. 2011. "Tuberculosis." *Lancet (London,*



- England*) 378(9785):57–72. Retrieved November 1, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/21420161>).
- Lechner, Marcus et al. 2011. "Proteinortho: Detection of (Co-)Orthologs in Large-Scale Analysis." *BMC bioinformatics* 12:124. Retrieved August 26, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/21526987>).
- Lee, Ki-Young and Bong-Jin Lee. 2016. "Structure, Biology, and Therapeutic Application of Toxin-Antitoxin Systems in Pathogenic Bacteria." *Toxins* 8(10). Retrieved December 2, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/27782085>).
- Lee, Robyn S. and Marcel A. Behr. 2016. "The Implications of Whole-Genome Sequencing in the Control of Tuberculosis." *Therapeutic advances in infectious disease* 3(2):47–62. Retrieved January 12, 2017 (<http://www.ncbi.nlm.nih.gov/pubmed/27034776>).
- Lee, Seung Heon. 2016. "Tuberculosis Infection and Latent Tuberculosis." *Tuberculosis and respiratory diseases* 79(4):201–6. Retrieved December 6, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/27790271>).
- Leplae, Raphaël et al. 2011a. "Diversity of Bacterial Type II Toxin-Antitoxin Systems: A Comprehensive Search and Functional Analysis of Novel Families." *Nucleic acids research* 39(13):5513–25. Retrieved November 1, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/21422074>).
- Leplae, Raphaël et al. 2011b. "Diversity of Bacterial Type II Toxin-Antitoxin Systems: A Comprehensive Search and Functional Analysis of Novel Families." *Nucleic acids research* 39(13):5513–25. Retrieved September 30, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/21422074>).
- Liu, Qingyun et al. 2016. "Genetic Features of Mycobacterium Tuberculosis Modern Beijing Sublineage." *Emerging microbes & infections* 5(2):e14. Retrieved January 13, 2017 (<http://www.ncbi.nlm.nih.gov/pubmed/26905026>).
- Lobato-Márquez, Damián, Ramón Díaz-Orejas, and Francisco García-Del Portillo. 2016. "Toxin-Antitoxins and Bacterial Virulence." *FEMS microbiology reviews* 40(5):592–609. Retrieved December 1, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/27476076>).
- Lucía Martínez, Martha, Mancel Enrique Martínez Duran Director Vigilancia Análisis del Riesgo en Salud Pública Oscar Eduardo Pacheco García, Subdirector de Análisis del Riesgo Respuesta Inmediata DOCUMENTO ELABORADO POR Santiago Fadul Pérez Martha Patricia López Pérez Profesional Especializado Equipo de Micobacterias, and Santiago Fadul Pérez Martha Patricia López Pérez Profesional Especializado Equipo de Micobacterias. 2016. "Protocolo de Vigilancia En Salud Pública TUBERCULOSIS Hernán Quijada Bonilla PROTOCOLO DE VIGILANCIA EN SALUD PUBLICA TUBERCULOSIS."
- Luo, Ruibang et al. 2012. "SOAPdenovo2: An Empirically Improved Memory-Efficient Short-Read de Novo Assembler." *GigaScience* 1(1):18. Retrieved November 20, 2014 (<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3626529&tool=pmcentrez&rendertype=abstract>).
- Marmiesse, Magali et al. 2004. "Macro-Array and Bioinformatic Analyses Reveal Mycobacterial 'Core' Genes, Variation in the ESAT-6 Gene Family and New Phylogenetic Markers for the Mycobacterium Tuberculosis Complex." *Microbiology (Reading, England)* 150(Pt 2):483–96. Retrieved October 4, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/14766927>).
- McEvoy, Christopher R. E. et al. 2012. "Comparative Analysis of Mycobacterium Tuberculosis Pe and Ppe Genes Reveals High Sequence Variation and an Apparent

- Absence of Selective Constraints” edited by O. Neyrolles. *PLoS ONE* 7(4):e30593. Retrieved November 1, 2016 (<http://dx.plos.org/10.1371/journal.pone.0030593>).
- Medini, Duccio, Claudio Donati, Hervé Tettelin, Vega Masignani, and Rino Rappuoli. 2005. “The Microbial Pan-Genome.” *Current opinion in genetics & development* 15(6):589–94. Retrieved October 4, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/16185861>).
- Van Melderen, Laurence. 2010. “Toxin-Antitoxin Systems: Why so Many, What For?” *Current opinion in microbiology* 13(6):781–85. Retrieved October 6, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/21041110>).
- Miallau, Linda et al. 2013. “Comparative Proteomics Identifies the Cell-Associated Lethality of *M. Tuberculosis* RelBE-like Toxin-Antitoxin Complexes.” *Structure (London, England : 1993)* 21(4):627–37. Retrieved October 31, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/23523424>).
- Narayanan, Sujatha and Uday Deshpande. 2013. “Whole-Genome Sequences of Four Clinical Isolates of *Mycobacterium Tuberculosis* from Tamil Nadu, South India.” *Genome announcements* 1(3). Retrieved October 23, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/23788533>).
- Nguyen, Van Anh Thi et al. 2016. “*Mycobacterium Tuberculosis* Lineages and Anti-Tuberculosis Drug Resistance in Reference Hospitals across Viet Nam.” *BMC microbiology* 16(1):167. Retrieved November 4, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/27464737>).
- Page, Rebecca and Wolfgang Peti. 2016. “Toxin-Antitoxin Systems in Bacterial Growth Arrest and Persistence.” *Nature Chemical Biology* 12(4):208–14. Retrieved December 9, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/26991085>).
- Pandey, Deo Prakash and Kenn Gerdes. 2005. “Toxin-Antitoxin Loci Are Highly Abundant in Free-Living but Lost from Host-Associated Prokaryotes.” *Nucleic acids research* 33(3):966–76. Retrieved October 6, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/15718296>).
- Periwal, Vinita et al. 2015. “Comparative Whole-Genome Analysis of Clinical Isolates Reveals Characteristic Architecture of *Mycobacterium Tuberculosis* Pangenome.” *PloS one* 10(4):e0122979. Retrieved October 5, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/25853708>).
- Peters, Julian S. et al. 2016. “Identification of Quantitative Proteomic Differences between *Mycobacterium Tuberculosis* Lineages with Altered Virulence.” *Frontiers in microbiology* 7:813. Retrieved June 20, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/27303394>).
- Ramage, Holly R., Lynn E. Connolly, and Jeffery S. Cox. 2009. “Comprehensive Functional Analysis of *Mycobacterium Tuberculosis* Toxin-Antitoxin Systems: Implications for Pathogenesis, Stress Responses, and Evolution.” *PLoS genetics* 5(12):e1000767. Retrieved September 30, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/20011113>).
- Realpe, Teresa et al. 2014. “Population Structure among *Mycobacterium Tuberculosis* Isolates from Pulmonary Tuberculosis Patients in Colombia.” *PloS one* 9(4):e93848. Retrieved November 4, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/24747767>).
- Rodríguez, Juan Germán, Camilo Pino, Andreas Tauch, and Martha Isabel Murcia. 2015. “Complete Genome Sequence of the Clinical Beijing-Like Strain *Mycobacterium Tuberculosis* 323 Using the PacBio Real-Time Sequencing Platform.” *Genome Announcements* 3(2):e00371-15. Retrieved January 12, 2017 (<http://www.ncbi.nlm.nih.gov/pubmed/25931600>).

- Rose, Graham et al. 2013. "Mapping of Genotype-Phenotype Diversity among Clinical Isolates of Mycobacterium Tuberculosis by Sequence-Based Transcriptional Profiling." *Genome biology and evolution* 5(10):1849–62. Retrieved October 31, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/24115728>).
- Sala, Ambre, Patricia Bordes, and Pierre Genevaux. 2014. "Multiple Toxin-Antitoxin Systems in Mycobacterium Tuberculosis." *Toxins* 6(3):1002–20. Retrieved September 2, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/24662523>).
- Singhal, Parul et al. 2016. "A Study on Pre-XDR & XDR Tuberculosis & Their Prevalent Genotypes in Clinical Isolates of Mycobacterium Tuberculosis in North India." *The Indian journal of medical research* 143(3):341–47. Retrieved November 29, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/27241648>).
- van Soolingen, D., P. E. de Haas, P. W. Hermans, and J. D. van Embden. 1994. "DNA Fingerprinting of Mycobacterium Tuberculosis." *Methods in enzymology* 235:196–205. Retrieved July 5, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/8057895>).
- Sreenu, Vattipally B., Pankaj Kumar, Javaregowda Nagaraju, and Hampapathalu A. Nagarajam. 2007. "Simple Sequence Repeats in Mycobacterial Genomes." *Journal of biosciences* 32(1):3–15. Retrieved November 1, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/17426376>).
- Steenken, W., W. H. Oatway, and S. A. Petroff. 1934. "BIOLOGICAL STUDIES OF THE TUBERCLE BACILLUS : III. DISSOCIATION AND PATHOGENICITY OF THE R AND S VARIANTS OF THE HUMAN TUBERCLE BACILLUS (H(37))." *The Journal of experimental medicine* 60(4):515–40. Retrieved November 1, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/19870319>).
- Thwaites, Guy et al. 2008. "Relationship between Mycobacterium Tuberculosis Genotype and the Clinical Phenotype of Pulmonary and Meningeal Tuberculosis." *Journal of clinical microbiology* 46(4):1363–68. Retrieved October 19, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/18287322>).
- Tiwari, Prabhakar et al. 2015. "MazF Ribonucleases Promote Mycobacterium Tuberculosis Drug Tolerance and Virulence in Guinea Pigs." *Nature communications* 6:6059. Retrieved July 6, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/25608501>).
- Walker, Timothy M. et al. 2013. "Whole-Genome Sequencing to Delineate Mycobacterium Tuberculosis Outbreaks: A Retrospective Observational Study." *The Lancet Infectious Diseases* 13(2):137–46. Retrieved January 13, 2017 (<http://www.ncbi.nlm.nih.gov/pubmed/23158499>).
- Wang, Xiaoxue et al. 2012. "A New Type V Toxin-Antitoxin System Where mRNA for Toxin GhoT Is Cleaved by Antitoxin GhoS." *Nature chemical biology* 8(10):855–61. Retrieved January 23, 2017 (<http://www.ncbi.nlm.nih.gov/pubmed/22941047>).
- Wen, Yurong, Ester Behiels, and Bart Devreese. 2014a. "Toxin-Antitoxin Systems: Their Role in Persistence, Biofilm Formation, and Pathogenicity." *Pathogens and Disease* 70(3):240–49. Retrieved December 13, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/24478112>).
- Wen, Yurong, Ester Behiels, and Bart Devreese. 2014b. "Toxin-Antitoxin Systems: Their Role in Persistence, Biofilm Formation, and Pathogenicity." *Pathogens and disease* 70(3):240–49. Retrieved January 18, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/24478112>).
- WHO. 2016. "World Health Statistics 2016: Monitoring Health for the SDGs." *WHO*. Retrieved November 1, 2016 ([http://www.who.int/gho/publications/world\\_health\\_statistics/2016/en/](http://www.who.int/gho/publications/world_health_statistics/2016/en/)).
- Xiao, Tongyang et al. 2016. "[Polymorphisms of Toxin-Antitoxin-Chaperone System of

- Mycobacterium Tuberculosis Complex in China].” *Zhonghua liu xing bing xue za zhi = Zhonghua liuxingbingxue zazhi* 37(3):394–97. Retrieved October 31, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/27005544>).
- Yamaguchi, Y., JH Park, and M. Inouye. 2011. “Toxin-Antitoxin Systems in Bacteria and Archaea.” *Annual review of genetics*. Retrieved January 18, 2016 (<http://www.annualreviews.org/doi/abs/10.1146/annurev-genet-110410-132412>).
- Yamaguchi, Yoshihiro and Masayori Inouye. 2016. “Toxin – Antitoxin Systems in Bacteria and Archaea.” 97–107.
- Yamaguchi, Yoshihiro, Jung-Ho Park, and Masayori Inouye. 2011. “Toxin-Antitoxin Systems in Bacteria and Archaea.” *Annual Review of Genetics* 45(1):61–79. Retrieved December 11, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/22060041>).
- Yimer, Solomon A. et al. 2015. “Mycobacterium Tuberculosis Lineage 7 Strains Are Associated with Prolonged Patient Delay in Seeking Treatment for Pulmonary Tuberculosis in Amhara Region, Ethiopia.” *Journal of clinical microbiology* 53(4):1301–9. Retrieved June 20, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/25673798>).
- Yruela, Inmaculada, Bruno Contreras-Moreira, Carlos Magalhães, Nuno S. Osório, and Jesús Gonzalo-Asensio. 2017. “Mycobacterium Tuberculosis Complex Exhibits Lineage-Specific Variations Affecting Protein Ductility and Epitope Recognition.” *Genome Biology and Evolution* evw279. Retrieved January 13, 2017 (<http://www.ncbi.nlm.nih.gov/pubmed/28062754>).
- Yuen, Courtney M., Ekaterina V Kurbatova, Eleanor S. Click, J. Sean Cavanaugh, and J. Peter Cegielski. 2013. “Association between Mycobacterium Tuberculosis Complex Phylogenetic Lineage and Acquired Drug Resistance.” *PloS one* 8(12):e83006. Retrieved November 4, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/24376623>).
- Zaychikova, Marina V et al. 2015. “Mycobacterium Tuberculosis Type II Toxin-Antitoxin Systems: Genetic Polymorphisms and Functional Properties and the Possibility of Their Use for Genotyping.” *PloS one* 10(12):e0143682. Retrieved June 20, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/26658274>).
- Zhu, Ling, Jared D. Sharp, Hiroshi Kobayashi, Nancy A. Woychik, and Masayori Inouye. 2010. “Noncognate Mycobacterium Tuberculosis Toxin-Antitoxins Can Physically and Functionally Interact.” *The Journal of biological chemistry* 285(51):39732–38. Retrieved November 6, 2016 (<http://www.ncbi.nlm.nih.gov/pubmed/20876537>).