



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**POSGRADO EN CIENCIAS BIOLÓGICAS**  
FACULTAD DE CIENCIAS  
BIOMEDICINA

**INVESTIGACIÓN DE REDES EPISTÁTICAS EN *Mycobacterium tuberculosis* USANDO  
GENOMAS COMPLETOS PARA LA IDENTIFICACIÓN DE GENES RELEVANTES**

**TESIS**

QUE PARA OPTAR POR EL GRADO DE:

**DOCTORA EN CIENCIAS**

PRESENTA:

**ALI BERENICE POSADA REYES**

**TUTOR PRINCIPAL DE TESIS: DR. SALVADOR FONSECA CORONADO**  
FACULTAD DE CIENCIAS, UNAM

**COMITÉ TUTOR: DR. SANTIAGO ÁVILA RÍOS**  
INSTITUTO NACIONAL DE ENFERMEDADES RESPIRATORIAS, INER

**COMITÉ TUTOR: DRA. YALBI ITZEL BALDERAS MARTÍNEZ**  
INSTITUTO NACIONAL DE ENFERMEDADES RESPIRATORIAS, INER



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.





**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**POSGRADO EN CIENCIAS BIOLÓGICAS**  
FACULTAD DE CIENCIAS  
BIOMEDICINA

**INVESTIGACIÓN DE REDES EPISTÁTICAS EN *Mycobacterium tuberculosis* USANDO  
GENOMAS COMPLETOS PARA LA IDENTIFICACIÓN DE GENES RELEVANTES**

**TESIS**

QUE PARA OPTAR POR EL GRADO DE:

**DOCTORA EN CIENCIAS**

PRESENTA:

**ALI BERENICE POSADA REYES**

**TUTOR PRINCIPAL DE TESIS: DR. SALVADOR FONSECA CORONADO**  
FACULTAD DE CIENCIAS, UNAM

**COMITÉ TUTOR: DR. SANTIAGO ÁVILA RÍOS**  
INSTITUTO NACIONAL DE ENFERMEDADES RESPIRATORIAS, INER

**COMITÉ TUTOR: DRA. YALBI ITZEL BALDERAS MARTÍNEZ**  
INSTITUTO NACIONAL DE ENFERMEDADES RESPIRATORIAS, INER

COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS  
FACULTAD DE CIENCIAS  
DIVISIÓN ACADÉMICA DE INVESTIGACIÓN Y POSGRADO  
OFICIO FCIE/DAIP/361/2022  
ASUNTO: Oficio de Jurado

**M. en C. Ivonne Ramírez Wence**  
**Directora General de Administración Escolar, UNAM**  
**Presente**

Me permito informar a usted que en la reunión ordinaria del Comité Académico del Posgrado en Ciencias Biológicas, celebrada el día **06 de junio de 2022** se aprobó el siguiente jurado para el examen de grado de **DOCTORA EN CIENCIAS** de la estudiante **POSADA REYES ALI BERENICE** con número de cuenta **410009033** con la tesis titulada: “**Investigación de redes epistáticas en *Mycobacterium tuberculosis* usando genomas completos para la identificación de genes relevantes.**”, realizada bajo la dirección del (la) **DR. SALVADOR FONSECA CORONADO:**

Presidente: **DR. ERNESTO PÉREZ RUEDA**  
Vocal: **DRA. LESLIE CHÁVEZ GALÁN**  
Vocal: **DR. JORGE ALBERTO BARRIOS PAYÁN**  
Vocal: **DR. VICENTE DIEGO ORTEGA DEL VECCHYO**  
Secretario: **DR. SANTIAGO ÁVILA RÍOS**

Sin otro particular, me es grato enviarle un cordial saludo.

**ATENTAMENTE**  
**“POR MI RAZA HABLARÁ EL ESPÍRITU”**  
Ciudad Universitaria, Cd. Mx., a 15 de agosto de 2022

**COORDINADOR DEL PROGRAMA**



**DR. ADOLFO GERARDO NAVARRO SIGÜENZA**



## **AGRADECIMIENTOS INSTITUCIONALES**

Al Posgrado en Ciencias Biológicas (PCB) de la Universidad Nacional Autónoma de México (UNAM) por todo el apoyo académico otorgado para mi formación científica.

Al Consejo Nacional de Ciencia y Tecnología (Conacyt) por concederme la beca de manutención para realizar mis estudios de doctorado (CVU 631936).

A mi Tutor Principal, el Dr. Salvador Fonseca Coronado y a los miembros de mi Comité Tutor: el Dr. Santiago Ávila Ríos y la Dra. Yalbi Itzel Balderas Martínez, por aceptar la dirección de esta tesis y por su apoyo académico durante todo el proyecto doctoral.

## **AGRADECIMIENTOS A TÍTULO PERSONAL**

A mis padres, gracias por todo su esfuerzo, apoyo y amor durante todo mi desarrollo académico, hoy agradezco profundamente que siempre estuviéramos juntos, hoy conservo esos momentos y los atesoro en lo profundo de mi corazón. Gracias por tan bellos momentos.

A mi esposo, Carlos, gracias por ser mi fuerza, paz, poesía de amor y calma en la vida. Agradezco que llegaras a mi vida.

A mi hermano, Josué, por tu fuerza para enfrentar el camino. Gracias hermano por apoyarme para continuar con mi sueño académico y personal.

A mi tía Ofelia, gracias por comprender mi camino, apoyar mis sueños personales y académicos, por escucharme y alentarme en todo momento para continuar.

A todos mis familiares y amigos, gracias por todo su apoyo.

Al Dr. Salvador Fonseca Coronado, Dr. Santiago Ávila Ríos, Dra. Yalbi Itzel Balderas Martínez, Dr. Alejandro Escobar Gutiérrez y al Dr. Pablo Vinuesa, gracias por todo su apoyo académico y confianza para navegar en el conocimiento perteneciente al mundo de la microbiología, en el que fue todo un reto aprender de las poblaciones bacterianas; sin embargo, con su colaboración y apoyo disfruté mucho el proceso. Gracias por enseñarme y permitirme aprender de ustedes.

Al Laboratorio de Ingeniería Genómica del Centro de Ciencias Genómicas, UNAM, por la estancia académica (febrero - diciembre 2019). Dr. Pablo Vinuesa, Dra. Susana Brom Klanner y Dr. David René Romero Camarena, gracias por recibirme durante la estancia académica y permitirme trabajar en su laboratorio.

Al Laboratorio de cómputo de alto rendimiento de la Coordinación del departamento de Matemáticas, en especial al M. en C. Miguel Ángel Pérez León, por el soporte computacional

A los miembros del honorable jurado: Dr. Ernesto Pérez Rueda, Dra. Leslie Chávez Galán, Dr. Jorge Alberto Barrios Payán, Dr. Vicente Diego Ortega del Vecchyo y Dr. Santiago Ávila Ríos, gracias por su revisión, aportaciones y comentarios a la tesis.

Por último, gracias a la UNAM, mi *alma mater*, por contribuir en mi formación académica, con mucho honor, mi corazón es águila mexicana y cóndor andino.



# DEDICATORIA

*In memoriam Delfino Posada Márquez†*  
(21 de mayo de 1967 – 6 de mayo de 2020)

Con amor ante tu ausencia, dedico este trabajo para ti, papá.

Hoy, a la distancia, conservo tu amor y tus sabias palabras,  
en una suave armonía, que oscilan en el viento.

Te recuerdo, con cada sonrisa cada vez que veo a mamá, y  
tu singular carisma heredado a mi hermano.

Ellos atesoran, junto a mí, tu recuerdo y juntos mantenemos  
tu memoria viva en cada segundo.

Gracias por todo, papá.

B. Posada

# ÍNDICE

ABREVIATURAS.....	1
RESUMEN.....	3
ABSTRACT .....	4
CAPÍTULO 1. INTRODUCCIÓN.....	5
1.1. Tuberculosis en el mundo antiguo.....	5
1.2. La enfermedad de tuberculosis .....	7
1.3. Inmunología en la infección por <i>Mycobacterium tuberculosis</i> .....	8
1.4. Epidemiología de la Tuberculosis.....	12
1.5. Tratamiento de la Tuberculosis .....	13
1.5.1. Resistencia a los antibióticos en Tuberculosis .....	15
1.6. Complejo <i>Mycobacterium tuberculosis</i> .....	16
1.7. <i>Mycobacterium tuberculosis</i> .....	19
1.7.1. Descripción del genoma de <i>Mycobacterium tuberculosis</i> .....	19
1.7.2. Mecanismos de resistencia antimicrobiana en el genoma de <i>Mycobacterium tuberculosis</i> .....	21
1.8. Epistasis.....	24
1.8.1. Epistasis en <i>Mycobacterium tuberculosis</i> .....	25
CAPÍTULO 2. OBJETIVOS .....	27
2.1. Objetivo general .....	27
2.2. Objetivos particulares .....	27
CAPÍTULO 3. HIPÓTESIS .....	27
CAPÍTULO 4. ANTECEDENTES.....	27
CAPÍTULO 5. METODOLOGÍA COMPUTACIONAL.....	29
5.1. Conjunto de datos .....	29
5.2. Generación de Alineamiento de genoma completo.....	31
5.3. Análisis filogenético .....	32
5.4. Predicción de la caracterización de resistencia genética a los antibióticos.....	33
5.5. Detección computacional de interacciones epistáticas.....	33
5.5.1. Detección de interacciones epistáticas .....	33
5.5.2. Análisis de enriquecimiento funcional .....	36
5.5.3. Redes de Interacciones Epistáticas .....	36

CAPÍTULO 6. RESULTADOS .....	37
6.1. Detección de interacciones epistáticas.....	39
6.2. Análisis de la Red Epistática <i>glgB</i> y <i>oppA</i> , genes blanco de co-selección .....	40
6.3. Enriquecimiento de términos de GO para redes epistáticas.....	43
6.4. Distribución alélica en <i>loci</i> de genes que interactúan con <i>glgB</i> y <i>oppA</i> .....	44
CAPÍTULO 7. DISCUSIÓN.....	50
7.1. Estimación de filogenia y resistencia a los antibióticos .....	50
7.2. Detección de interacciones epistáticas en los genes blanco de co-selección .....	50
7.3. Enriquecimiento de términos de GO para redes epistáticas.....	52
7.4. Genes reportados por el catálogo de la OMS en interacción con <i>glgB</i> y <i>oppA</i> . .....	55
CAPÍTULO 8. CONCLUSIONES .....	57
CAPÍTULO 9. PERSPECTIVAS .....	58
CAPÍTULO 10.REFERENCIAS .....	59
CAPÍTULO 11.ANEXO I .....	71

## ÍNDICE DE TABLAS

Tabla 1. Principales fármacos contra Tuberculosis y sus efectos adversos.....	14
Tabla 2. Mecanismos de acción de fármacos y genes de resistencia en el genoma de <i>Mycobacterium tuberculosis</i> .....	22
Tabla 3. Valores de las evaluaciones de modelos de filogenia con RAxML Next Generation..	37
Tabla 4. Estadísticas de las interacciones con valores atípicos finales.....	40
Tabla 5. Genes de resistencia a los antibióticos reportados por el catálogo de la OMS encontrados en la red epistática y que interactúan con <i>glgB</i> y <i>oppA</i> .....	49

## ÍNDICE DE FIGURAS

Figura 1. Infección por <i>M. tb</i> . Imagen obtenida de Pai Madhukar y colaboradores, 2016.....	10
Figura 2. Esquema propuesto de la ruta evolutiva del bacilo tuberculoso. ....	18
Figura 3. Enfoque bioinformático para el estudio de interacciones epistáticas en <i>M. tb</i> . ....	29
Figura 4. Distribución geográfica de 234 cepas del conjunto de datos.....	30
Figura 5. Imagen ilustrativa del pan-genoma completo alineado.....	32
Figura 6. Árbol filogenético del pan-genoma completo de <i>M. tb</i> visualizado por iTOL. ....	38
Figura 7. Red epistática creada con las interacciones con valores atípicos de MI.....	42
Figura 8. Enriquecimiento de términos GO de valores atípicos. ....	43
Figura 9. Enriquecimiento de términos de los genes que interactúan con <i>glgB</i> y <i>oppA</i> .....	47
Figura 10. Distribución alélica de loci de genes que interactúan con <i>glgB</i> . ....	48
Figura 11. Distribución alélica de loci de genes que interactúan con el gen <i>oppA</i> .....	48
Figura 12. Interacciones epistáticas entre los genes de resistencia a los antibióticos <i>fprA</i> y <i>embC</i> , y los genes blancos tentativos de co-selección <i>glgB</i> y <i>oppA</i> .....	49

## ABREVIATURAS

ADN	Ácido desoxirribonucleico
AIC	<i>Akaike information criterion</i>
AMG	Aminoglucósido
AMK	Amikacina
ARN	Ácido ribonucleico
BCG	Bacilo de Calmette-Guérin
BIC	<i>Bayesian information criterion</i>
BQD	Bedaquilina
CARD	<i>Comprehensive Antibiotic Resistance DataBase</i>
CAS	<i>Central Asian strain</i>
CM	Capreomicina
CRISPR	<i>Clustered regularly interspaced short palindromic repeats</i>
DLM	Delamanida
DOTS	<i>Directly observed treatment, short-course</i>
EAI	<i>Indo-Oceanic</i>
EDM	Etionamida
ETB	Etambutol
ESAT-6	<i>Early secreted antigen 6 kilodaltons</i>
ESX-1	<i>Secretion system 1</i>
FLQ	Fluoroquinolona
FO	<i>Maximum likelihood estimation of the base frequencies</i>
F <sub>γ</sub> R	Receptor para Fc de inmunoglobulinas G
G	<i>Gamma model rate heterogeneity</i>
GAR	<i>Genotypic antibiotic resistance</i>
GTR	<i>General Time Reversible model</i>
IFN- $\gamma$	Interferón gama
IL	Interleucinas
INH	Isoniazida
IS	<i>Insertion sequences</i>
JC	Jukes Cantor
KM	Kanamicina
LAM	<i>Lipoarabinomannan</i>
LB	Linfocitos B
LD	<i>Linkage disequilibrium</i>
LT	Linfocitos T
LV	Levofloxacina
M1	Macrófagos inflamatorios
M2	Macrófagos antiinflamatorios
<i>M. bovis</i>	<i>Mycobacterium bovis</i>
MDR	<i>Multidrug-resistant</i>
M $\Phi$	Macrófago

MΦA	Macrófago alveolar
MF	Moxifloxacina
MI	<i>Mutual information</i>
<i>M. tb</i>	<i>Mycobacterium tuberculosis</i>
MTBC	Complejo <i>Mycobacterium tuberculosis</i>
MRE	<i>Majority rules extended</i>
OMS	Organización Mundial de la Salud
PAMS	<i>Pathogen-associated molecular pattern</i>
PAS	P-Ácido aminosalicílico
PZA	Pirazinamida
pb	Pares de bases
R4	<i>Four free rates</i>
RIF	Rifampicina
S	Streptomycin
SNP	<i>Single nucleotide polymorphism</i>
TB	Tuberculosis
TDR	<i>Totally drug-resistant</i>
TGF-β	Factor de crecimiento transformante beta
Th1	Célula T helper tipo 1
TNF-α	Factor de necrosis tumoral alfa
VIH	Virus de inmunodeficiencia humana
VNTRs	<i>Variable number tandem repeats</i>
WHO	<i>World Health Organization</i>
WGS	<i>Whole genome sequencing</i>
XDR	<i>Extensively drug-resistant</i>

## RESUMEN

*Mycobacterium tuberculosis* (*M. tb*) es una bacteria que causa la enfermedad de tuberculosis (TB) en humanos alrededor del mundo. La identificación de las interacciones epistáticas entre diferentes loci bajo presión selectiva en *M. tb* podría ser crucial para la comprensión de la enfermedad y las bases de la adquisición de resistencia a los antibióticos.

El objetivo de este proyecto es predecir bioinformáticamente redes de interacciones epistáticas en *M. tb* a partir del análisis de genomas complejos, para identificar genes involucrados en diferentes rasgos biológicos.

En este proyecto, se analizaron diferentes interacciones epistáticas aplicando para su detección SpydrPick, una aproximación libre de modelo (*model-free*), usando un alineamiento de pan-genoma completo creado con un conjunto de 254 genomas de referencia de *M. tb*. Para el estudio de las interacciones epistáticas, se construyó una red en la que se caracterizó la presencia de dos genes altamente conectados, el gen *glgB*, que codifica para la enzima ramificadora de  $\alpha$ -1,4-glucano, y el gen *oppA*, que codifica para una proteína de unión a oligopéptidos. Ambos se describen en este trabajo como blancos tentativos de co-selección en *M. tb* ya que fueron asociados en la red con genes relevantes de virulencia, patogénesis, moduladores del sistema de transporte, de la respuesta inmune y de resistencia a antibióticos.

También se describen aplicaciones farmacológicas en la resistencia genotípica a los antibióticos inherente a las posiciones de *glgB* y *oppA* en interacción epistática con *fprA* y *embC*, dos genes recientemente incluidos como genes de resistencia en el catálogo de la Organización Mundial de la Salud. Así, en esta investigación se presenta la identificación de interacciones epistáticas en genes relevantes de co-selección para mejorar la comprensión de *M. tb* y esta estrategia podría ser aplicada a diferentes poblaciones de bacterias.



## ABSTRACT

*Mycobacterium tuberculosis* (*M. tb*) is a bacterium that causes tuberculosis (TB) disease in humans worldwide. Identification of epistatic interactions among different *loci* under selective pressure in *M. tb* may be crucial for understanding the disease and the molecular basis of antibiotic resistance acquisition.

The aim of this project is to predict bioinformatically networks of epistatic interactions in *M. tb*, through the analysis of complete reference genomes, to identify genes involved in different biological traits.

In this project, different epistatic interactions were analyzed by applying a model-free approach, SpydrPick, using a complete pan-genome alignment created from a set of 254 *M. tb* reference genomes. For the study of the epistatic interactions, a network was built in which the presence of two highly connected genes was identified, the *glgB* gene, which encodes for the  $\alpha$ 1,4-glucan branching enzyme, and the *oppA* gene, which encodes for an oligopeptide-binding protein. Both genes are described in this work as tentative targets of co-selection in *M. tb* since they were associated in the network with relevant genes of virulence, pathogenesis, transport system modulators of the immune response and antibiotic resistance.

Furthermore, this project describes potential pharmacological applications related to the genotypic antibiotic resistance inherent to mutations of *glgB* and *oppA* as they epistatically interact with *fprA* and *embC*, two genes recently included as antibiotic-resistant genes in the catalog of the World Health Organization. Thus, in this project we present the identification of epistatic interactions in relevant genes under co-selection, which may contribute to improve the understanding of *M. tb*. In addition, this strategy has the potential to be applied to different bacterial populations.

# CAPÍTULO 1. INTRODUCCIÓN

## 1.1. Tuberculosis en el mundo antiguo

La TB es una enfermedad que ha estado presente a lo largo de la historia humana y se le ha descrito de diferentes formas en papiros, pinturas, óperas y literatura (Campbell et al., 2006; Cartes Parra, 2013; Paneque et al., 2018). Además, se le ha denominado de maneras distintas en las diferentes culturas, por ejemplo, fue llamada: “phythisis” (por la cultura griega), “tabes” (romana), “sosha” (india), “schachepheh” (hebrea) y “chakyoncay” (inca), todas haciendo referencia a los términos “secar”, “consumir” o “descomponerse” debido al aspecto debilitado y caquéctico<sup>1</sup> de los afectados (CDC, 2021).

El término “tisis” es mencionado por primera vez en la literatura griega, alrededor del 406 a.C., por Hipócrates. Posteriormente, el aumento de la enfermedad durante la Edad Media y Renacimiento hizo que la enfermedad se extendiera a nivel mundial a partir del siglo XVIII (Augusto Salomone, 2019). En este siglo, la TB era denominada “la plaga blanca” debido a la palidez de los pacientes y en el siglo XIX comúnmente se llamaba “consunción”; hasta que Johann Schönlein, profesor de medicina alemán, en 1834, fue quien unificó todas las denominaciones de la enfermedad y la llamó **tuberculosis** (Balakrishnan, 2017).

La TB ha sido una enfermedad muy presente en todas las sociedades, por lo que diversas figuras de la cultura han plasmado el padecimiento en sus obras. Tal es el caso de la ópera “La Traviata” de Giuseppe Fortunino Francesco Verdi, cuya protagonista, Violetta Valéry, en el acto III fallece a consecuencia de la tisis. Por otro lado, en la literatura varios protagonistas padecen la tisis, como Margarita Gautier, en “La dama de las camelias” de Alejandro Dumas. También en la novela de “Los miserables” de Víctor Hugo, Fantine muere de TB. Por otra parte, en la película inspirada en la novela de John le Carré “El jardinero fiel”, se desarrolla una trama alrededor de las pruebas para un fármaco antituberculoso realizado por una farmacéutica en África y muestra el tema de una posible pandemia mundial de TB debida a la aparición de cepas

---

<sup>1</sup> Estado de salud general deteriorado caracterizado por la pérdida de peso y pérdida muscular.

resistentes a los tratamientos antibióticos conocidos hasta el momento (Augusto Salomone, 2019).

La antigua asociación de la TB con la humanidad ha sido apoyada por datos arqueológicos. La evidencia fósil más antigua hasta ahora de TB en humanos proviene de los restos de 9, 000 años de antigüedad que pertenecieron a una mujer y un bebé que fueron recuperados de un yacimiento en el Mediterráneo Oriental. Los datos arqueológicos sugieren que la TB surgió como una zoonosis<sup>2</sup> y su aparición podría ser anterior a la domesticación de animales (Galagan, 2014).

Se ha postulado que la TB se originó hace millones de años y que pudo haber infectado a los primeros homínidos en África Oriental aproximadamente 3 millones de años atrás (Fatima et al., 2020). Por otro lado, en el estudio descrito por Rothschild y colaboradores en el 2001, se descubrió la existencia de TB en el Pleistoceno tardío mucho antes de la domesticación del ganado y mencionan que la enfermedad estaba muy diseminada en ese periodo sugiriendo teorías de su llegada al Norte de América por lo menos 17, 000 años antes del año presente. En ese estudio también describen que existe la posibilidad de que los bóvidos que emigraron a América del Norte a través del estrecho de Bering podrían haber sido los posibles vectores para la dispersión de la conocida plaga blanca (Rothschild et al., 2001). En América del Sur, los primeros casos de TB registrados en restos humanos se encuentran en Perú y el norte de Chile y datan de los 700 años de nuestra era, casos de TB en América del Norte comienzan a registrarse después del 900 de nuestra era (Vågane et al., 2022).

Como se mencionó anteriormente, la evidencia de TB en huesos revela su presencia en sociedades pasadas; sin embargo, debe recordarse que identificarla es complicado debido a distintos factores como el ambiente, que puede no ser propicio para la preservación de los restos humanos durante largos períodos de tiempo, así como diferentes ceremonias que implican el entierro de restos humanos en ambientes acuáticos o el proceso de cremación (Plomp et al., 2022).

---

<sup>2</sup> Transmisión de una enfermedad de animales a humanos.

La TB es una enfermedad infecciosa mortal que ha estado presente y ha persistido a lo largo de la historia de la humanidad. Actualmente existen indicios de la presencia de esta enfermedad en animales actuando como reservorios antes de la presencia del hombre actual. Su importancia no sólo se limita a definir una enfermedad de importancia clínica, los mecanismos de la enfermedad pueden ayudarnos a comprender la evolución biológica de los organismos presentes en los procesos de adaptación de la tierra y las enfermedades asociadas a este proceso evolutivo.

## 1.2. La enfermedad de tuberculosis

La enfermedad infecciosa de TB es causada principalmente por el bacilo *M. tb* y fue identificada por Robert Koch en 1882. Éste es un patógeno que afecta generalmente pulmones, aunque puede causar enfermedad en cualquier parte del cuerpo en humanos, por ejemplo, riñones, columna vertebral y cerebro (CDC, 2014; WHO, 2020).

La infección por *M. tb* se transmite a través del aire cuando una persona infectada tose, estornuda, habla o escupe y las gotas de saliva que contienen bacilos tuberculosos se dispersan en el aire y pueden ser inhalados por una persona cercana. La TB presenta un espectro clínico dinámico, desde una infección asintomática hasta una enfermedad potencialmente mortal, que generalmente se clasifica en TB latente<sup>3</sup> y TB activa que es transmisible (Pai et al., 2016).

Los pacientes con TB activa pueden presentar los siguientes síntomas: fiebre, cansancio, falta de apetito y pérdida de peso, y en una enfermedad avanzada los pacientes presentan tos persistente y hemoptisis<sup>4</sup> (WHO, 2021). Para describir la infección por *M. tb* en el cuerpo humano, se presentan enseguida las respuestas inmunológicas y algunos factores involucrados en la enfermedad.

---

<sup>3</sup> Estado asintomático no transmisible.

<sup>4</sup> Tos con sangre.

### 1.3. Inmunología en la infección por *Mycobacterium tuberculosis*

La vía de transmisión de *M. tb* es a través del tracto respiratorio por inhalación. Se reporta que se desplaza hacia la parte inferior del tracto respiratorio, donde existe una afinidad celular (tropismo) por los macrófagos alveolares (MΦAs) (Figura 1). La bacteria ingresa a la célula por fagocitosis<sup>5</sup> mediada por diferentes receptores como el receptor para Fc de inmunoglobulinas G (FcγR) y el receptor de complemento CR3, que reconocen a *M. tb* facilitando la fagocitosis (Allard et al., 2018).

Los macrófagos (MΦs) son células altamente dinámicas del sistema inmunológico que responden diferencialmente durante la infección, la activación y polarización en macrófagos inflamatorios (M1), que secretan citocinas<sup>6</sup> inflamatorias (IL1-β, TNF-α, Factor de necrosis tumoral alfa, e IL-12) o anti-inflamatorios (M2), que se clasifican por la expresión de moléculas de membrana CD206, el receptor de manosa de MΦ y la secreción de interleucina<sup>7</sup> IL-10, que participa en la reparación de tejidos. Esta polarización inflamatoria y reparadora de los tejidos se ha descrito que podría tener aplicaciones importantes en terapias avanzadas dirigidas al hospedero, pero debido a la complejidad de la respuesta desarrollada por los MΦs ante la infección activa por *M. tb*, es necesario aumentar los estudios de las estructuras celulares presentes en la respuesta contra el bacilo en los pulmones (Arish y Naz, 2022).

Una vez dentro del MΦ, *M. tb* bloquea activamente la fusión del fagosoma con los lisosomas, asegurando su supervivencia. Posteriormente, la actividad del sistema de secreción ESX-1<sup>8</sup> de *M. tb* puede interrumpir la membrana del fagosoma provocando la liberación de productos bacterianos, incluyendo el ADN (ácido desoxirribonucleico) de las micobacterias dentro del citosol en el MΦ. Por otro lado, la muerte celular, que experimentan los MΦs infectados es importante para continuar con la respuesta ante el agente infeccioso y así desarrollar inmunidad adaptativa (Arish y Naz, 2022).

---

<sup>5</sup> Internalización de materia particulada por las células.

<sup>6</sup> Proteínas secretadas por el sistema inmune.

<sup>7</sup> Interleucinas (IL), proteínas de bajo peso molecular.

<sup>8</sup> Sistema de secreción tipo 1 (*secretion system 1*).

Después de infectar los MΦAs en las vías respiratorias, *M. tb* accede al intersticio pulmonar. Aún se desconoce cómo la micobacteria accede al parénquima pulmonar (tejido esencial de los pulmones), por lo que se han descrito dos posibles mecanismos: uno que infecta directamente las células epiteliales y el segundo por transmigración de *M. tb* a los MΦs a través del epitelio (Figura 1a). Una vez que *M. tb* se localiza en el parénquima, conduce al reclutamiento de un mayor número de células al sitio de la infección, derivada de la capacidad de *M. tb* para evadir la fagocitosis, generando la formación de una estructura multicelular llamada granuloma (Pai et al., 2016). Un granuloma es un conjunto de MΦs que incluye células gigantes multinucleadas, células epiteliales, linfocitos, neutrófilos y células espumosas.

Desde la perspectiva del hospedero, el granuloma es una estructura de confinamiento bacteriano con el potencial de aislar al patógeno del resto del cuerpo y evitar un daño inflamatorio orgánico; sin embargo, desde la perspectiva bacteriana, constituye un sitio para permanecer en estado de latencia por tiempos prolongados y en un nicho con células fagocíticas para infectar y replicarse (Ge et al., 2021).

Al interior del granuloma se genera un ambiente pro-inflamatorio con la producción de citocinas como TNF- $\alpha$ , IL-1B e IL-6 producidas por MΦs, o interferón gamma (IFN- $\gamma$ ) producido principalmente por linfocitos T (LT); mientras que el tejido circundante adquiere propiedades anti-inflamatorias debidas a citocinas como el factor de crecimiento transformante beta (TGF- $\beta$ ) o la IL-10. En la parte periférica del granuloma se localizan linfocitos generalmente de tipo CD4 T helper tipo 1 (Th1), que sintetizan también citocinas pro-inflamatorias como IFN- $\gamma$  o IL-12. También se observan LT CD8<sup>+</sup> y, en menor proporción, linfocitos B (LB). Esta conformación origina que el granuloma se rodee de una cápsula de tejido fibroso para contener la diseminación (Bini et al., 2015).

*M. tb* genera diferentes proteínas de secreción como ESAT-6<sup>9</sup> y CFP-10<sup>10</sup> liberadas a través de ESX-1, ambas implicadas en diferentes mecanismos de virulencia. Se ha reportado que la proteína de secreción ESAT-6 induce la producción de IFN- $\alpha$ , que contribuye al proceso

---

<sup>9</sup> Early secretory antigen 6 kilodaltons

<sup>10</sup> Culture filtrate protein 10

de inflamación; por otra parte, la proteína chaperona<sup>11</sup> CFP-10 puede unirse a la superficie de los MΦs y estimular la secreción de citocinas inflamatorias como TNF- $\alpha$  (Conrad et al., 2017).

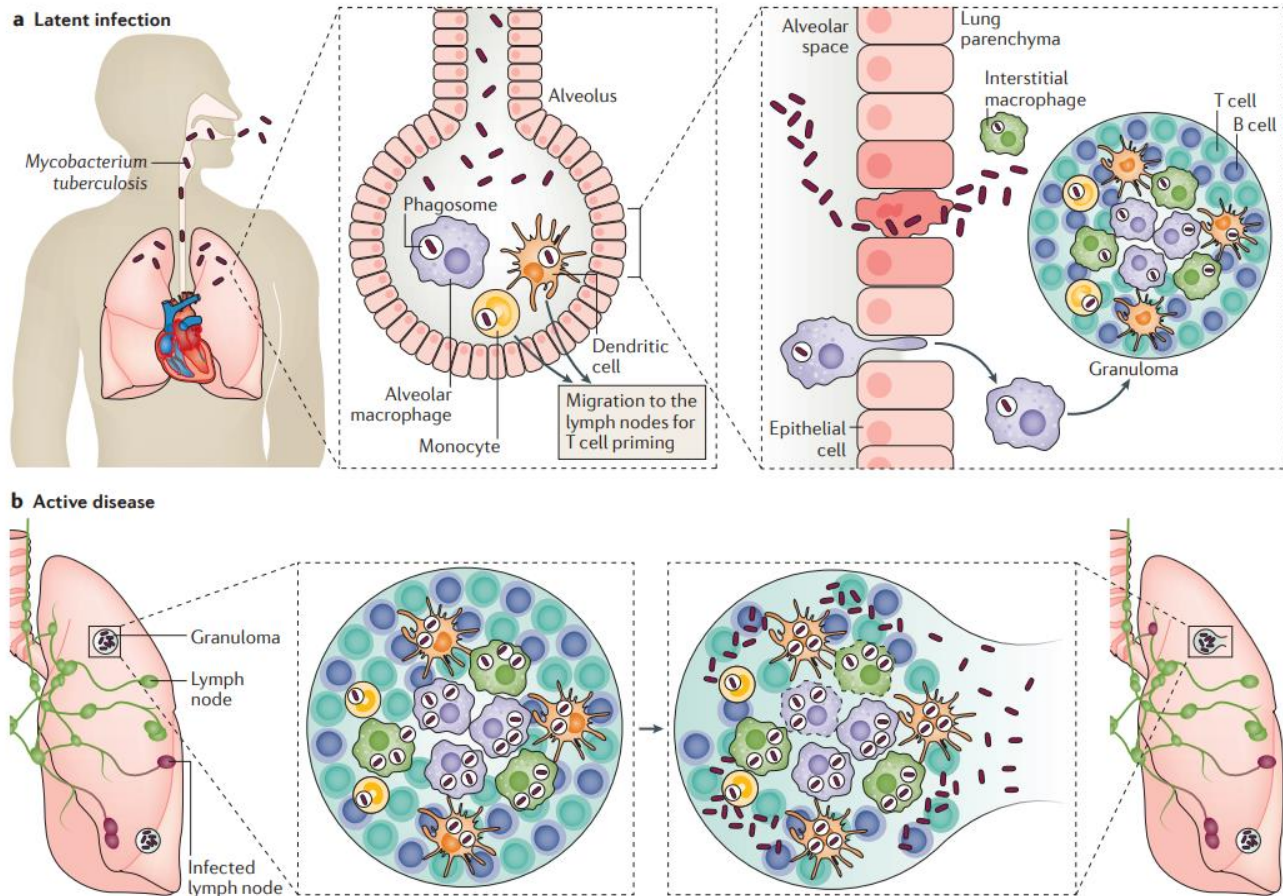


Figura 1. Infección por *M. tb*. Imagen obtenida de Pai Madhukar y colaboradores, 2016.

a) La infección cuando *M. tb* ingresa a los pulmones por inhalación donde se dirige al espacio alveolar y es fagocitada por los MΦAs. Esta primera línea de defensa no logra eliminar la bacteria, por lo que *M. tb* invade el tejido intersticial pulmonar, ya sea por las bacterias que infectan directamente el epitelio alveolar o los MΦAs infectados que migran al parénquima pulmonar. Posteriormente, las células dendríticas o los MΦs transportan a *M. tb* a los ganglios linfáticos pulmonares para la presentación antigénica de los LTs, este evento conduce al reclutamiento de células inmunes, incluyendo LTs y LBs, al parénquima pulmonar para formar un granuloma.

b) Las bacterias se replican dentro del granuloma. Si la carga bacteriana aumenta, el granuloma no podrá contener la infección y las bacterias se diseminarán eventualmente a otros órganos incluido el cerebro. En esta fase, las bacterias pueden ingresar al torrente sanguíneo o volver a ingresar al tracto respiratorio: el hospedero infectado ahora es infeccioso, sintomático, por lo que la enfermedad se convierte en TB activa.

<sup>11</sup> Conjunto de proteínas cuya función es la de ayudar al plegamiento, ensamblaje y transporte de otras proteínas.

En la mayoría de las personas con TB latente, la respuesta de los MΦs desencadena la respuesta inmune innata de otras células como las células dendríticas y la respuesta adaptativa de los LTs (Figura 1a). La respuesta inflamatoria de estas células es suficiente para mantener una infección asintomática controlada. Sin embargo, por razones que aún no están completamente claras, la infección puede progresar en semanas hasta décadas a enfermedad activa. Se han descrito algunos factores epidemiológicos, desde el punto de vista del hospedero, que se han asociado con el riesgo de TB activa, los cuales son: infección por VIH (virus de inmunodeficiencia humana), anticuerpos neutralizantes anti-TNF- $\alpha$  y alteraciones congénitas en moléculas de la inmunidad. Estudiar estos factores puede ser crucial para comprender y controlar los mecanismos de infección de TB activa (Pai et al., 2016).

Para describir la susceptibilidad genética, se han desarrollado diversos estudios para analizar genes asociados a TB activa, así, se ha reportado que de 100 millones de niños vacunados con BCG (bacilo de Calmette-Guérin) sólo un pequeño número desarrolla la enfermedad diseminada de BCG. En ellos ha sido posible identificar mutaciones en genes (IFN- $\gamma$ , IL12) que codifican proteínas que son cruciales para esta fase de la enfermedad (Bustamante et al., 2014).

La progresión hacia formas activas de TB puede estar asociada a coinfecciones, ya que se ha descrito que la sindemia<sup>12</sup> de VIH-TB es un factor de riesgo para la progresión de la enfermedad activa en un individuo, con un aumento en la incidencia de decesos por TB activa en pacientes con VIH, en tanto que en caso sin coinfección, las proteínas antigénicas conformacionales y los antígenos procesados presentados a los LT CD4+ inmunodominantes no varían entre cepas y linajes (Comas et al., 2010; Corbett E., 2006).

La importancia de los antígenos inmunodominantes de *M. tb* queda de manifiesto en la efectividad de la vacuna BCG para proteger de las formas graves de la infección mediante la inmunización con el bacilo completo atenuado (*M. bovis*) (Abel et al., 2014). Después de casi un siglo de la vacuna BCG se conoce que la protección se genera en gran medida por inmunidad celular (LT CD4+ y LT CD8+) y en menor medida por inmunidad humoral (LB y anticuerpos),

---

<sup>12</sup> Dos o más epidemias que comparten factores sociales.



por lo que actualmente hay varios candidatos a vacunas con diversos grados de protección, aunque BCG sigue siendo la única aplicada (Tameris et al., 2013).

La exposición a *M. tb* conduce a tres panoramas: 1) Eliminación del patógeno, 2) Persistencia del patógeno (TB latente a TB activa) 3) Enfermedad. Aún se desconocen muchos mecanismos de respuesta inmune y la identificación de los factores desde la perspectiva de la respuesta inmune en el humano. Estudios dirigidos a la bacteria podrían ayudarnos a la comprensión de la TB por lo que es necesario continuar investigando dichos temas.

Una vez descritas las respuestas inmunológicas generales de la infección por TB, a continuación se presentan los datos epidemiológicos reportados por la Organización Mundial de la Salud (OMS)<sup>13</sup>.

#### **1.4. Epidemiología de la Tuberculosis**

La OMS reporta que una cuarta parte de la población mundial está infectada de TB latente. Se calcula que en 2020 se enfermaron 9.9 millones de personas en todo el mundo, con un estimado de 1.5 millones de muertes, de las cuales 214, 000 tenían sindemia con VIH. Del total de 9.9 millones de personas infectadas, 5.5 millones representan infecciones en hombres, 3.3 millones en mujeres y 1.1 millones en niños (WHO, 2020).

Las personas infectadas por VIH tienen 18 veces más probabilidad de desarrollar TB activa (con un intervalo de confianza de 15-21), mientras que las personas con desnutrición presentan un riesgo tres veces mayor. Otros factores asociados con la enfermedad son el alto consumo de alcohol y tabaquismo, que aumentan el riesgo de contraer TB en un factor de riesgo de 3.3 y 1.6, respectivamente (WHO, 2020).

En México, la OMS reporta que la incidencia en el año 2020 por TB fue de 31, 000 [23, 000 – 39, 000]<sup>14</sup> y 24 [18 – 30] en tasa por 100, 000 habitantes, de los cuales 3, 900 [3, 000 - 5, 000] son sindemia con VIH con tasa de 3 [2.3 – 3.9] por 100, 000 habitantes. También se reporta

---

<sup>13</sup> En inglés *World Health Organization*, WHO.

<sup>14</sup> Los [] indican intervalo de valores.

para el país que la tasa de cobertura de tratamiento es de 57% [45% - 75%] y en el año 2020 se notificaron 18, 339 casos, de los cuales el 61% fueron hombres, 36% mujeres y 3% niños (WHO, 2020).

Uno de los objetivos de Desarrollo Sostenible para 2035 de la OMS consiste en ponerle fin a la pandemia por TB. La estrategia de la OMS plantea reducir las muertes en un 90% y la incidencia de la enfermedad en un 80% para 2030 en comparación con las cifras de 2015. Sin embargo, la pandemia de COVID-19 ha revertido los avances reportados por la OMS en la última década en la lucha contra TB debido a fallas en los servicios esenciales de prevención y atención (WHO, 2021). En este momento, otro tema de interés sobre la TB es la resistencia a los antibióticos, por lo que en el siguiente subcapítulo procedo a describir en qué consiste el tratamiento.

## **1.5. Tratamiento de la Tuberculosis**

El tratamiento estándar comprende la combinación de antimicrobianos clasificados como fármacos de primera línea, los cuales son: isoniazida (INH), rifampicina (RIF), pirazinamida (PZA) y etambutol (ETB), administrados durante los primeros dos meses de la fase intensiva, seguidos por la fase de continuación administrando por cuatro meses INH y RIF. Este tratamiento puede rápidamente disminuir las bacterias en los pulmones de una persona para reducir la transmisión y así terminar con el proceso de infección (Tornheim y Dooley, 2019).

La resistencia a estos antimicrobianos de primera línea puede ocurrir, y cuando las cepas presentan resistencia a INH y RIF son definidas como multirresistentes (*Multidrug-resistant, MDR*); actualmente la persistencia de cepas MDR se ha notificado en todos los países. Por otro lado, existe *M. tb* extremadamente resistente a los medicamentos, que se clasifica así por su resistencia no sólo a los fármacos de primera línea, sino también a fluoroquinolona (FLQ) y cualquiera de los tres aminoglucósidos (AMG) inyectables de segunda línea: capreomicina (CM), kanamicina (KM) y amikacina (AMK). A estas cepas se les define como cepas extremadamente resistentes (*Extensibely drug-resistant, XDR*). Las opciones terapéuticas varían para la TB latente y activa, así como para la enfermedad sensible y resistente a los medicamentos (Seung et al., 2015).

Como parte de la estrategia mundial para el control de la TB denominada tratamiento breve bajo observación directa (*directly observed treatment, short-course, DOTS*), estos regímenes se convirtieron en el estándar de atención incluso en entornos con recursos limitados a partir de 1993 (Frieden T et al., 1993). Cuando la terapia se toma correctamente, tiene una tasa de curación del 85%. Sin embargo, este largo tratamiento es a menudo asociado con efectos secundarios significativos como el riesgo de toxicidad hepática, reacciones de hipersensibilidad y neuropatía periférica en el paciente, además de que se complica con interacciones de otros fármacos administrados, por ejemplo, con agentes antirretrovirales para tratar el VIH (Zhang, 2007). En la Tabla 1, se describen los efectos adversos principales de los fármacos contra TB utilizados en la terapia estándar. Más adelante se mencionarán los mecanismos de los fármacos en *M. tb*.

Tabla 1. Principales fármacos contra Tuberculosis y sus efectos adversos.

Se observan los efectos adversos principales farmacológicos de los antimicrobianos contra *M. tb*. Información obtenida de (Pascual-Pareja et al., 2018).

Fármaco	Clasificación	Vía de administración	Efectos adversos relevantes
Isoniazida (INH)	Primera línea	Oral	Hepatitis. Neuritis periférica.
Rifampicina (RIF)			Hepatitis. Fiebre. Anemia hemolítica. Trombocitopenia. Insuficiencia renal. Reacciones de hipersensibilidad.
Pirazinamida (PZA)			Hepatitis. Hiperuricemia.
Etambutol (ETB)			Neuritis óptica.
Estreptomina (S)	Segunda línea	Intravenosa Intramuscular	Ototoxicidad. Nefrotoxicidad.
Capreomicina (CM)			
Amikacina (AMK)			
Kanamicina (KM)			
Etionamida (EDM)		Oral	Náuseas y vómitos. Hepatitis. Hipotiroidismo. Sabor metálico.
Ácido Para - amino salicílico (PAS)			Gastroenteritis. Hepatitis.
Levofloxacina (LV)			Oral Intravenosa
Moxifloxacina (MF)	Oral		

### 1.5.1. Resistencia a los antibióticos en Tuberculosis

La resistencia a los medicamentos es un fenómeno biológico que se ha observado en *M. tb* desde el descubrimiento del primer fármaco antimicrobiano: la estreptomicina (S). Los pacientes a los que se les inyectó el antibiótico pronto comenzaron a presentar bacilos resistentes a la S en el laboratorio (Pyle, 1947). Unos años después, se encontró que el uso combinado de S y el fármaco posteriormente descubierto en 1948, el ácido para-amino salicílico (PAS), logró mejores resultados en el tratamiento de la TB.

El descubrimiento de la INH, en 1952, resultó en una terapia con tres fármacos (INH, S y PAS) que podían curar la TB en las décadas de 1950 y 1960. Sin embargo, esa terapia antituberculosa requería una duración de 18 meses. El descubrimiento de la PZA (1952), ETB (1962) y RIF (1963) sentó las bases para los regímenes de medicamentos del tratamiento actual. En las siguientes décadas, existió desinterés por continuar los incentivos de la industria farmacéutica para generar nuevos fármacos antituberculosos. Este escenario limitó los ingresos potenciales para apoyar los gastos de investigación; sin embargo, con el aumento de las cepas de resistencia, en los últimos años se ha incrementado la acción para el desarrollo de fármacos antituberculosos (Yew et al., 2011).

La aparición de cepas MDR y XDR ha amenazado cada vez más el control global de la TB y para empeorar los escenarios de resistencia a los medicamentos, Migliori y colaboradores (2012) informaron sobre cepas totalmente resistentes a los medicamentos desde el 2012 (*Totally drug-resistant*, TDR).

La resistencia farmacológica en *M. tb* se encuentra asociada a la acumulación de mutaciones genéticas en el bacilo, que se generan de manera espontánea dentro del cromosoma bacteriano (este tema se describe en el siguiente capítulo). Sin embargo, existen factores que contribuyen como: el suministro irregular de medicamentos, indicaciones inapropiadas en la prescripción médica y pobre adherencia al tratamiento<sup>15</sup> por parte del paciente. Estas han contribuido con el desarrollo de TB resistente a los medicamentos, y la

---

<sup>15</sup> Falla en el cumplimiento de las instrucciones terapéuticas.

posterior transmisión de cepas resistentes del paciente a otros miembros de la comunidad termina por profundizar el problema (Yew et al., 2011).

La secuenciación de genoma completo es actualmente una tecnología avanzada para descubrir los mecanismos de resistencia a los medicamentos en *M. tb*. Esta metodología aumentó la cantidad de secuencias genómicas y su estudio se unió a los esfuerzos por comprender mejor los mecanismos evolutivos en la bacteria; por lo que se requiere más investigación para desentrañar la importancia de algunas mutaciones genéticas asociadas a la resistencia a los medicamentos. El presente proyecto doctoral se enmarca en este tipo de estudios asociados al análisis del genoma completo de *M. tb*. A continuación, se presenta la descripción de los bacilos causantes de TB.

### **1.6. Complejo *Mycobacterium tuberculosis***

La TB es causada principalmente por *M. tb*; no obstante, este bacilo comprende a un grupo de bacterias Gram-positivas estrechamente relacionadas, colectivamente conocido como complejo *Mycobacterium tuberculosis* (MTBC). MTBC comprende los patógenos que infectan a humanos *M. tb* y *Mycobacterium africanum*, así como variantes adaptadas que infectan a diversas especies animales. Estos patógenos incluyen diversos hospederos animales, como *Mycobacterium bovis* (que afecta ganado), *Mycobacterium caprae* (cabras y ovejas), *Mycobacterium microti* (roedores) y *Mycobacterium pinnipedii* (focas y leones marinos). Al contrario de otras bacterias, el MTBC no adquiere islas de patogenicidad, ni produce ninguna toxina; sin embargo, es capaz de infectar, sobrevivir y transmitirse eficientemente entre diferentes hospederos (Comas I & Gagneux S, 2011).

Los métodos más frecuentes para estudiar la dinámica de transmisión de la TB son técnicas basadas en marcadores moleculares reportados con tasas de evolución rápidas, como secuencias de inserción (IS6110 RFLP), repeticiones en tándem de número variable (*variable number tandem repeats*, VNTRs) o espigotipado, método de hibridación de transferencia de línea inversa que usa sondas de captura diseñadas en espaciadores y unidas a la superficie de una membrana unida a un producto de PCR obtenido de repeticiones palindrómicas cortas agrupadas regularmente interespaciadas conocido como CRISPR (*clustered regularly*

*interspaced short palindromic repeats*). Estos métodos tienen una utilidad limitada cuando se aplican a largo plazo para evolución. MTBC exhibe una estructura de población clonal sin evidencia de transferencia horizontal de genes (Lundgren et al., 2015). Los primeros métodos de genotipado aplicados a cuestiones evolutivas en MTBC se basaron en deleciones genómicas, como marcadores filogenéticos para revisar la clasificación de entre las diferentes subespecies que infectan a humanos y animales dentro del MTBC (Brosch et al., 2002).

Actualmente se describen siete linajes adaptados a humanos. Los linajes que comprenden a *M. tb* son del 1 al 4 y el linaje 7. *M. africanum* comprende los linajes 5 y 6. El linaje 2 (conocido como linaje asiático) y linaje 4 (euroamericano) se encuentran distribuidos en todo el mundo. El linaje 5 y el linaje 6 se reportan muy restringidos a la especie *M. africanum*, localizados geográficamente en África Occidental.

El linaje 7 de la especie *M. tb* se distribuye predominantemente en Etiopía. Los linajes restantes muestran una distribución intermedia entre regiones, por ejemplo, el linaje 1 es conocido como Indo-Océánico (*Indo-Oceanic*, EAI) y el linaje 3 (cepa de Asia Central, *Central Asian strain*, CAS) que domina en partes de África Oriental, Centro y sur de Asia (Gagneux, 2017). Los linajes 2, 3 y 4 son denominados “modernos” mientras que el resto de los linajes son llamados “antiguos” y genéticamente se refiere a la eliminación de la región conocida como TbD1, que se encuentra presente en otras especies como *M. africanum*, *M. bovis* y todas las especies adaptadas a animales (Brosch et al., 2002).

Los polimorfismos de nucleótido único (*single nucleotide polymorphism*, SNP) en las poblaciones genéticas de MTBC están presentes en bajas frecuencias y las mutaciones poblacionales se reportan como perjudiciales para MTBC. Mutaciones en frecuencias bajas también pueden explicarse por la expansión de la población y, en el caso de linajes adaptados a humanos del MTBC, tales expansiones pudieron ocurrir junto con las mutaciones de las poblaciones humanas modernas.

En la Figura 2 se muestra el esquema de la ruta evolutiva propuesta por Brosch y colaboradores (2002). MTBC exhibe una alta proporción de cambios no sinónimos a sinónimos

en comparación con otras bacterias, que se ha propuesto que puede justificarse en un efecto de fuerte deriva genética ligada a cuellos de botella durante la transmisión (Gagneux, 2017).

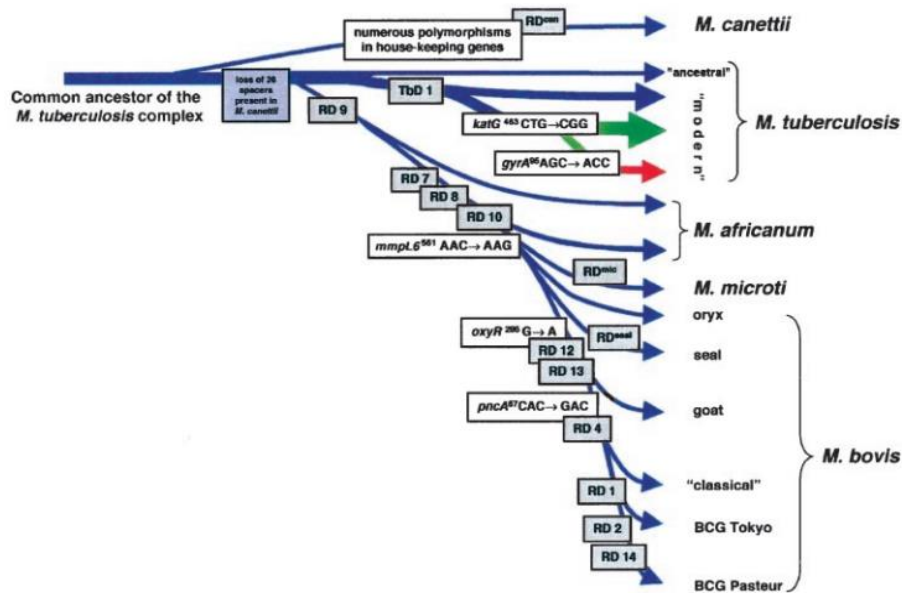


Figura 2. Esquema propuesto de la ruta evolutiva del bacilo tuberculoso.

Imagen obtenida de Brosch y colaboradores, 2002.

En la figura, se ilustra la pérdida sucesiva de ADN en ciertos linajes (recuadros grises). El esquema muestra la presencia o ausencia de regiones conservadas o con deleciones en polimorfismos de secuencia en cinco genes seleccionados. Las distancias entre las ramas no corresponden a las diferencias filogenéticas reales calculadas. Las flechas azules indican que las cepas se caracterizan por los genes *katG*<sup>463</sup> CTG (Leu) y *gyrA*<sup>95</sup> ACC (Thr) típico de los organismos del grupo 1. Las flechas verdes indican que las cepas pertenecen al grupo 2 caracterizadas por *katG*<sup>463</sup> CGG (Arg) y *gyrA*<sup>95</sup> ACC(Thr). La flecha roja indica que las cepas pertenecen al grupo 3, caracterizado por *katG*<sup>463</sup> CGG(Arg) y *gyrA*<sup>95</sup> AGC (Ser).

Lo anterior se une con datos experimentales y estudios fisiológicos en miembros del MTBC que han demostrado que la enfermedad pulmonar en la infección puede ser establecida por cargas bajas como de una a tres bacterias. Este cuello de botella de la infección es una vía a través de la cual la deriva génica da forma a la evolución de *M. tb*, cuyo efecto sería amplificado en cada transmisión. Las poblaciones de *M. tb* también experimentan cuellos de botella posteriores a la transmisión, relacionados con la diseminación dentro del hospedero, así como con el tratamiento farmacológico. Estos cambios a nivel genético probablemente contribuyan a los bajos niveles de diversidad genética y al exceso de variantes de baja frecuencia (Morales-Arce et al., 2021).

La secuenciación de siguiente generación se ha utilizado para resaltar las limitaciones de las técnicas de genotipado para diferenciar entre cepas estrechamente relacionadas y se infiere que la secuenciación se utilizará cada vez más para la epidemiología molecular rutinaria de la TB en el futuro (Comas I & Gagneux S, 2011).

Los avances técnicos en la secuenciación de siguiente generación pueden contribuir a posicionar filogenéticamente a algunos restos paleontológicos de la evidencia molecular más antigua de MTBC, como el encontrado en el bisonte de 17, 000 años (Rothschild et al., 2001). Podemos resumir que el panorama de la diversidad genética del MTBC puede conducir a una estimación del antepasado del mismo a través de los ancestros que causaron la enfermedad de TB en humanos y animales.

El MTBC presenta alta clonalidad<sup>16</sup> y falta de transferencia horizontal de genes<sup>17</sup>, lo que causa un fuerte vínculo entre los sitios en el genoma del MTBC: las variantes nucleotídicas no son independientes entre sí. Mutaciones con fuerte selección confieren resistencia a los antibióticos. Pacientes con TB en tratamiento farmacológico pueden desarrollar nuevas mutaciones de resistencia presentes en el fondo genético de la cepa. Lo anterior se complementará en el siguiente capítulo para una mayor comprensión de los mecanismos de resistencia, pero primero se describen las características generales del genoma de *M. tb*.

## **1.7. *Mycobacterium tuberculosis***

### **1.7.1. Descripción del genoma de *Mycobacterium tuberculosis***

Poco más de 100 años después de que la bacteria fuera aislada por Robert Koch, se reveló que la cepa H37Rv (la cepa de laboratorio más utilizada) tiene un genoma con alto contenido de GC de 4.4 Mb (millones de pares de bases nucleotídicas) con ~4, 000 genes que codifican proteínas y 50 genes que codifican ARN estable (Cole S T et al., 1998).

---

<sup>16</sup> Grupo de individuos de un organismo que se encuentran constituidos de la misma semejanza fenotípica.

<sup>17</sup> El evento de un organismo por el cual se adquiere material genético de otra célula que no es de su ancestro.



Dentro del genoma de *M. tb*, se atribuyen funciones precisas al 40% de los genes que codifican proteínas, 34% de los genes se describe que tienen alguna información funcional, mientras que el 16% restante tienen función desconocida, posiblemente específica de las especies de micobacterias. *M. tb* contiene elementos abundantes en secuencias de inserción (*Insertion sequences*, IS<sup>18</sup>) y contiene 56 copias de ocho familias de IS diferentes (Gordon et al., 1999).

Uno de los hallazgos más notables fue el descubrimiento de que ~9% del genoma codifica dos nuevas familias de proteínas ricas en glicina, las proteínas PE (proteínas ricas en glicina–alanina) y PPE (proteínas ricas en glicina-asparagina). La información funcional sobre estas proteínas aún es muy limitada y se ha sugerido que podrían desempeñar un papel en la variación antigénica. Otra característica del genoma de *M. tb* es la presencia de 250 genes que codifican enzimas implicadas en metabolismo de ácidos grasos, en comparación con sólo 50 en *E. coli*, que contribuyen a la presencia en la envoltura celular de una amplia diversidad de moléculas lipofílicas, además de que *M. tb* puede producir 100 enzimas involucradas en la  $\beta$ -oxidación de ácidos grasos (Domenech Pila, 2001).

El conocimiento de *M. tb* H37Rv permitió conocer la diversidad genómica de los otros miembros del MTBC que hoy en día comparten > 99.95% de identidad a nivel de ADN (Gordon et al., 1999). Además, se describió un conjunto de siete regiones con deleciones en los cromosomas de *M. bovis* y *M. bovis*-BCG. La pérdida de genes ocurre con una alta frecuencia como resultado de la homologación de eventos de recombinación entre copias de IS6110 que flanquean los genes en la orientación directa (Domenech Pila, 2001).

La diversidad genómica de *M. tb* y otros miembros del MTBC a nivel intrahospedero actualmente se puede explorar mejor con la secuenciación masiva, la cual permite la detección de subpoblaciones y la identificación de las variantes de baja frecuencia que pueden indicar la presencia de infecciones mixtas (es decir, coinfección con distintas cepas dentro del MTBC) o microevolución que conduce a subpoblaciones relacionadas con la resistencia a los medicamentos, por lo que se necesitan más investigaciones y mejorar metodologías

---

<sup>18</sup> Elementos genéticos pequeños de longitud < 2.5 kb.

computacionales para comprender e interpretar mejor la diversidad dentro del hospedero (Meehan et al., 2019).

### **1.7.2. Mecanismos de resistencia antimicrobiana en el genoma de *Mycobacterium tuberculosis***

Los estudios de secuenciación de genoma completo (*Whole genome sequencing*, WGS) de *M. tb* han aumentado considerablemente la comprensión del patógeno. La WGS ha contribuido con una descripción más completa de las características genómicas que causan la resistencia a los antibióticos del bacilo y también ha identificado las nuevas mutaciones específicas dentro del genoma que permiten que *M. tb* escape a los efectos de los fármacos (Cohen et al., 2019).

Se estima que *M. tb* tiene una lenta evolución con [0.3 – 0.6] SNP/genoma/año. Por otro lado, los datos experimentales sugieren que la resistencia a los medicamentos puede evolucionar dentro de un paciente individual durante el tratamiento de la TB (Cohen et al., 2015; Walker et al., 2013). La evolución de la resistencia a los medicamentos dentro de un hospedero no siempre es lineal, y puede implicar infecciones complejas entre poblaciones heterogéneas de *M. tb*. La diversidad genética puede existir antes de que surja un clon dominante, lo que se convierte en un desafío para predecir con precisión la heterorresistencia<sup>19</sup> a los medicamentos. La ambigüedad de las muestras que contienen cepas mixtas o la detección de la heterorresistencia siguen siendo un reto computacional (Cohen et al., 2019).

Las técnicas basadas en el cultivo siguen siendo la prueba estándar para el diagnóstico y para las pruebas de susceptibilidad a los medicamentos de la TB, pero estos procesos requieren mucho tiempo y capacidad de laboratorio especializado. Actualmente el uso de pruebas moleculares rápidas para el diagnóstico de TB ha aumentado a nivel mundial (Cohen et al., 2019), particularmente el uso de Xpert MTB/RIF (Cepheid, Sunnyvale, CA). Esta es una prueba basada en la técnica molecular de PCR para detectar la presencia de resistencia de RIF. Generalmente, la mayor parte de la resistencia a los antimicrobianos ocurre para los

---

<sup>19</sup> Definida como la coexistencia de poblaciones de patógenos que tienen diferentes nucleótidos en un *locus* de resistencia a fármacos específicos.

medicamentos de primera línea y se explica por mecanismos y mutaciones que se muestran en la Tabla 2.

Tabla 2. Mecanismos de acción de fármacos y genes de resistencia en el genoma de *Mycobacterium tuberculosis*.

Clasificación de los antibióticos caracterizados por el mecanismo de acción de los fármacos.

Fármaco	Mecanismo de acción del fármaco	Genes de resistencia y función de genes
Rifampicina (RIF)	Inhibición de la transcripción de genes por interacción de ARN polimerasa (ARN polimerasa ADN dependiente)	<i>rpoB</i> Subunidad b de la ARN polimerasa
Isoniazida (INH)	Inhibición de la síntesis de ácidos micólicos	<i>katG</i> Enzima catalasa peroxidasa
		<i>inhA</i> Enoil ACP reductasa
Pirazinamida (PZA)	Inhibe la síntesis en la membrana bacteriana	<i>pncA</i> Enzima Pirazinamidasa
Etambutol (ETB)	Inhibición de síntesis de componentes de la pared celular de micobacterias	Operon <i>embCAB</i> Arabinosiltransferasa
Levofloxacin (LV)	Inhibición de síntesis de ADN y la aparición de superhélices al actuar en la topoisomerasa.	<i>gyrA</i> DNA girasa subunidad A
Moxifloxacin (MF)		<i>gyrB</i> DNA girasa subunidad B
Amikacina (AMK)	Inhibición de la síntesis de proteínas (Metilación del 16S ARN)	<i>Rrs</i> 16S rARN (ARN ribosomal 16S)
Capreomicina (CM)		<i>eis</i> N-Acetil-transferasa
Kanamicina (KM)		
Ácido Para-aminosalicílico (PAS)	Inhibición del ácido fólico	<i>folC</i> Dihidrofolato sintetasa
		<i>folA</i> Dihidrofolato reductasa

Varios estudios se han comprometido a caracterizar mutaciones asociadas a resistencia en el genoma de *M. tb*. Un ejemplo de investigación asociada es la que se realizó en México por Juárez-Eusebio y colaboradores en el 2017, que reportó el crecimiento de casos de cepas MDR en los estados de Nuevo León, Tamaulipas, Guerrero, Veracruz y Baja California donde se encontraron mutaciones asociadas a los genes *rpoB*, *katG*, *inhA*, *rrs*, *eis*, *gyrA*, *gyrB* y *tlyA*. En el estudio se reporta que se observaron discrepancias entre las pruebas fenotípicas y los

polimorfismos y así establecen la necesidad de implementar un sistema de vigilancia epidemiológica molecular en el país (Juarez-Eusebio et al., 2017).

Desde las primeras aplicaciones del WGS a *M. tb* en 1998, estas técnicas han acelerado en gran medida la comprensión de los mecanismos de resistencia a los fármacos antituberculosos. La creciente disponibilidad de secuencias de genoma completo de *M. tb* han proporcionado información valiosa sobre la evolución de mutaciones de resistencia. La adquisición de nuevas mutaciones de resistencia como la transmisión clonal son factores críticos en la propagación de *M. tb* resistente a los antibióticos. Recientemente el uso de WGS fue aplicado para identificar la alta prevalencia del linaje 4 (94%) con una presencia de sub linajes fuertemente asociados a cepas MDR. En este estudio de Jiménez-Ruano y colaboradores también se caracterizó a los linajes 1 y 2 en menor porcentaje con una representatividad de 3% cada uno, los investigadores comentan de la necesidad de aumentar el número de estudios relacionado con el análisis de WGS de *M. tb* en México (Jiménez-Ruano et al., 2021).

Es crucial detectar y describir las mutaciones asociadas a resistencia en *M. tb*; sin embargo, se ha explicado que son sólo una parte del problema. Conocer los mecanismos evolutivos que impulsan la resistencia a los medicamentos puede mejorar las herramientas de diagnóstico, así como mejorar estrategias para preservar los nuevos protocolos de tratamiento.

El camino evolutivo que conduce a la resistencia a los medicamentos es fuertemente influenciado por dos factores: la epistasis (interacciones entre genes) y la bacteria (Trauner et al., 2014). Por lo que es necesario que continúen investigaciones centradas en las interacciones entre genes que describan asociaciones vinculadas a resistencia. Así, el presente proyecto de investigación es un esfuerzo por describir interacciones epistáticas en *M. tb*. En el siguiente subcapítulo se describe que es la epistasis.

## 1.8. Epistasis

El término “epistasis” fue descrito por William Bateson en 1906 para referir a la interacción entre genes independiente a los rasgos complejos. En el trabajo de Bateson, se explica la idea de que un fenotipo es caracterizado por un genotipo dependiente de genotipos presentes en dos o más *loci*<sup>20</sup> en el genoma (Sackton y Hartl, 2016). Bateson aplicó estudios de patrones de herencia en los rasgos categóricos descritos por Mendel, este tipo de epistasis aplicada por Bateson es denominada epistasis clásica (Aylor & Zeng, 2008).

En genética cuantitativa, epistasis es utilizada para describir en términos generales la combinación de variantes genéticas en relación con rasgos fenotípicos (Wei et al., 2014). Epistasis, en este sentido, representa una relación funcional entre genes en el contexto de un rasgo complejo (Campbell et al., 2018) y podría tener implicaciones importantes en la variación y heredabilidad. En genética cuantitativa los factores genéticos representan sólo una parte del riesgo asociado a una enfermedad compleja (Boyle et al., 2017).

Conectar variantes genotípicas a su efecto fenotípico es complicado para la mayoría de los rasgos naturales (de los Campos et al., 2018). Esta dificultad es consecuencia de la alta poligenicidad<sup>21</sup>, es decir, de la dependencia de un rasgo a dos o más factores genéticos involucrados. Además, variaciones de los rasgos naturales son causadas por alelos de baja frecuencia. Este problema fue evaluado en 1918 por Fisher, que describe desviaciones de varianza aditiva en un modelo genético estadístico usando la regresión lineal para estudiar la selección natural relacionada con los rasgos complejos, esto se denominó epistasis estadística (Fisher, 1919).

La epistasis estadística es ilustrada usando funciones de penetrancia en un modelo de probabilidad condicional:  $P(D/G)$ , donde  $P$  es la probabilidad de una enfermedad ( $D$ ) dado un genotipo en particular ( $G$ ); es decir, es un modelo de probabilidad de una enfermedad dado un fenotipo en particular o combinación de genotipos y sus respectivas frecuencias. La

---

<sup>20</sup> Locus, del latín *locus*, hace referencia a lugar o posición genética, y deriva en el término en plural *loci*.

<sup>21</sup> La cualidad de tener múltiples genes. La herencia poligénica ocurre cuando una característica está controlada por dos o más genes.

interpretación estadística del modelo depende en gran medida del genotipo y del contexto en el que se encuentra (Hermisson et al., 2003).

El estudio de la epistasis en poblaciones bacterianas ha permitido explicar la evolución adaptativa de las especies en el contexto de los rasgos complejos que las constituyen, por ejemplo, se ha descrito la codependencia de la tasa de recombinación y de algunos elementos de la estructura de la población, por lo que los avances actuales en modelos estadísticos demuestran cada vez más la existencia de interacciones epistáticas no descritas (Arnold et al., 2018).

Además, la epistasis es relevante para el estudio de la resistencia a los antibióticos. Poblaciones bacterianas, que proporcionan un reservorio que puede transmitirse mediante la transferencia horizontal de genes, hacen que los microorganismos se vuelvan resistentes a un determinado fármaco (Brown y Wright, 2016). La capacidad de identificar y determinar mutaciones específicas puede afectar la epistasis entre alelos resistentes y es esencial para comprender cómo la resistencia a los antimicrobianos persiste en poblaciones bacterianas (Balaban et al., 2019; Durão et al., 2015).

### **1.8.1. Epistasis en *Mycobacterium tuberculosis***

La epistasis puede representar cambios en la diversidad genética, en la evolución, en la resistencia farmacológica y en funciones metabólicas. La evolución en la resistencia a múltiples antibióticos es debida a la adquisición secuencial y acumulación de mutaciones que confieren resistencia en el cromosoma de *M. tb*. Las mutaciones en *M. tb* principalmente son relacionadas con el blanco farmacológico, por ejemplo, RIF con el gen *rpoB*, FQ con el gen *gyrA*; o causar una sobreexpresión del blanco, por ejemplo, INH/ETB promotor de la región del gen *inhA*. Estas mutaciones son relevantes clínicamente y determinan la resistencia clínica de los aislados bacterianos (Gilpin & Mirzayev, 2019; Ramaswamy & Musser, 1998).

Las mutaciones en *katG* S315T han precedido a las mutaciones del gen *rpoB* que conduce a la adquisición de resistencia a la RIF. La combinación de mutaciones en *katG*315 y *rpoB*531 con mutación compensatoria de resistencia a RIF se ve favorecida en el aislado clínico

caracterizado como cepa MDR, esta combinación de mutaciones que interactúan entre ellas, se ha descrito como interacciones epistáticas asociadas a resistencia antimicrobiana (Borrell et al., 2013; Trauner et al., 2014).

Los avances recientes en la escala y diversidad de la genómica de poblaciones con datos para *M. tb* proporcionan el potencial para revelar patrones genéticos dentro del genoma completo. Utilizando métodos estadísticos combinados con recientes avances en biología estructural computacional se han identificado los *loci* polimórficos bajo fuertes presiones co-evolutivas o interacciones epistáticas (Sackton & Hartl, 2016; Skwark et al., 2017).

Por otro lado, se han descrito secuencias altamente conservadas con variaciones secuenciales concentradas en genes PE/PPE, identificando 97 interacciones epistáticas con la metodología computacional *machine learning*, que genera un modelo predictivo a partir de datos de entrenamiento (Kavvas et al., 2018). Sin embargo, hasta el momento no se han caracterizado otros genes vinculados a interacciones epistáticas en el genoma de *M. tb*.

Los estudios mencionados anteriormente de interacciones entre mutaciones en *M. tb* que resultan en resistencia a diversos fármacos han sugerido que la epistasis puede estar relacionada con resistencia a múltiples fármacos. El papel de las interacciones epistáticas entre muchas regiones del genoma bajo selección en *M. tb* sigue siendo desconocido y aumentar el estudio de epistasis en *M. tb* contribuirá a mejorar nuestro conocimiento sobre la bacteria y la enfermedad.

## **CAPÍTULO 2. OBJETIVOS**

### **2.1. Objetivo general**

Predecir bioinformáticamente redes de interacciones epistáticas en *M. tb* a partir del análisis de genomas completos para identificar mutaciones asociadas a diferentes genes involucrados en diferentes rasgos biológicos.

### **2.2. Objetivos particulares**

- Seleccionar genomas completos de referencia pertenecientes a *M. tb*.
- Generar un alineamiento sobre los genomas completos de referencia de *M. tb*.
- Predecir con herramientas bioinformáticas las interacciones epistáticas en *M. tb*.
- Crear una red de interacciones epistáticas de *M. tb*.
- Asociar las interacciones epistáticas con sus respectivos genes y dominios, para relacionarlos con sus funciones biológicas.

## **CAPÍTULO 3. HIPÓTESIS**

El análisis bioinformático de genomas completos permitirá identificar redes de interacciones epistáticas en *M. tb* entre genes asociados a diferentes rasgos biológicos.

## **CAPÍTULO 4. ANTECEDENTES**

La adquisición de la resistencia a los antibióticos de las cepas MDR y XDR requiere tratamiento prolongado con medicamentos costosos, eficacia limitada y aumento de efectos adversos; así, a menudo se asocia a estas cepas con mala respuesta al tratamiento (Seung et al., 2015). Además, se han descrito los mecanismos de acción de resistencia a antibióticos y la identificación de determinantes de resistencia que hoy pueden ser utilizados como marcadores moleculares fiables para la susceptibilidad a fármacos (WHO, 2021). Se ha definido que las mutaciones son sólo una parte del problema de resistencia, ya que puede estar fuertemente



influenciado por mecanismos de interacciones entre genes (epistasia) y puede transmitirse de un hospedero a otro, lo que muestra que los antecedentes genéticos de la cepa pueden determinar el curso de la evolución de la resistencia farmacológica (Trauner et al., 2014).

Hasta el momento, han sido descritas algunas interacciones epistáticas tentativas para *M. tb* en el contexto de la presión selectiva debida a fármacos. En el trabajo descrito por Borrell y colaboradores en el 2013, se reportan interacciones genéticas tentativas descubiertas por la metodología de PCR dirigida a la detección puntual de una región genética específica y amplificación en los genes *rpoB* y *gyrA*. Por otro lado, Kavvas y colaboradores en el 2018 describen un análisis computacional con la metodología *machine learning* que permitió detectar 97 potenciales interacciones epistáticas asociadas a resistencia farmacológica usando datos de WGS de la base de datos PATRIC. Sin embargo, aún existe poca información acerca de este fenómeno de epistasia considerando el análisis de genomas completos y su estudio para identificar interacciones epistáticas asociadas a diferentes mecanismos biológicos como virulencia, patogénesis, resistencia a los antibióticos, moduladores del sistema de transporte y de la respuesta inmune.

Un método que logra la identificación de interacciones epistáticas tentativas en poblaciones bacterianas usando genoma completo es el desarrollado por Pensar y colaboradores (2019). Su método, SpydrPick, permite la detección de genes blanco de co-selección en ausencia de caracterización fenotípica. El método ha sido probado en las poblaciones bacterias de *Streptococcus pneumoniae* y *Neisseria meningitidis*; sin embargo, no se había probado en *M. tb* hasta el presente trabajo doctoral.

Así, este trabajo presenta un estudio de interacciones epistáticas tentativas caracterizadas con metodología bioinformática en genomas completos y esperamos que en un futuro se comprueben estas interacciones a nivel molecular o celular para evaluar sus respectivos efectos fenotípicos que permitan el desarrollo de fármacos, vacunas o contribuyan al diagnóstico de resistencia farmacológica oportuna de *M. tb*.

## CAPÍTULO 5. METODOLOGÍA COMPUTACIONAL

En la Figura 3, se muestra una descripción general del enfoque computacional para crear redes epistáticas a partir del análisis de genomas completos de *M. tb* y el estudio de interacciones epistáticas. Los pasos son descritos con detalle en los siguientes subcapítulos.

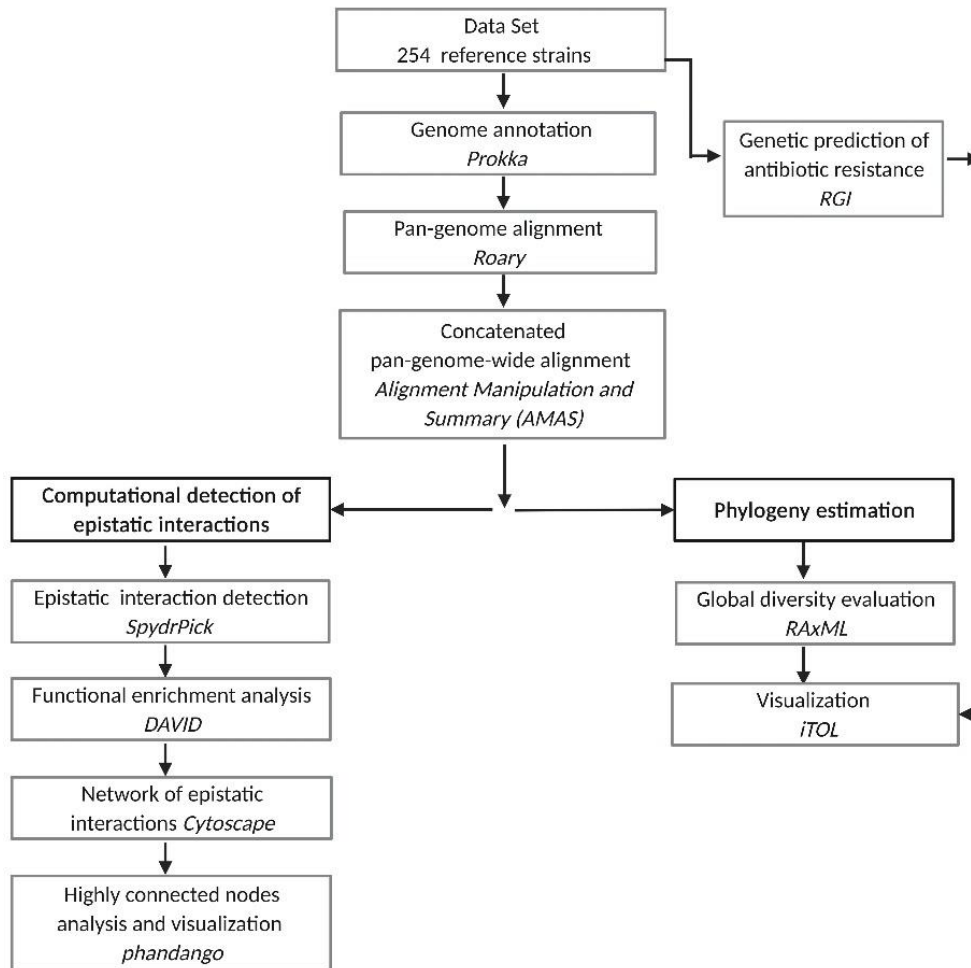


Figura 3. Enfoque bioinformático para el estudio de interacciones epistáticas en *M. tb*.

### 5.1. Conjunto de datos

En el presente trabajo de investigación, se usaron 254 cepas de referencia de *M. tb* obtenidas de la base de datos NCBI Refseq, que estaban disponibles en la fecha del 4 de noviembre de 2020. Dentro de los criterios de inclusión para la selección de cepas se encuentran las características de genomas completos > 4 Mb y considerar únicamente a la especie *M. tb*.

El origen geográfico de las cepas incluye a los siguientes países, en mayor porcentaje: Perú 29% (74 cepas), India 14% (37), Guatemala 6% (16), Suecia 6% (16), Bélgica 5% (13), Moldavia 5% (13) y Sudáfrica 4.7% (12). En menor porcentaje: China 3% (7), Colombia 2.3% (6), Hong Kong 2.3% (6), Filipinas 2% (5) y Estados Unidos de América 2% (5). Y en mínimo porcentaje (menor a 2%, esto es, menor a 4 cepas): Vietnam, Corea del sur, Japón, Rusia, Australia, Kazajistán, Dinamarca, Australia, Canadá, Malasia, Bielorrusia, Taiwán, Nueva Zelanda y Ruanda. Los porcentajes anteriores describen un total de 234 cepas que incluían el país en su archivo de anotación (archivo GFF) descargado del NCBI. Las 20 cepas restantes carecen de información del país en el archivo GFF y representan el 8% del conjunto de datos. En la Figura 4, se observa la distribución geográfica.

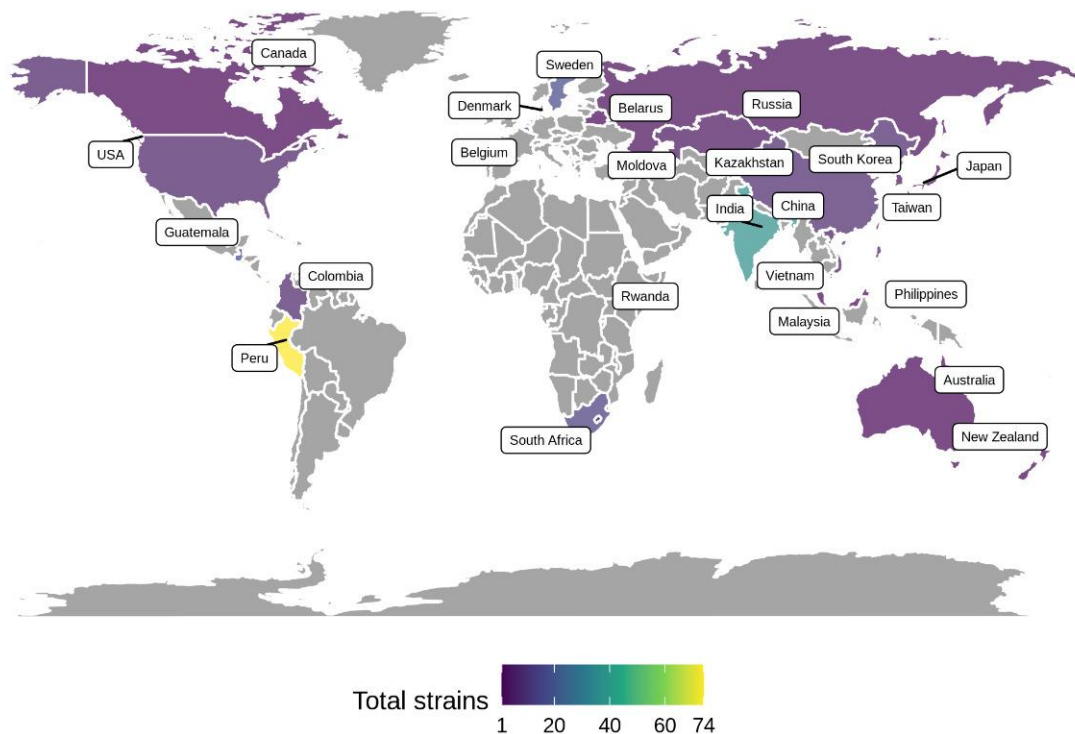


Figura 4. Distribución geográfica de 234 cepas del conjunto de datos.

Los archivos GFF también incluían el tipo de cepa y mostraban alta heterogeneidad de tipos. Por otro lado, el 49% (126 cepas) son del año 2000 al 2019, el 3% (7) son del año 1951 al 1998 y el 48% (121) tenían ausencia de información del atributo año. El 36% (92) refieren

obtención de la muestra en esputo<sup>22</sup>, el 5% (12) otro atributo (por ejemplo, *clinical*), el 25% (64) obtenidas de pulmón y el 34% (86) con ausencia de información. A pesar de que los archivos GFF ayudaron a la caracterización de las cepas, se observó falta de información en varios atributos. Esta carencia limita una caracterización completa y resalta la importancia de mejorar las anotaciones de cepas en las bases de datos.

La longitud en pares de bases (pb) de la cepa más pequeña es de un tamaño de 4, 359, 659 pb y corresponde a la cepa TBMENG-03 (acceso genebank GCF\_003293655) muestra del 2016 de esputo del Nacional JALMA Instituto de Lepra y otras Mycobacterias en la India; mientras que la cepa más grande de los datos es la 3-0096P6C4 (acceso genebank GCF\_014899985), muestra de pulmón de la Fundación de Enfermedades Tropicales en Filipinas con 4, 450, 340 pb. La longitud promedio de las 254 cepas fue de 4, 412, 106 pb.

## 5.2. Generación de Alineamiento de genoma completo

Siguiendo la estrategia de Pensar y colaboradores (2019), se creó un alineamiento de genoma completo de las 254 cepas. Primero se utilizó la herramienta Prokka (Seemann, 2014) para anotar genes y características de interés en el conjunto de cepas. Posteriormente este alineamiento anotado (formato GFF3) fue la entrada para crear un *pan-genoma*<sup>23</sup> de las cepas con la herramienta Roary (Page et al., 2015). Esta herramienta extrae las secuencias de genes de la entrada y luego identifica grupos para obtener un alineamiento de genes individuales.

Roary considera dos categorías de genes: núcleo (esenciales) y accesorios (dispensables). Se considera que un gen se encuentra en la categoría de núcleo cuando el gen se encuentra en todos los miembros del genoma procariótico, y se ha reportado que los genes que representan el núcleo se encuentran en un mayor porcentaje en todos los genomas de la especie (McInerney et al., 2017). En la investigación, se siguió el enfoque de Pensar y colaboradores (2019), quienes fijan este porcentaje predeterminado en 95% para identificar

---

<sup>22</sup> Moco u otra materia que se expulsa desde los pulmones al toser.

<sup>23</sup> El pan-genoma se refiere al conjunto de genes encontrados en una especie.

genes núcleo de las cepas. La salida de Roary es un conjunto de archivos con alineamiento de genes individuales, un archivo por gen.

Estos archivos fueron concatenados en una matriz utilizando la herramienta AMAS, *Alignment Manipulation and Summary* (Borowiec, 2016). Esta matriz está formada por el gen 1 unido por la derecha con el gen 2 y así sucesivamente con el resto de genes (Figura 5). Así, las columnas de esta matriz de salida son los genes y el número de filas es el número de cepas utilizadas para generar la alineación de todo el genoma (pan-genoma).

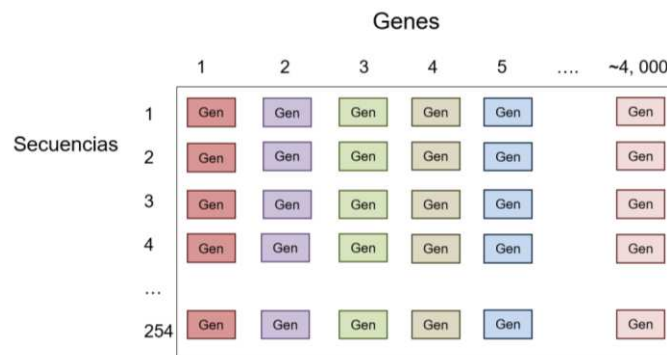


Figura 5. Imagen ilustrativa del pan-genoma completo alineado

En la imagen se ilustra la matriz de genes alineados. Las filas se refieren a las secuencias y el número de columnas corresponde al número de genes generando en conjunto un pan-genoma completo alineado.

### 5.3. Análisis filogenético

El alineamiento de pan-genoma completo fue usado para estimar la diversidad genética mediante el desarrollo de un árbol filogenético, por lo que se usó la herramienta RAxML *Next Generation* (Kozlov et al., 2019). Los modelos y parámetros que se utilizaron se mostrarán más adelante (sección 6). Se utilizó una prueba no paramétrica Bootstrap con 1, 000 réplicas. El árbol fue visualizado posteriormente con iTOL v 6.4.1 (Letunic y Bork, 2021).

## 5.4. Predicción de la caracterización de resistencia genética a los antibióticos

Se realizó la predicción de un resistoma<sup>24</sup> para las 254 cepas usando la herramienta *Resistance Gene Identifier* (RGI) v. 5.1.1 (Alcock et al., 2020). RGI usa la base de datos de referencia para la caracterización de resistencia a los antibióticos *Comprehensive Antibiotic Resistance Database*, CARD.

Usando el archivo de salida de RGI, se anotaron las cepas para la caracterización de antibióticos de resistencia MDR para los fármacos de INH, RIF, y para la resistencia XDR para los fármacos INH, RIF, FLQ, KM, AMK, CM. Estas anotaciones fueron incorporadas en la visualización de la filogenia con la herramienta iTOL.

## 5.5. Detección computacional de interacciones epistáticas

### 5.5.1. Detección de interacciones epistáticas

Se utilizó SpydrPick (Pensar et al., 2019) para detectar interacciones epistáticas en el alineamiento de pan-genoma completo. SpydrPick es un método computacional libre de modelo eficiente para el análisis a gran escala de pan-genomas de bacteria. Este método facilita la detección de blancos de co-selección relacionados con la virulencia, patogénesis y resistencia a los antimicrobianos. El potencial de métodos como este es la detección de interacciones epistáticas en ausencia de datos fenotípicos.

La aproximación de SpydrPick está basada en calcular la información mutua (*Mutual information*, MI) entre dos *loci* polimórficos. La MI es una medida de información entre variables aleatorias, y es la cantidad de información que una variable,  $X$ , contiene sobre otra variable,  $Y$ . La MI también se define como la reducción de incertidumbre en  $X$  después de observar  $Y$ , en otras palabras, MI manifiesta la reducción de la incertidumbre de  $X$  debido al conocimiento de  $Y$  (Cover y Thomas, 1991). La MI se puede ver como una medida de asociación de correlación entre  $X$  e  $Y$  (Chanda et al., 2020): si dos variables,  $X$  e  $Y$ , son independientes, entonces la medida de MI es cero. MI es formalmente definida como:

---

<sup>24</sup> Conjunto de genes de resistencia.

$$MI(X, Y) = \sum_{x \in (X)} \sum_{y \in (Y)} p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right).$$

En la fórmula anterior  $p(x, y)$  es la probabilidad conjunta, y  $p(x)$  y  $p(y)$  son las probabilidades marginales de  $X$  e  $Y$ .

La MI se ha utilizado con éxito para detectar la co-selección en la genómica de poblaciones bacterianas a escala de genoma completo (Pensar et al., 2019). Una característica relevante introducida por el enfoque de SpydrPick es la corrección de la estructura poblacional. Esto se aplica mediante una estrategia de reajuste usando un peso de las secuencias calculado en función de qué tan diferentes son entre ellas en la alineación de pan-genoma completo (Pensar et al., 2019).

SpydrPick detecta interacciones directas e indirectas entre *loci*. Una interacción directa se da entre dos posiciones ( $P1 \rightarrow P2$ ), mientras que una interacción indirecta ocurre cuando las dos posiciones ( $P1$  y  $P2$ ) también están vinculadas a través de una tercera posición ( $P1 \rightarrow P3 \rightarrow P2$ ). En el caso de interacciones indirectas ( $P1 \rightarrow P2$ ), SpydrPick elimina la interacción si la MI no es mayor que la MI de las otras dos interacciones ( $P1 \rightarrow P3$  y  $P3 \rightarrow P2$ ).

SpydrPick realiza un análisis para detectar valores atípicos de la MI de las interacciones. En este caso, un valor atípico está alejado de los demás valores y es un valor mayor a los valores típicos. Un primer criterio para encontrar los valores atípicos es que la distancia en pb entre las posiciones de los *loci* polimórficos debe ser mayor que un parámetro de desequilibrio de ligamiento (*linkage disequilibrium*, LD).

En este caso, un LD fuerte se refiere a una estrecha distancia genética entre dos posiciones de nucleótidos. Debido a que un fuerte LD oculta una posible señal de presión de selección co-evolutiva compartida, SpydrPick descarta pares de posiciones con LD fuerte y así selecciona interacciones con valores atípicos más fiables. De acuerdo con la documentación de SpydrPick (<https://github.com/santeripuranen/SpydrPick>), los valores típicos de LD para bacterias están en el rango de 500 a 20, 000 pb y el enfoque predeterminado para filtrar pares con fuerte LD consiste en usar un simple límite basado en la distancia de 20, 000 pb. Un

segundo criterio para detectar valores atípicos es que la MI debe ser mayor que un umbral obtenido con la prueba de valores atípicos de Tukey  $Q_3 + 1.5 \times (Q_3 - Q_1)$  (Tukey, 1977).

La salida de SpydrPick es una tabla de interacciones epistáticas que incluye el par de posiciones en la alineación de pan-genoma completo de los dos *loci* polimórficos que interactúan, la distancia genómica entre las dos posiciones, el tipo de interacción (directa/indirecta) y el valor de MI. Cuando SpydrPick detecta valores atípicos, entrega otra tabla sólo con las interacciones epistáticas que tienen estos valores atípicos, incluyendo los mismos campos y tres campos adicionales: el valor de MI sin gaps, el efecto del gap y si el valor atípico es considerado un valor atípico extremo. Los valores atípicos extremos se determinan de acuerdo al siguiente criterio:  $(MI > Q_3 + 3 \times (Q_3 - Q_1))$ .

En el alineamiento de entrada, SpydrPick categoriza cualquier carácter diferente de A, C, G y T como un gap. Los gaps son considerados en el cálculo de la MI, por lo que X e Y tienen un espacio de resultados de cinco categorías (A, C, G, T y gap). Dado que los gaps pueden ser no informativos, SpydrPick calcula para cada par de posiciones de los valores atípicos otra puntuación considerando sólo aquellas cepas sin gaps en ambas posiciones. Esta puntuación se denomina MI sin gaps (MI\_sin\_gaps). Utilizando esta puntuación se calcula el efecto del gap como:  $(1 - MI\_sin\_gaps / MI) \times 100$ . Esto cuantifica el efecto positivo o negativo de los gaps sobre el valor de MI.

Comparar valores de MI sin gaps de manera significativa es difícil debido a que el conjunto de cepas sin gaps en las dos posiciones varía entre pares de posiciones (Pensar et al., 2019). Así, siguiendo el análisis realizado por Pensar et al. (2019), usamos el valor normal de MI, dejando el análisis de la MI sin gaps para un estudio futuro.

Los *loci* de las interacciones epistáticas se anotaron con el identificador del gen y nombre del gen. Los nombres de los genes se obtuvieron del archivo generado por AMAS utilizando un script en R. Luego, usando la biblioteca Biomart de R (Drost H & Paszkowski J, 2017), se recuperó el ENTREZ id buscando el nombre del gen en el archivo GFF del genoma de referencia H37Rv de *M. tb*. Si la herramienta Roary no detecta ningún nombre del gen durante la creación del pan-genoma, entonces Roary da un nombre genérico único formado por el prefijo *group* y



un número consecutivo. Estos nombres genéricos aparecen en el archivo de AMAS, pero no existen en el archivo GFF del genoma de referencia, por lo que el ENTREZ id no pudo recuperarse para estos nombres genéricos.

Por otro lado, Prokka indica con un sufijo numérico anotaciones diferentes para el mismo gen, como *carB\_1* y *carB\_2* (cadena larga de carbamoil-fosfasa sintetasa). Estos nombres tampoco se encontraban en el archivo GFF del genoma de referencia. En estos casos, eliminamos el sufijo numérico para encontrar el nombre del gen en el archivo del genoma de referencia. Así, por ejemplo, encontramos el ENTREZ id 886253 para el gen *carB*.

### **5.5.2. Análisis de enriquecimiento funcional**

Se utilizó la base de datos para la anotación, visualización y descubrimiento integrado v.6.8 (*database for annotation, visualization and integrated discovery*, DAVID) (Huang et al., 2009) para obtener una anotación funcional de términos de Gene Ontology (GO) y las vías metabólicas de KEGG de los genes que participaron en las interacciones epistáticas. Específicamente se utilizó el Script de Python para generar un reporte de enriquecimiento (*chart report*) del servicio web de DAVID (Jiao et al., 2012).

### **5.5.3. Redes de Interacciones Epistáticas**

El conjunto de interacciones epistáticas puede verse como un modelo de complejas relaciones que pueden ser analizadas y visualizadas como una red. Por esto, se utilizó Cytoscape (Kohl et al., 2011) para estudiar el conjunto de interacciones epistáticas como una red epistática. Cytoscape incluye una herramienta de análisis de redes que calcula varias medidas, como el grado de un nodo<sup>25</sup> (*node degree*) y la centralidad de intermediación<sup>26</sup> (*betweenness centrality*).

Otra herramienta útil de Cytoscape es el conjunto de algoritmos de disposición de nodos que calculan automáticamente la disposición de la red (*yFile Layout Algorithm App*). Estos algoritmos organizan visualmente la red alineando y rotando grupos de nodos.

---

<sup>25</sup> Es el número de aristas conectadas al nodo (punto de intersección, conexión o unión).

<sup>26</sup> Medida de centralidad en un grafo basada en los caminos más cortos para cada par de nodos en un grafo conectado.

## CAPÍTULO 6. RESULTADOS

Véase la sección Anexo I (artículo publicado) para verificar los resultados y consultar material suplementario.

La alineación de pan-genoma completo generado con Roary resultó en un total de 6, 205 genes individuales, incluidos 3,659 genes núcleo. Después de concatenar todos los genes individuales con AMAS, un alineamiento de todo el genoma de 6, 751, 593 pb fue obtenido.

Basado en el criterio de información de Akaike (*Akaike information Criterion, AIC*) y el criterio de información bayesiano (*Bayesian Information Criterion, BIC*), que son dos criterios para penalizar modelos complejos, se seleccionó una filogenia usando un modelo GTR con cuatro tasas de libertad (GTR+R4+FO).

La comparación de modelos se proporciona en la Tabla 3. Se alcanzó una convergencia usando el criterio *extended majority rule*, MRE (Pattengale et al., 2010) con un umbral de corte del 3% mediante bootstrapping<sup>27</sup> después de 400 árboles.

Tabla 3. Valores de las evaluaciones de modelos de filogenia con RAxML Next Generation.

En la tabla se visualizan los valores de las evaluaciones de los diferentes modelos de filogenia. De acuerdo con los criterios AIC y BIC, se seleccionó el mejor modelo considerando su valor más bajo: GTR+R4+FO (indicado en fuente negrita). AIC= Akaike's an information criterion; BIC= Bayesian information criterion; JC = Jukes-Cantor; G= Gamma model rate heterogeneity; GTR= Simple GTR model; FC = Empirical base frequencies; FO = ML estimation of the base frequencies; R4= four free rates instead of GAMMA-distributed rates.

Modelo	log Likelihood	AIC	BIC
<b>GTR+R4+FO</b>	<b>-10090860.066867</b>	<b>20182758.133734</b>	<b>20189881.415436</b>
GTR+G+FO	-10222603.873729	20446235.747459	20453290.404096
GTR+G+FC	-10223046.653685	20447121.30737	20454175.964008
GTR	-10286598.064064	20574222.128129	20581263.059754
JC+G	-10603345.407673	21207702.815346	21214647.67188
JC	-10667194.195936	21335398.391872	21342329.523394

<sup>27</sup> Es una técnica informática para evaluar la precisión de cualquier estimación estadística.

La predicción de la resistencia a los antibióticos por RGI reportó que el 100% de las 254 cepas son caracterizadas como MDR genóticamente; es decir que reportan en la base de datos de CARD algún gen que contiene resistencia farmacológica (Figura 6) e igualmente se reporta que el 15% contenían genes vinculados a la caracterización de XDR. El árbol filogenético de la Figura 6 muestra la diversidad nucleotídica de *M. tb* en nuestra colección de 254 cepas, que incluye a la cepa de referencia H37Rv (resaltada en color amarillo).

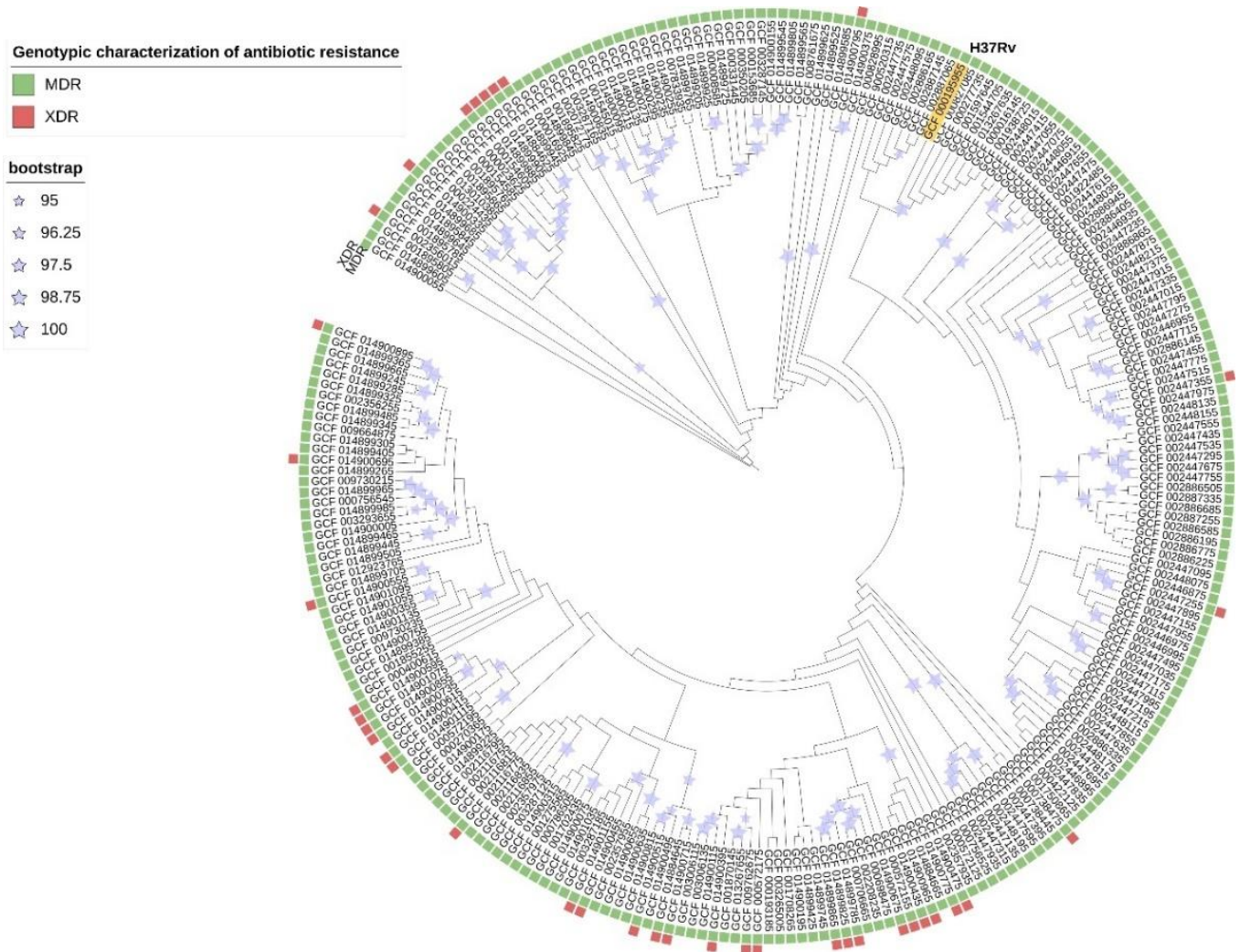


Figura 6. Árbol filogenético del pan-genoma completo de *M. tb* visualizado por iTOL.

En el árbol en forma circular se muestran la longitud de ramas y el valor de los Bootstrap. La longitud es proporcional a la topología de los nucleótidos. Los valores de bootstrap de los clados corresponden a los mostrados en las leyendas. Los colores corresponden a las diferentes caracterizaciones genótipicas de la resistencia a los antibióticos (MDR y XDR).

Esta misma filogenia junto con características de las cepas como la institución y la resistencia farmacológica serán incluidas más adelante en el análisis de distribución alélica de

los genes relevantes asociados a interacciones epistáticas usando la herramienta Phandango (Hadfield et al., 2018).

## 6.1. Detección de interacciones epistáticas

Se detectaron 10, 573 interacciones epistáticas con valores atípicos (5,484 directas y 5, 089 indirectas). Estas interacciones describen *loci* polimórficos bajo la presión co-evolutiva más fuerte. Se proporciona una tabla con la lista completa de interacciones con valores atípicos en el material suplementario del artículo del Anexo I.

Después de la anotación de genes, se generó una nueva tabla donde se excluyeron las interacciones con genes asociados con nombres genéricos obtenidos por Roary. La nueva tabla de valores atípicos incluyó 1, 940 interacciones epistáticas entre 107 genes únicos.

De este conjunto de genes, sólo se encontraron 70 en el archivo GFF del genoma de referencia y se asociaron con su ENTREZ id. Se recuperaron sólo las interacciones que incluían estos 70 genes y se obtuvo una tabla final con 890 interacciones epistáticas. Las interacciones que no fueron consideradas en este estudio serán incluidas en una futura investigación.

La tabla final de interacciones con valores atípicos incluye las dos posiciones de los *loci* polimórficos que interactúan en la alineación de todo el genoma, el ENTREZ id y nombre de los genes para cada posición, la distancia entre las dos posiciones (pb), el tipo de interacción (directa = 1, indirecta = 0), el valor de MI y si el valor atípico es valor extremo (sí =1, no=0). Esta tabla final de valores atípicos se utilizó para el análisis de enriquecimiento, reconstrucción de redes y análisis.

SpydrPick reportó interacciones a larga distancia superando los dos millones de pb (Tabla 4), este dato confirma que se encontraron interacciones epistáticas a lo largo de todo el genoma completo. La distancia mínima en pb entre las posiciones de *loci* polimórficos que interactúan superó el criterio de LD de 20, 000 pb. La media de la distancia entre *loci* polimórficos en las interacciones atípicas fue de 846, 454 pb, considerando que es mayor que la mediana (721, 980 pb), hay un mayor número de distancias inferiores a la media. Por otro

lado, el rango de valores de MI fue corto, de 0.4130 a 0.5020 (Tabla 4). Hay un mayor número de valores bajos de MI ya que la media (0.4509) fue más alta que la mediana (0.4202).

Todos los *loci* en las 890 interacciones epistáticas se encontraron en los polimorfismos de un sólo nucleótido (*Single Nucleotide Polymorphism, SNP*) identificados con la herramienta SNP-sites v. 2.5.1 en la alineación de todo el genoma. Esta herramienta puede identificar rápidamente los SNP de una alineación multi-FASTA (Page et al., 2016) . Este paso adicional fue desarrollado para confirmar los polimorfismos implicados en las interacciones epistáticas detectadas.

Tabla 4. Estadísticas de las interacciones con valores atípicos finales.

Estadísticas de la distancia entre pares de posiciones de interacciones de sitios polimórficos (*loci*) y estadísticas de valores de MI.

Estadística	Valores
Mínima distancia (pb)	20,870
Máxima distancia (pb)	2,328,291
Mediana de distancia (pb)	721, 980
Media de distancia (pb)	846, 454
Valor mínimo de MI	0.4130
Valor máximo de MI	0.5020
Mediana de MI	0.4202
Media de MI	0.4509

## 6.2. Análisis de la Red Epistática *glgB* y *oppA*, genes blanco de co-selección

Se analizó la red de interacciones epistáticas obtenida de Cytoscape para averiguar aquellos genes con un alto grado de conectividad (*node degree*), es decir, un alto nivel de conectividad del gen con otros genes. Los genes más conectados fueron ***glgB*** (ENTREZ: 886893, degree = 56), que codifica para la enzima ramificadora  $\alpha$ -1,4-glucano (GlgB) y ***oppA*** (ENTREZ: 886985, degree= 37), oligopéptido unido a proteína (OppA) (Figura 7).

Estos dos genes también tienen el valor más alto de centralidad (*glgB* = 0.493, *oppA* = 0.219). La centralidad es más alta para aquellos nodos que unen a subredes (comunidades) que para aquellos que se encuentran dentro de las subredes. En este estudio, observamos tres subredes conectadas por estos dos genes.

Un patrón interesante es que cada subred tiene una distribución diferente de valores de MI (ver gradiente de color de los enlaces en la Figura 7). La subred en la parte inferior tiene valores mayores de MI que los otros (mediana de MI = 0.502). La subred superior tiene un valor de mediana de MI = 0.463 y la subred de en medio tiene valor de mediana de MI = 0.420. Se requiere un estudio adicional para encontrar la causa de este patrón. Además, futuros análisis serán necesarios para identificar los patrones asociados con las subredes aisladas (*mmpL1-mmpS4* y *lipR-ponA1*).

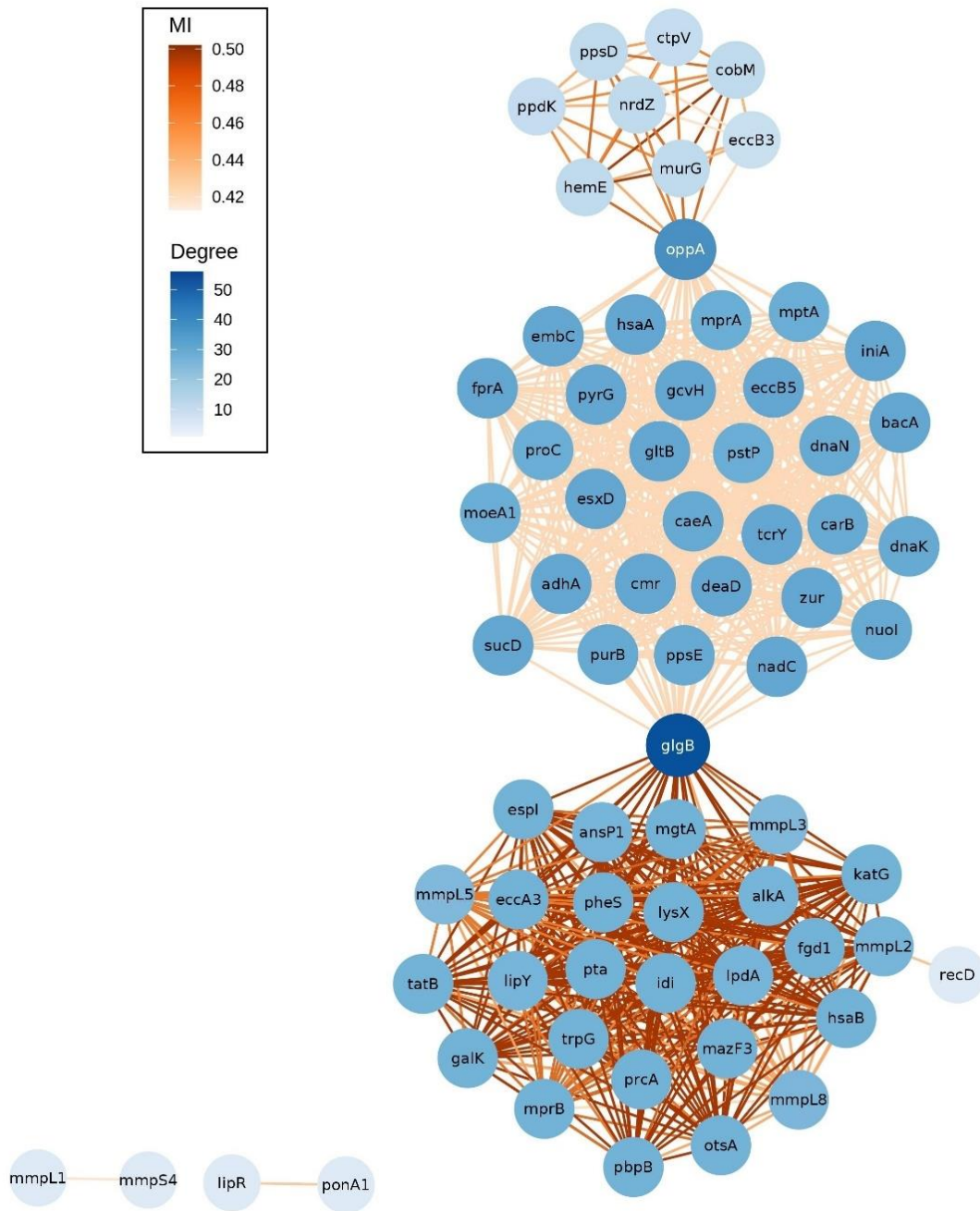


Figura 7. Red epistática creada con las interacciones con valores atípicos de MI.

El gradiente de color de los nodos muestra el grado de conectividad (*degree*). El gradiente de color de los enlaces muestra el valor de MI. Así, se muestra que los genes *glgB* y *oppA* tienen mayor conectividad entre las tres subredes.

### 6.3. Enriquecimiento de términos de GO para redes epistáticas

Se utilizó la lista de ENTREZ ids de los genes de la red para realizar un análisis de enriquecimiento funcional con DAVID. Del reporte de DAVID, sólo consideramos los términos como relevantes con valor  $p\text{-value} < 0.05$ . Los procesos biológicos de patogénesis (GO: 0009405) y organización de la pared celular (GO: 0071555) fueron enriquecidos en un subconjunto de genes (Figura 8).

Pared celular (GO: 0005618), membrana plasmática (GO: 0005886), citosol (GO: 0005829) y componentes integrales de la membrana plasmática (GO: 0005887) fueron los componentes celulares más abundantes; en este caso, el 63% de los genes están en la membrana plasmática (Figura 8). En cuanto a las funciones moleculares, se obtuvo enriquecimiento en algunos genes para la unión de ATP (GO: 0005524) y actividad de la fosfoproteína fosfatasa (GO: 0004721).

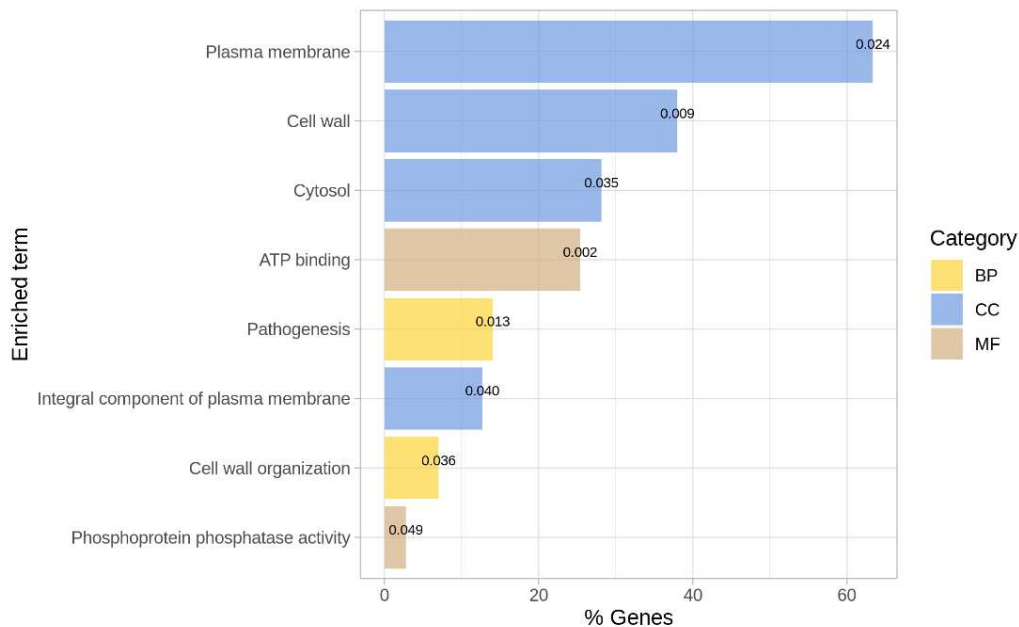


Figura 8. Enriquecimiento de términos GO de valores atípicos.

El valor de  $p\text{-value}$  indicado para cada barra es significativo para cada término. BP = Proceso biológico, CC = componente celular; MF = función molecular.



Los términos GO enriquecidos de genes que interactúan con *glgB* y *oppA* se muestran usando gráficos circulares (Figura 9). Además, en el material suplementario del Anexo I, se incluyen los genes que interactúan con *glgB*, su producto y enriquecimiento con términos GO en forma de tabla, la misma información se proporciona para *oppA*.

Los gráficos circulares se generaron utilizando las posiciones de inicio y final de los genes reportados en el archivo generado por AMAS, por lo que la posición y el tamaño de los genes en el gráfico y las posiciones de las interacciones corresponde a la alineación de todo el pan-genoma. Se resaltaron las interacciones de *glgB* y *oppA* en rojo para distinguirlas de las interacciones de otros genes (mostradas en gris). Tanto *glgB* como *oppA* interactúan epistáticamente con genes enriquecidos con los procesos biológicos de patogénesis y organización de la pared celular (Figura 9D y H).

#### **6.4. Distribución alélica en *loci* de genes que interactúan con *glgB* y *oppA***

Para observar los patrones de los alelos de los *loci* polimórficos que interactúan, se muestran sus distribuciones en las interacciones epistáticas de *glgB* y *oppA* usando Phandango (Hadfield et al., 2018). SpydrPick detectó que tres *loci* de *glgB* (837, 764; 839, 047 y 839, 053) interactúan con 57 *loci* polimórficos de 56 genes. Para *oppA*, dos *loci* polimórficos (5, 934, 914 y 5, 936, 231) fueron encontrados interactuando con 38 *loci* de 37 genes. Las tablas con los datos de las interacciones se encuentran en el material suplementario del Anexo I.

Los *loci* y los nombres de los genes que interactúan se muestran como etiquetas de columnas en la Figura 10 para *glgB* y en la Figura 11 para *oppA*, los *loci* están organizados en secciones con bordes. Cada sección incluye las interacciones para cada *locus* con el que interactúan. El color del borde para cada sección corresponde al color de las secciones en los archivos suplementarios. En la Figura 10, la primera sección comienza con los *loci* 837, 764 y 839, 047 de *glgB* (*glgB\_837764* y *glgB\_839047*) seguido de los 29 *loci* polimórficos que interactúan con ellos, es decir, estos dos *loci* interactúan epistáticamente con cada uno de los 29 *loci*.

Al observar la distribución de alelos de pares de *loci* que interactúan, confirmamos que SpydrPick es capaz de detectar, utilizando el valor de MI, patrones predecibles de alelos en los dos *loci*. Por ejemplo, la segunda sección en la Figura 10, que representa la distribución de los alelos del *locus* 839, 047 de *glgB* (*glgB\_839047*) y el *locus* 818, 177 del gen *gcvH* (*gcvH\_818177*), muestra que cuando hay una C en el *locus* *glgB\_839047*, hay una A en el *locus* *gcvH\_818177*, y cuando hay una T en el *locus* *glgB\_839047*, hay una C en el *locus* *gcvH\_818177*. Esto se representa cuantitativamente por el valor de MI = 0.4201.

La tercera sección de la Figura 10 expone la distribución de alelos de las interacciones entre el *locus* 839, 053 de *glgB* (*glgB\_839053*) y 27 *loci* de 26 genes (dos *loci* del gen *mmpL3* interactúan con el *locus* *glgB\_839053*). En esta sección, observamos varios gaps (-) en la distribución alélica del *locus* 5, 762, 846 de *mprB* (histidina-proteína quinasa/fosfatasa de dos componentes *mprB*, MI = 0.4743), en el *locus* 5, 711, 087 de *mmpL5* (proteína transportadora de membrana *mmpL5*, MI = 0.4670) y en el *locus* 5, 715, 652 de *mmpL8* (proteína transportadora integral de membrana *mmpL8*, MI= 0.4435).

En la Figura 11, se presenta la distribución de alelos de los *loci* que interactúan con los *loci* 5, 934, 914 y 5, 936, 231 del gen *oppA*. La primera sección contiene las interacciones con el *locus* *oppA\_5934914*. Se describen varias regiones con gaps, por ejemplo, en los *loci* *ctpV\_375337* (MI=0.4516) y *eccB3\_562698* (MI = 0.417019). La Figura 11 también presenta patrones bien definidos de distribución de alelos entre *loci* que interactúan.

La primera columna de la Figura 10 y Figura 11 indica la institución que sometió los genomas de referencia (incluimos instituciones con menos de seis cepas en la categoría *otro*) y la segunda columna describe la resistencia genotípica a antibióticos (*genotypic antibiotic resistance*, GAR) que se caracterizó con RGI previamente. La filogenia estimada se incluye a la izquierda para mostrar la diversidad de las cepas. Por ejemplo, el clado inferior del árbol filogenético resalta por su cambio de nucleótidos con relación al resto de las cepas.

Se observó que la mayoría de las interacciones epistáticas reportadas en la Figura 10 y Figura 11 muestran cambios de sólo dos alelos en cada posición. Muy pocas posiciones mostraron cambios de tres alelos y presencia de gaps dentro de la distribución de la población bacteriana. Por ejemplo, la interacción de *oppA* con el gen *ppdk* en la posición 6, 164, 526 (*ppdk\_6164526*) describe cambios en la frecuencia alélica en su mayoría con T y en menor frecuencia con C y G (Figura 11).

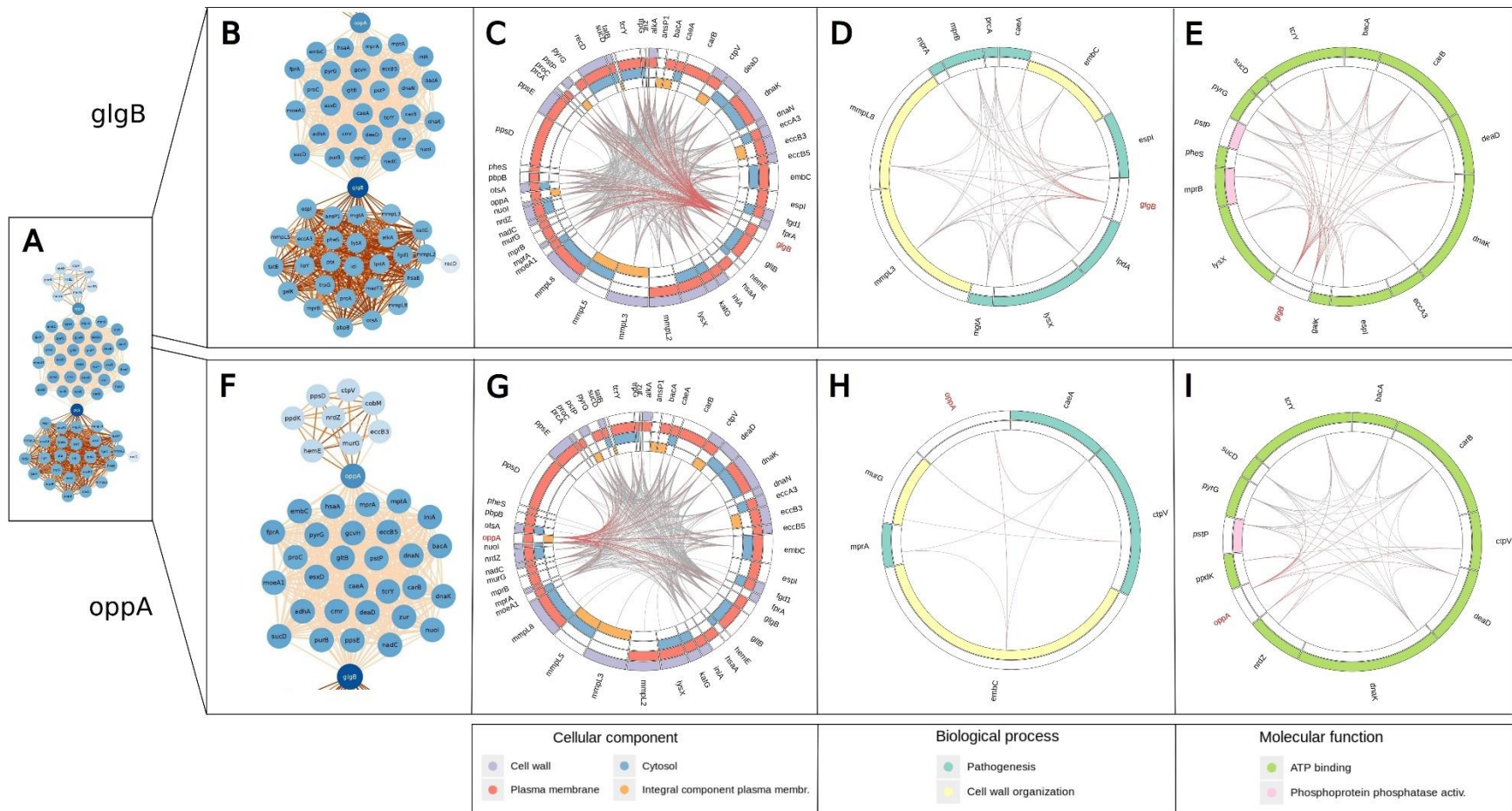


Figura 9. Enriquecimiento de términos de los genes que interactúan con *glgB* y *oppA*.

Las posiciones y el tamaño de los genes y los *loci* de las interacciones corresponden a la alineación de todo el pan-genoma. Las líneas rojas indican interacciones de *glgB* y *oppA*, mientras que las grises indican interacción de otros genes. (A) Red epistática. (B) Subred de interacciones de *glgB*. (C) Componentes celulares enriquecidos de genes que interactúan con *glgB*. (D) Procesos biológicos enriquecidos de genes que interactúan con *glgB*. (E) Enriquecimiento de funciones moleculares de los genes que interactúan con *glgB*. (F) Subred de interacciones de *oppA*. (G) Componentes celulares enriquecidos de genes que interactúan con *oppA*. (H) Enriquecimiento de procesos biológicos de genes que interactúan con *oppA*. (I) Funciones moleculares enriquecidas de genes que interactúan con *oppA*.

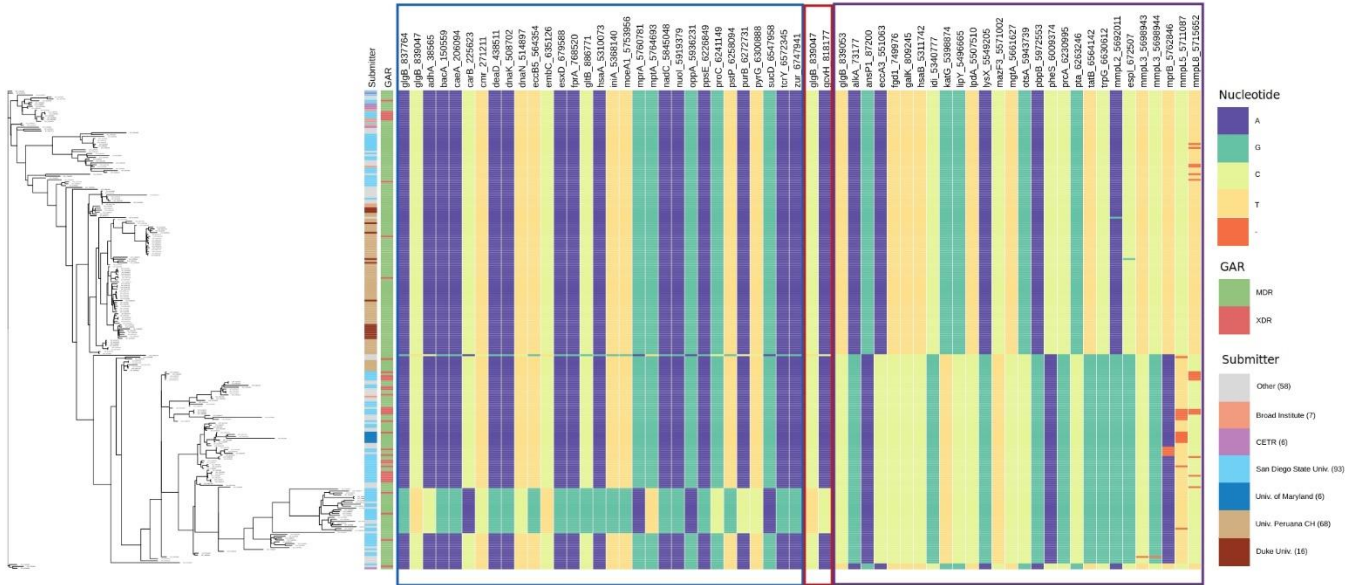


Figura 10. Distribución alélica de *loci* de genes que interactúan con *glgB*.

Las etiquetas de las columnas indican el nombre del gen y el *locus*. La filogenia estimada se incluye en la izquierda. Los *loci* que interactúan se organizan en secciones con bordes. El color del borde corresponde a las secciones en los archivos suplementarios con las interacciones de *glgB*. GAR= resistencia genotípica a antibióticos. Submitter= Institución.

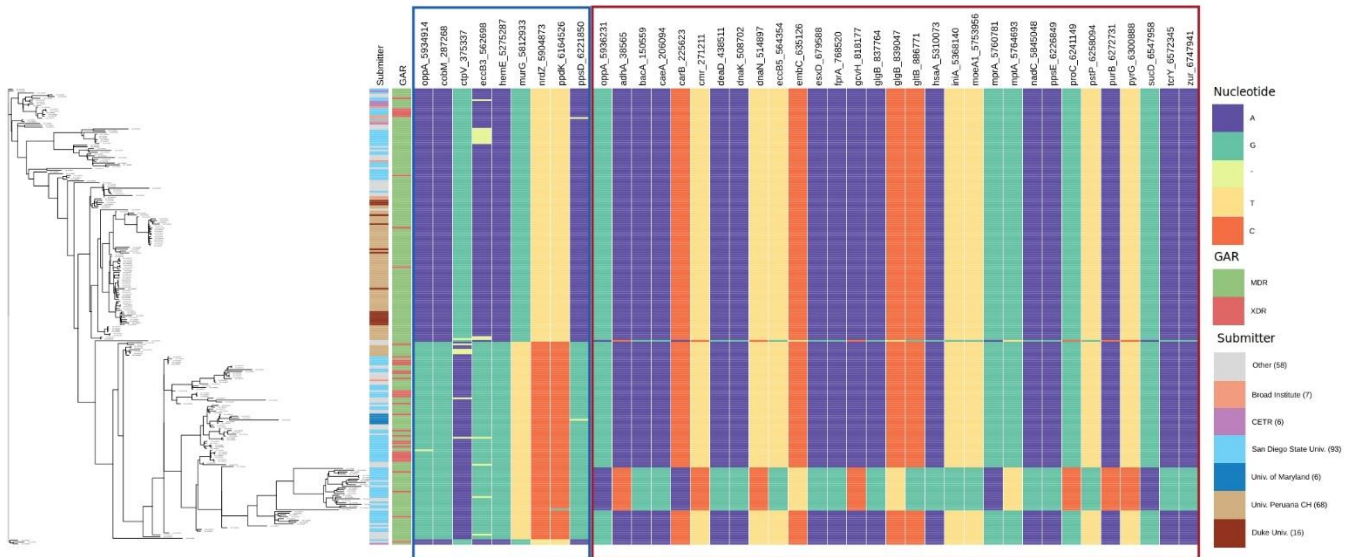


Figura 11. Distribución alélica de *loci* de genes que interactúan con el gen *oppA*.

Las etiquetas de las columnas indican el nombre del gen y el *locus*. La filogenia estimada se muestra a la izquierda de la figura. Los *loci* que interactúan se organizan en secciones con bordes. El color del borde corresponde a las secciones en los archivos suplementarios con las interacciones de *oppA*. GAR= resistencia genotípica a antibióticos. Submitter = institución.

La aplicación de esta metodología también permitió la identificación dentro de la red de varios *loci* en genes asociados a resistencia de antimicrobianos en *M. tb*. Recientemente, la OMS publicó el primer catálogo de variantes genéticas asociadas a resistencia (OMS, 2021). Esto ha permitido la identificación de múltiples posiciones relacionadas con este rasgo y su clasificación en cinco grupos. Por ello, se utilizó esta clasificación para encontrar genes en la red epistática (Tabla 5).

De los genes que identificamos en el catálogo de la OMS, se observó que el gen *fprA* (resistente a AMK y CM) y *embC* (resistente a ETB) interactúan con ambos genes, *glgB* y *oppA* (Tabla 5). Además, interactúan entre ellos, formando un clique<sup>28</sup> de cuatro genes (Figura 12).

Tabla 5. Genes de resistencia a los antibióticos reportados por el catálogo de la OMS encontrados en la red epistática y que interactúan con *glgB* y *oppA*.

Genes de resistencia a los antibióticos reportados por la OMS que interactúan con *glgB* y *oppA*. AMK = amikacina; BQD = bedaquilina; CM = capreomicina; CFZ = clofazimina; DLM= delamanida; ETB= etambutol; INH = isoniazid.

Gen	Gen de resistencia a los antibióticos (fármaco contra <i>M. tb</i> )
<i>glgB</i> y <i>oppA</i>	<i>fprA</i> (AMK)
<i>glgB</i> y <i>oppA</i>	<i>fprA</i> (CM)
<i>glgB</i> y <i>oppA</i>	<i>embC</i> (ETB)
<i>glgB</i>	<i>katG</i> (INH)
<i>glgB</i>	<i>fgd1</i> (DLM)
<i>glgB</i>	<i>mmpL5</i> (BQD)
<i>glgB</i>	<i>mmpL5</i> (CFZ)

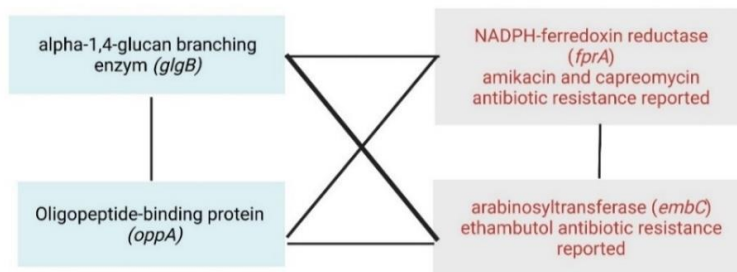


Figura 12. Interacciones epistáticas entre los genes de resistencia a los antibióticos *fprA* y *embC*, y los genes blancos tentativos de co-selección *glgB* y *oppA*.

En la imagen se muestra un mecanismo fuerte de interacciones.

<sup>28</sup> Un clique representa una red donde todos los nodos están completamente conectados entre sí, creando un fuerte mecanismo de interacciones.

## **CAPÍTULO 7. DISCUSIÓN**

### **7.1. Estimación de filogenia y resistencia a los antibióticos**

En este trabajo se caracterizó la presencia genotípica de MDR en el 100% de las cepas, lo que sugiere que una proporción importante de las variantes actualmente circulantes presentan un alto nivel de resistencia genotípica a los tratamientos de primera línea. Esto representa una dificultad en la respuesta exitosa al tratamiento aumentando la diseminación de cepas resistentes.

Por lo tanto, la detección de interacciones epistáticas para caracterizar *loci* polimórficos bajo la co-selección más fuerte es de suma importancia para la vigilancia molecular con herramientas bioinformáticas que nos ayuden a describir las cepas genotípicamente en un corto período de tiempo. Actualmente, se informa que el 3.4% de los nuevos pacientes con TB y el 20% de los pacientes con antecedentes de tratamiento para ésta fueron detectados como MDR en todo el mundo (WHO, 2021).

En el presente estudio, se describe que existen datos heterogéneos de instituciones que sometieron las secuencias y las cepas presentan considerable diversidad nucleotídica. Esta clasificación permitió continuar con el análisis de interacciones y unir estos resultados posteriormente en gráficas de distribución alélica (Figura 10 y Figura 11) que ayudaron a identificar patrones alélicos en las interacciones epistáticas.

### **7.2. Detección de interacciones epistáticas en los genes blanco de co-selección**

Gracias a la detección de interacciones epistáticas visualizadas en la Figura 7, se puede considerar a los genes *glgB* y *oppA* como blancos relevantes de co-selección, mismos que pueden estar asociados con varios genes relacionados con posibles resistencias farmacológicas (Figura 12).

El gen *glgB* codifica para una enzima involucrada en la biosíntesis de  $\alpha$ -glucano, que tiene un papel importante en la virulencia y patogenia de *M. tb*. Se reportó que las enzimas que participan en la síntesis de glucanos capsulares están involucradas en la interacción bacteria y hospedero, contribuyendo en la patogenicidad de la bacteria; mientras que los glucanos citosólicos son importantes fuentes de energía para la hibernación bacteriana y contribuyen al desarrollo de la TB latente. Lo anterior ha ganado importante atención como blanco farmacológico para *M. tb* (Dkhar et al., 2015).

La importancia del metabolismo de GlgB ha sido previamente descrita (de Smet et al., 2000); sin embargo, las interacciones con otros genes no habían sido reportadas. Por lo que las interacciones epistáticas mostradas en la Figura 7 abren la posibilidad de estudiar molecularmente estas interacciones para su caracterización fenotípica teniendo una aplicación potencial en el desarrollo de fármacos.

Por otro lado, el gen *oppA* codifica para una proteína de unión a oligopéptido y es responsable de la importación de péptidos. El sistema Opp es un casete transportador de unión a ATP, que ayuda en la absorción de péptidos, brindando los nutrientes esenciales como fuente de carbono, nitrógeno y aminoácidos. El sistema Opp afecta muchos procesos celulares incluyendo internalización de péptidos de detección de producción de biopelículas, modificación de la superficie celular y resistencia a los antibióticos (Hopfe et al., 2011).

La relevancia de la caracterización del sistema transportador de péptidos ha sido descrita por Dasgupta y colaboradores (2010) y recientemente se observó que este transportador de péptidos modula la respuesta inmune innata de MΦs infectados en *M. tb* (de Oliveira & Balan, 2020), pero las interacciones epistáticas de *oppA* con otros genes seguían siendo desconocidas hasta el presente proyecto de investigación.

Una bacteria es capaz de adaptar su respuesta a las condiciones del hospedero, tales como estrés oxidativo, hipoxia, fuente de carbono y nitrógeno. Por esta razón, evaluar las interacciones mediante experimentos bioinformáticos es necesario para la identificación de nuevas interacciones epistáticas en genes que han sido reportados previamente en bases de



datos, como el catálogo de la OMS, o para la comprensión de las interacciones epistáticas en *M. tb* antes del desarrollo de nuevas terapias.

### 7.3. Enriquecimiento de términos de GO para redes epistáticas

Se ha descrito que la funcionalidad de los componentes celulares de *M. tb* actúan como reguladores de la respuesta inmune en el hospedero. Por lo tanto, la descripción que se hace en este trabajo de genes involucrados en los componentes celulares, y que se encuentran bajo co-evolución, es crucial para comprender las interacciones de las bacterias con las moléculas en el hospedero que regulan la respuesta inmune (Li et al., 2020).

En estudios recientes, la relevancia de la estructura y de los genes relacionados con la biogénesis de *M. tb* que codifican a los glicoconjugados ha sido descrita con particular énfasis en la importancia de la membrana celular (Angala et al., 2014). Además, se ha subrayado previamente que la producción de ATP es crucial para la resistencia a los antibióticos en las bacterias (Black et al., 2014).

Tanto *glgB* como *oppA* interactúan epistáticamente con genes enriquecidos con los procesos biológicos de patogénesis y organización de la pared celular (Figura 9D y H). Uno de estos genes es *embC*, que codifica para una arabinosiltransferasa involucrada en la biosíntesis de un componente principal de la célula micobacteriana, el componente celular de lipoarabinomananas (LAM, *lipoarabinomannan*). El característico LAM cubierto de manosa de *M. tb* actúa como un patrón molecular asociado a patógenos (PAMPS *pathogen – associated molecular pattern*) modulando la activación de las células fagocíticas y la magnitud de la respuesta inmune inflamatoria del hospedero; mientras que también representa uno de los componentes principales en la organización de la pared celular (Goude et al., 2008) .

Además, se ha descrito que *embC* es expresado como parte de un ARNm policistrónico<sup>29</sup> controlado por una región promotora expresada diferencialmente en función de la fase de persistencia estacionaria o inducida por hipoxia de los bacilos, destacando el importante papel de esta proteína en las funciones biológicas de *M. tb* y la compleja interacción involucrada en

---

<sup>29</sup> ARN que transfiere el código genético procedente del ADN del núcleo celular a un ribosoma en el citoplasma.

la regulación de la pared celular (Goude et al., 2008). La interacción epistática de *embC* con *glgB* y *oppA* seguirá siendo una pregunta interesante que deberá abordarse en futuras investigaciones.

Acerca de las funciones moleculares, *oppA* y *glgB* interactúan con varios genes enriquecidos con la unión de ATP y con tres genes enriquecidos con actividad de fosfoproteína fosfatasa (Figura 9E e I). De estos genes, *bacA* es otro gen que interactúa con *glgB* y *oppA* en la red. El gen *bacA* codifica para una proteína de la familia tipo IV de exportadores tipo transportador ABC. Tiene función como importador de compuestos hidrofílicos de múltiples solutos como: vitamina B12, bleomicina y aminoglucósidos (de Oliveira & Balan, 2020).

También se ha demostrado que este transportador está implicado en el mantenimiento de la infección crónica en modelos murinos por mediar el transporte de una molécula que puede modular directa o indirectamente la respuesta proinflamatoria del hospedero. A pesar de tener estructuras diferentes, BacA y OppA comparten su capacidad para transportar una amplia gama de sustratos; en particular, participan en la capacidad de importar péptidos relacionados con la respuesta inmune innata sugiriendo una compleja regulación e interacción de estos transportadores, garantizando la necesidad de realizar estudios en el futuro cercano (de Oliveira & Balan, 2020; Rempel et al., 2020).

Los términos GO enriquecidos que pueden estar relacionados con los asociados con ATP sintasa en micobacterias son de particular interés porque contribuyen a la producción eficiente de ATP, esta enzima ha sido validada como un objetivo potencial con aplicaciones farmacológicas. Además, el ATP micobacteriano sintasa y sus características pueden proporcionar información sobre adaptaciones del metabolismo energético bacteriano (Lu et al., 2014).

*M. tb* puede sobrevivir en MΦs humanos durante un tiempo prolongado, para esta y otras cepas de micobacterias patógenas, el bloqueo de la hidrólisis de ATP es relevante ya que puede representar una adaptación a los fagosomas humanos (Lu et al., 2014). Por lo tanto, el estudio de las interacciones epistáticas asociadas con la producción de ATP en *M. tb* puede ayudar a

comprender cómo la bacteria sobrevive en condiciones de baja energía, como limitación de oxígeno y nutrientes dentro del hospedero.

En cuanto a las interacciones de genes enriquecidos con actividad fosfoproteína fosfatasa (Figura 9E e I), tanto *oppA* como *glgB* establecen interacción epistática con el gen *pstP*, que codifica la proteína serina/treonina fosfatasa (PstP) de *M. tb*. Esta proteína tiene un papel central en el control de procesos vitales como un regulador negativo de la actividad cinasa y fosforilación de serina y treonina (Iswahyudi et al., 2019). Se han descrito otras funciones de PstP como regulador de la síntesis de la pared celular y la división celular por desfosforilación de sustratos clave y también se ha asociado con dependencia de  $Mn^{2+}$ , elemento necesario para la biosíntesis de la pared celular (Sharma et al., 2016).

PstP está presente como una fosfatasa transmembrana (Boitel et al., 2003), pero aún no se ha descrito si durante la infección y la activación de los mecanismos de la respuesta inmune innata (p. ej., activación del estallido respiratorio) la bacteria degradada retiene la actividad de fosfatasa en los fragmentos de membrana; estos podrían contribuir a la desfosforilación de las vías de señalización y a la evasión de la respuesta inmune innata.

Se enriquecieron cuatro componentes celulares para genes que interactúan con *glgB* (Figura 9C) y *oppA* (Figura 9G). La mayoría de los genes codifican proteínas de la membrana plasmática. Por ejemplo, cuatro genes del grupo de proteínas largas de membrana plasmática (*membrane protein large*, MmpL: *mmpL2*, *mmpL3*, *mmpL5* y *mmpL8*) tienen interacciones epistáticas con *glgB* y *oppA*. Estas proteínas exportan componentes de la membrana celular al espacio periplásmico (como los lípidos asociados a la virulencia y moléculas de bajo peso molecular como los sideróforos<sup>30</sup>), contribuyendo en un alto nivel a la persistencia de *M. tb* en el hospedero (Melly y Purdy, 2019). Será necesario un nuevo estudio para investigar las asociaciones fenotípicas entre las interacciones epistáticas y los sistemas de importación y exportación en la membrana.

---

<sup>30</sup> Compuesto secretado por los microorganismos quelante de hierro.

#### **7.4. Genes reportados por el catálogo de la OMS en interacción con *glgB* y *oppA*.**

Un interesante descubrimiento fue la identificación de interacciones epistáticas entre *glgB* y *oppA* con genes asociados a resistencia y recientemente reportados por el catálogo de la OMS (Tabla 5). Así, los genes *katG*, *fprA*, *embC*, *fgd1* y *mmpL5* se relacionaron con *glgB* y *oppA*, genes blancos tentativos de co-selección.

Se encontró una interacción entre *glgB* y *katG*. Al respecto se ha reportado, que el gen *katG* confiere monorresistencia a INH y mutaciones son comunes debido a que esta ha estado en uso clínico desde los años 1950. Sin embargo, las pruebas de resistencia a este fármaco sólo se están implementando en pruebas especializadas (por ejemplo, Xpert MTB/XDR) y no están disponibles de forma rutinaria de tal manera que, si existe la resistencia, no se detecta. En estos pacientes se presenta un alto riesgo de fracaso al tratamiento (Sulis y Pai, 2020).

*M. tb* depende en gran medida de la disponibilidad de hierro y cofactores para su crecimiento y supervivencia. Los sistemas de reducción del bacilo son llamados oxidorreductasas férricas, que utilizan cofactores redox intracelulares para reducir el  $\text{Fe}^{3+}$  a la forma ferrosa ( $\text{Fe}^{2+}$ ) para la biosíntesis de proteínas de hierro. El gen *fprA* codifica proteínas para la reducción del hierro, la activación de la reductasa ribonucleótida, la respuesta al estrés del oxígeno (dentro del fagosoma del hospedero) y la reducción de citocromos P450 (Fischer et al., 2002). Sin embargo, hasta ahora no se habían descrito las interacciones epistáticas tentativas de *fprA* con otros genes.

La bedaquilina (BQD) y delamanida (DLM) son ejemplos de moléculas farmacológicas desarrolladas en el 2005 y 2003 respectivamente para mejorar el tratamiento de cepas MDR y XDR. En este sentido, recientemente se han reportado mutaciones en el gen *fgd1* encontradas en casos de resistencia fenotípica al fármaco DLM (Reichmuth et al., 2020). Lo anterior abre las pautas para continuar con la caracterización de interacciones epistáticas en este gen.

Por otro lado, el gen *mmpL5* codifica una proteína transmembrana transportadora y se ha sugerido que genera resistencia a varios fármacos, por ejemplo, a los azoles, transportando el fármaco fuera de la célula como un mecanismo de salida (Briffotiaux et al., 2017). Continuar con el estudio de interacciones epistáticas en las que estén involucradas la proteína MmpL podría aportar nuevo conocimiento referente a los mecanismos de transporte de la bacteria.

Las interacciones epistáticas descubiertas entre los genes mencionados nos motivaron a visualizar futuros estudios para probar nuevas hipótesis experimentales y dilucidar sus efectos biológicos (fenotípicos) y farmacológicos. Así, el método aplicado de SpydrPick permitió caracterizar nuevas interacciones entre genes relevantes en *M. tb* y podría tener un enfoque muy exitoso para conducir a nuevo conocimiento de co-selección en poblaciones bacterias.

## CAPÍTULO 8. CONCLUSIONES

- El enfoque desarrollado permitió identificar nuevas interacciones epistáticas con implicaciones en virulencia, patogenia, moduladores del sistema de transporte, de la respuesta inmune y resistencia genotípica a los antibióticos.
- En el análisis de la red epistática, se identificaron *glgB* y *oppA* como objetivos tentativos de co-selección, ya que, por ejemplo, estos dos genes interactúan epistáticamente con *fprA* y *embC*, genes reportados en el catálogo de la OMS como resistentes a etambutol (*embC*) y amikacina y capreomicina (*fprA*).
- Los resultados destacan la importancia de implementar enfoques computacionales para dilucidar nuevos genes asociados a interacciones epistáticas tentativas en *M. tb*.

Dado lo anterior, consideramos que se han cumplido los objetivos planteados al inicio de este trabajo y se puede confirmar que el análisis computacional de genomas completos permite identificar redes de interacciones epistáticas entre genes asociados a diferentes rasgos biológicos.

## CAPÍTULO 9. PERSPECTIVAS

Se enlistan algunas ideas para continuar con la presente investigación:

- Ampliar el análisis de la red de interacciones epistáticas para encontrar nuevos genes relevantes involucrados en diferentes rasgos biológicos.
- Comprobar las interacciones epistáticas descritas para los genes *gigB* y *oppA* con metodologías de biología molecular y así comprobar biológicamente dichas interacciones.
- Extender los estudios de interacciones epistáticas con otras metodologías bioinformáticas que ayuden a describir y comprender los mecanismos biológicos de *M. tb* y su asociación con la resistencia farmacológica.

Se espera que esta investigación contribuya a mejorar nuestro conocimiento sobre la *M. tb* y la TB.

## CAPÍTULO 10. REFERENCIAS

- Abel, L., El-Baghdadi, J., Bousfiha, A. A., Casanova, J. L., & Schurr, E. (2014). Human genetics of tuberculosis: A long and winding road. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 369, Issue 1645). Royal Society.  
<https://doi.org/10.1098/rstb.2013.0428>
- Allard B.A., Panariti, A. Martin J.G. (2018). Alveolar Macrophages in the resolution of inflammation, tissue repair, and tolerance to infection. *Front. Immunol.* 9:1777  
doi: 10.3389/fimmu.2018.01777
- Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A. L. v., Cheng, A. A., Liu, S., Min, S. Y., Miroshnichenko, A., Tran, H. K., Werfalli, R. E., Nasir, J. A., Oloni, M., Speicher, D. J., Florescu, A., Singh, B., ... McArthur, A. G. (2020). CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*, 48(D1), D517–D525. <https://doi.org/10.1093/nar/gkz935>
- Angala, S. K., Belardinelli, J. M., Huc-Claustre, E., Wheat, W. H., & Jackson, M. (2014). The cell envelope glycoconjugates of Mycobacterium tuberculosis. In *Critical Reviews in Biochemistry and Molecular Biology* (Vol. 49, Issue 5, pp. 361–399). Informa Healthcare.  
<https://doi.org/10.3109/10409238.2014.925420>
- Arish, M., & Naz, F. (2022). Macrophage plasticity as a therapeutic target in tuberculosis. *European Journal of Immunology*. <https://doi.org/10.1002/eji.202149624>
- Arnold, B. J., Gutmann, M. U., Grad, Y. H., Sheppard, S. K., Corander, J., Lipsitch, M., & Hanage, W. P. (2018). Weak epistasis may drive adaptation in recombining bacteria. *Genetics*, 208(3), 1247–1260. <https://doi.org/10.1534/genetics.117.300662>
- Augusto Salomone, C. (2019). TBC: su protagonismo en el arte y la cultura. *Revista Del Torax*.
- Aylor, D. L., & Zeng, Z. B. (2008). From classical genetics to quantitative genetics to systems biology: Modeling epistasis. *PLoS Genetics*, 4(3).  
<https://doi.org/10.1371/journal.pgen.1000029>
- Balaban, N. Q., Helaine, S., Lewis, K., Ackermann, M., Aldridge, B., Andersson, D. I., Brynildsen, M. P., Bumann, D., Camilli, A., Collins, J. J., Dehio, C., Fortune, S., Ghigo, J. M., Hardt, W. D., Harms, A., Heinemann, M., Hung, D. T., Jenal, U., Levin, B. R., ...



- Zinkernagel, A. (2019). Definitions and guidelines for research on antibiotic persistence. *Nature Reviews Microbiology*, 17(7), 441–448. <https://doi.org/10.1038/s41579-019-0196-3>
- Balakrishnan, V. S. (2017). Tuberculosis—know thy enemy. *The Lancet Respiratory Medicine*, 5(10), 778. [https://doi.org/10.1016/s2213-2600\(17\)30355-7](https://doi.org/10.1016/s2213-2600(17)30355-7)
- Bateson, W. (1906). The progress of genetics since the rediscovery of mendel's235 papers, *Progressus rei botanicae* 1, 368.
- Bini, E. I. (2015). Tesis Inmunidad y esteroides sexuales en la tuberculosis estudios clínicos y experimentales. Rosario, Argentina: Universidad Nacional de Rosario.
- Black, P. A., Warren, R. M., Louw, G. E., van Helden, P. D., Victor, T. C., & Kana, B. D. (2014). Energy metabolism and drug efflux in mycobacterium tuberculosis. In *Antimicrobial Agents and Chemotherapy* (Vol. 58, Issue 5, pp. 2491–2503). American Society for Microbiology. <https://doi.org/10.1128/AAC.02293-13>
- Boitel, B., Ortiz-Lombardía, M., Durán, R., Pompeo, F., Cole, S. T., Cerveñansky, C., & Alzari, P. M. (2003). PknB kinase activity is regulated by phosphorylation in two Thr residues and dephosphorylation by PstP, the cognate phospho-Ser/Thr phosphatase, in Mycobacterium tuberculosis. *Molecular Microbiology*, 49(6), 1493–1508. <https://doi.org/10.1046/j.1365-2958.2003.03657.x>
- Borowiec, M. L. (2016). AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ*, 2016(1). <https://doi.org/10.7717/peerj.1660>
- Borrell, S., Teo, Y., Giardina, F., Streicher, E. M., Klopper, M., Feldmann, J., Müller, B., Victor, T. C., & Gagneux, S. (2013). Epistasis between antibiotic resistance mutations drives the evolution of extensively drug-resistant tuberculosis. *Evolution, Medicine and Public Health*, 2013(1), 65–74. <https://doi.org/10.1093/emph/eot003>
- Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. In *Cell* (Vol. 169, Issue 7, pp. 1177–1186). Cell Press. <https://doi.org/10.1016/j.cell.2017.05.038>
- Briffotiaux, J., Huang, W., Wang, X., & Gicquel, B. (2017). MmpS5/MmpL5 as an efflux pump in Mycobacterium species. *Tuberculosis*, 107, 13-19.
- Brosch, R., Gordon, S. v, Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L. M., Pym, A. S., Samper, S., van

- Soolingen, D., & Cole, S. T. (2002). *A new evolutionary scenario for the Mycobacterium tuberculosis complex* (Vol. 99, Issue 6). [www.pnas.org/cgi/doi/10.1073/pnas.052548299](http://www.pnas.org/cgi/doi/10.1073/pnas.052548299)
- Brown, E. D., & Wright, G. D. (2016). Antibacterial drug discovery in the resistance era. In *Nature* (Vol. 529, Issue 7586, pp. 336–343). Nature Publishing Group.  
<https://doi.org/10.1038/nature17042>
- Bustamante, J., Boisson-Dupuis, S., Abel, L., & Casanova, J. L. (2014). Mendelian susceptibility to mycobacterial disease: Genetic, immunological, and clinical features of inborn errors of IFN- $\gamma$  immunity. In *Seminars in Immunology* (Vol. 26, Issue 6, pp. 454–470). Academic Press. <https://doi.org/10.1016/j.smim.2014.09.008>
- Campbell, I. A., & Bah-Sow, O. (2006). Pulmonary tuberculosis: diagnosis and treatment. *Bmj*, 332(7551), 1194-1197. <https://doi.org/10.1136/bmj.332.7551.1194>
- Campbell, R. F., McGrath, P. T., & Paaby, A. B. (2018). Analysis of Epistasis in Natural Traits Using Model Organisms. In *Trends in Genetics* (Vol. 34, Issue 11, pp. 883–898). Elsevier Ltd. <https://doi.org/10.1016/j.tig.2018.08.002>
- Cartes Parra, J. C. (2013). Breve Historia de la tuberculosis. *Revista Médica de Costa Rica y Centroamerica LXX* (605) 145-150.
- CDC. (2021). *Centros para el Control y la Prevención de Enfermedades*.  
<http://www.cdc.gov/tb/esp/>.
- CDC. (2014). *Preguntas y respuestas sobre la tuberculosis 2014 DEPARTAMENTO DE SALUD Y SERVICIOS HUMANOS DE EE. UU. Centros para el Control y la Prevención de Enfermedades*. <http://www.cdc.gov/tb/esp/>.
- Chanda, P., Costa, E., Hu, J., Sukumar, S., van Hemert, J., & Walia, R. (2020). Information theory in computational biology: Where we stand today. *Entropy*, 22(6).  
<https://doi.org/10.3390/E22060627>
- Cohen, K. A., Abeel, T., Manson McGuire, A., Desjardins, C. A., Munsamy, V., Shea, T. P., Walker, B. J., Bantubani, N., Almeida, D. v., Alvarado, L., Chapman, S. B., Mvelase, N. R., Duffy, E. Y., Fitzgerald, M. G., Govender, P., Gujja, S., Hamilton, S., Howarth, C., Larimer, J. D., ... Earl, A. M. (2015). Evolution of Extensively Drug-Resistant Tuberculosis over Four Decades: Whole Genome Sequencing and Dating Analysis of Mycobacterium tuberculosis Isolates from KwaZulu-Natal. *PLoS Medicine*, 12(9).  
<https://doi.org/10.1371/journal.pmed.1001880>

- Cohen, K. A., Manson, A. L., Desjardins, C. A., Abeel, T., & Earl, A. M. (2019). Deciphering drug resistance in *Mycobacterium tuberculosis* using whole-genome sequencing: Progress, promise, and challenges. In *Genome Medicine* (Vol. 11, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s13073-019-0660-8>
- Cole S T, Brosh R, Parkhill T G, Churcher D H, Gordon S V, Eigmeir K, Gas S, Barrt C E, Tejaia F, Badcock K, Basham D, Brown D, Chillingworth R, Connor R, Davies R, Devin K, Feltwell T, Gentle S, Hamlin N, ... Barrell G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 393.
- Comas, Í., Chakravarti, J., Small, P. M., Galagan, J., Niemann, S., Kremer, K., Ernst, J. D., & Gagneux, S. (2010). Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nature Genetics*, 42(6), 498–503. <https://doi.org/10.1038/ng.590>
- Comas I, & Gagneux S. (2011). A role for systems epidemiology in tuberculosis research. In *Trends in Microbiology* (Vol. 19, Issue 10, pp. 492–500). <https://doi.org/10.1016/j.tim.2011.07.002>
- Conrad, W. H., Osman, M. M., Shanahan, J. K., Chu, F., Takaki, K. K., Cameron, J., Hopkinson-Woolley, D., Brosch, R., & Ramakrishnan, L. (2017). Mycobacterial ESX-1 secretion system mediates host cell lysis through bacterium contact-dependent gross membrane disruptions. *Proceedings of the National Academy of Sciences of the United States of America*, 114(6), 1371–1376. <https://doi.org/10.1073/pnas.1620133114>
- Corbett E., M. B. , C. G. , D. C. K. (2006). Tuberculosis in sub-Saharan Africa: opportunities, challenges, and change in the era of antiretroviral treatment. *The Lancet*, 367. <http://www.theglobalfund.org/en/>
- Cover, T. M. (1991). *Elements of Information Theory*. Wiley-Interscience.
- de los Campos, G., Vazquez, A. I., Hsu, S., & Lello, L. (2018). Complex-Trait Prediction in the Era of Big Data. In *Trends in Genetics* (Vol. 34, Issue 10, pp. 746–754). Elsevier Ltd. <https://doi.org/10.1016/j.tig.2018.07.004>
- de Oliveira, M. C. B., & Balan, A. (2020). The atp-binding cassette (Abc) transport systems in mycobacterium tuberculosis: Structure, function, and possible targets for therapeutics. In *Biology* (Vol. 9, Issue 12, pp. 1–25). MDPI AG. <https://doi.org/10.3390/biology9120443>

- de Smet, K. A. L., Weston, † Anthony, Brown, I. N., Young, D. B., & Robertson, B. D. (2000). Three pathways for trehalose biosynthesis in mycobacteria. In *Microbiology* (Vol. 146). [www.sanger.ac.uk](http://www.sanger.ac.uk)
- Dasgupta, A., Sureka, K., Mitra, D., Saha, B., Sanyal, S., Das, A. K., ... & Basu, J. (2010). An oligopeptide transporter of *Mycobacterium tuberculosis* regulates cytokine release and apoptosis of infected macrophages. *PloS one*, 5(8), e12225.
- Dkhar, H. K., Gopalsamy, A., Loharch, S., Kaur, A., Bhutani, I., Saminathan, K., Bhagyaraj, E., Chandra, V., Swaminathan, K., Agrawal, P., Parkesh, R., & Gupta, P. (2015). Discovery of *Mycobacterium tuberculosis*  $\alpha$ -1,4-glucan branching enzyme (GlgB) inhibitors by structure-And ligand-based virtual screening. *Journal of Biological Chemistry*, 290(1), 76–89. <https://doi.org/10.1074/jbc.M114.589200>
- Domenech Pila, B. C. C. S. (2001). *Mycobacterium tuberculosis* in the post-genomic age. *Current Opinion in Microbiology*. <http://www.pasteur.fr/recherche/unites/Lgmb/>
- Drost, H. G., & Paszkowski, J. (2017). Biomart: Genomic data retrieval with R. *Bioinformatics*, 33(8), 1216–1217. <https://doi.org/10.1093/bioinformatics/btw821>
- Durão, P., Trindade, S., Sousa, A., & Gordo, I. (2015). Multiple resistance at no cost: Rifampicin and streptomycin a dangerous liaison in the spread of antibiotic resistance. *Molecular Biology and Evolution*, 32(10), 2675–2680. <https://doi.org/10.1093/molbev/msv143>
- Fatima, S., Kumari, A., Das, G., & Dwivedi, V. P. (2020). Tuberculosis vaccine: A journey from BCG to present. In *Life sciences* (Vol. 252, p. 117594). NLM (Medline). <https://doi.org/10.1016/j.lfs.2020.117594>
- Fischer, F., Raimondi, D., Aliverti, A., & Zanetti, G. (2002). *Mycobacterium tuberculosis* FprA, a novel bacterial NADPH-ferredoxin reductase. vol. 269. *Eur J Biochem*, 1432-1033.
- Fisher, R. A. (1919). The correlation between relatives on the supposition of medelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 399–433.
- Frieden T, Sterling T, Pablos- Mendez A, Kilburn J, Cauthen G, & Dooley S. (1993). The emergence of drug resistant tuberculosis in new york city. *The New England Journal of Medicine*, 328(8).

- Gagneux, S. (2017). Strain Variation in the Mycobacterium tuberculosis Complex: Its Role in Biology, Epidemiology and Control. In *Advances in Experimental Medicine and Biology*. <http://www.springer.com/series/5584>
- Galagan, J. E. (2014). Genomic insights into tuberculosis. In *Nature Reviews Genetics* (Vol. 15, Issue 5, pp. 307–320). Nature Publishing Group. <https://doi.org/10.1038/nrg3664>
- Ge, G., Jiang, H., Xiong, J., Zhang, W., Shi, Y., Tao, C., & Wang, H. (2021). Progress of the Art of Macrophage Polarization and Different Subtypes in Mycobacterial Infection. In *Frontiers in Immunology* (Vol. 12). Frontiers Media S.A. <https://doi.org/10.3389/fimmu.2021.752657>
- Gilpin, C., & Mirzayev, F. (2019). Tuberculosis Supranational Reference Laboratories: A Global Approach. In *Clinics in Chest Medicine* (Vol. 40, Issue 4, pp. 755–762). W.B. Saunders. <https://doi.org/10.1016/j.ccm.2019.07.005>
- Gordon, S. v, Heym, B., Parkhil, J., Cole', S. T., & Barrel13, B. (1999). New insertion sequences and a novel repeated sequence in the genome of Mycobacterium tuberculosis H37Rv. *Microbiology*, 145, 881–892.
- Goude, R., Amin, A. G., Chatterjee, D., & Parish, T. (2008). The critical role of embC in Mycobacterium tuberculosis. *Journal of Bacteriology*, 190(12), 4335–4341. <https://doi.org/10.1128/JB.01825-07>
- Hadfield, J., Croucher, N. J., Goater, R. J., Abudahab, K., Aanensen, D. M., & Harris, S. R. (2018). Phandango: An interactive viewer for bacterial population genomics. *Bioinformatics*, 34(2), 292–293. <https://doi.org/10.1093/bioinformatics/btx610>
- Hermisson, J., Hansen, T. F., & Wagner, G. P. (2003). the american naturalist may. In *Am. Nat* (Vol. 161, Issue 5).
- Hopfe, M., Dahlmanns, T., & Henrich, B. (2011). In Mycoplasma hominis the OppA-mediated cytoadhesion depends on its ATPase activity. *BMC Microbiology*, 11. <https://doi.org/10.1186/1471-2180-11-185>
- Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), 44–57. <https://doi.org/10.1038/nprot.2008.211>
- Iswahyudi, Mukamolova, G. v., Straatman-Iwanowska, A. A., Allcock, N., Ajuh, P., Turapov, O., & O'Hare, H. M. (2019). Mycobacterial phosphatase PstP regulates global serine threonine

- phosphorylation and cell division. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-44841-9>
- Jiao, X., Sherman, B. T., Huang, D. W., Stephens, R., Baseler, M. W., Lane, H. C., & Lempicki, R. A. (2012). DAVID-WS: A stateful web service to facilitate gene/protein list analysis. *Bioinformatics*, 28(13), 1805–1806. <https://doi.org/10.1093/bioinformatics/bts251>
- Jiménez-Ruano, A. C., Madrazo-Moya, C. F., Cancino-Muñoz, I., Mejía-Ponce, P. M., Licona-Cassani, C., Comas, I., Muñoz-Salazar, R., & Zenteno-Cuevas, R. (2021). Whole genomic sequencing based genotyping reveals a specific X3 sublineage restricted to Mexico and related with multidrug resistance. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-020-80919-5>
- Juarez-Eusebio, D. M., Munro-Rojas, D., Muñoz-Salazar, R., Laniado-Laborín, R., Martínez-Guarneros, J. A., Flores-López, C. A., & Zenteno-Cuevas, R. (2017). Molecular characterization of multidrug-resistant *Mycobacterium tuberculosis* isolates from high prevalence tuberculosis states in Mexico. *Infection, Genetics and Evolution*, 55, 384–391. <https://doi.org/10.1016/j.meegid.2016.09.012>
- Kavvas, E. S., Catoi, E., Mih, N., Yurkovich, J. T., Seif, Y., Dillon, N., Heckmann, D., Anand, A., Yang, L., Nizet, V., Monk, J. M., & Palsson, B. O. (2018). Machine learning and structural analysis of *Mycobacterium tuberculosis* pan-genome identifies genetic signatures of antibiotic resistance. *Nature Communications*, 9(1). <https://doi.org/10.1038/s41467-018-06634-y>
- Kohl, M., Wiese, S., and Warscheid, B. (2011). “Cytoscape: Software for Visualization and Analysis of Biological Networks,” in *Data Mining in Proteomics* (Springer), 291–303. doi:10.1007/978-1-60761-987-1\_18
- Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35(21), 4453–4455. <https://doi.org/10.1093/bioinformatics/btz305>
- Letunic, I., & Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49(W1), W293–W296. <https://doi.org/10.1093/nar/gkab301>

- Li, L., Lv, J., He, Y., & Wang, Z. (2020). Gene network in pulmonary tuberculosis based on bioinformatic analysis. *BMC Infectious Diseases*, *20*(1). <https://doi.org/10.1186/s12879-020-05335-6>
- Lundgren, M., Charpentier, E., & Fineran, P. C. (2015). *CRISPR Methods and Protocols Methods in Molecular Biology*. <http://www.springer.com/series/7651>
- Lu, P., Lill, H., & Bald, D. (2014). ATP synthase in mycobacteria: Special features and implications for a function as drug target. In *Biochimica et Biophysica Acta - Bioenergetics* (Vol. 1837, Issue 7, pp. 1208–1218). Elsevier B.V. <https://doi.org/10.1016/j.bbabi.2014.01.022>
- Martín, C., Marinova, D., Aguiló, N., & Gonzalo-Asensio, J. (2021). MTBVAC, a live TB vaccine poised to initiate efficacy trials 100 years after BCG. *Vaccine*, *39*(50), 7277–7285. <https://doi.org/10.1016/j.vaccine.2021.06.049>
- McInerney, J. O., McNally, A., & O'Connell, M. J. (2017). Why prokaryotes have pangenomes. In *Nature Microbiology* (Vol. 2). Nature Publishing Group. <https://doi.org/10.1038/nmicrobiol.2017.40>
- Meehan, C. J., Goig, G. A., Kohl, T. A., Verboven, L., Dippenaar, A., Ezewudo, M., Farhat, M. R., Guthrie, J. L., Laukens, K., Miotto, P., Ofori-Anyinam, B., Dreyer, V., Supply, P., Suresh, A., Utpatel, C., van Soolingen, D., Zhou, Y., Ashton, P. M., Brites, D., ... van Rie, A. (2019). Whole genome sequencing of Mycobacterium tuberculosis: current standards and open issues. In *Nature Reviews Microbiology* (Vol. 17, Issue 9, pp. 533–545). Nature Publishing Group. <https://doi.org/10.1038/s41579-019-0214-5>
- Melly, G., & Purdy, G. E. (2019). Mmpl proteins in physiology and pathogenesis of m. Tuberculosis. *Microorganisms*, *7*(3). <https://doi.org/10.3390/microorganisms7030070>
- Migliori, G. B., Sotgiu, G., D'Ambrosio, L., Centis, R., Lange, C., Bothamley, G., Cirillo, D. M., de Lorenzo, S., Guenther, G., Kliiman, K., Muetterlein, R., Spinu, V., Villar, M., Zellweger, J. P., Sandgren, A., Huitric, E., & Manissero, D. (2012). TB and MDR/XDR-TB in European Union and European Economic Area countries: Managed or mismanaged? *European Respiratory Journal*, *39*(3), 619–625. <https://doi.org/10.1183/09031936.00170411>
- Morales-Arce, A. Y., Sabin, S. J., Stone, A. C., & Jensen, J. D. (2021). The population genomics of within-host Mycobacterium tuberculosis. In *Heredity* (Vol. 126, Issue 1). Springer Nature. <https://doi.org/10.1038/s41437-020-00377-7>

- Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A., & Parkhill, J. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, *31*(22), 3691–3693.  
<https://doi.org/10.1093/bioinformatics/btv421>
- Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, J. A., & Harris, S. R. (2016). SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial Genomics*, *2*(4), e000056. <https://doi.org/10.1099/mgen.0.000056>
- Pai, M., Behr, M. A., Dowdy, D., Dheda, K., Divangahi, M., Boehme, C. C., Ginsberg, A., Swaminathan, S., Spigelman, M., Getahun, H., Menzies, D., & Raviglione, M. (2016). Tuberculosis. In *Nature Reviews Disease Primers* (Vol. 2). Nature Publishing Group.  
<https://doi.org/10.1038/nrdp.2016.76>
- Pascual-Pareja, J. F., Carrillo-Gómez, R., Hontañón-Antoñana, V., & Martínez-Prieto, M. (2018). Treatment of pulmonary and extrapulmonary tuberculosis. *Enfermedades Infecciosas y Microbiología Clínica*, *36*(8), 507–516.  
<https://doi.org/10.1016/j.eimc.2017.10.018>
- Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R. P., Moret, B. M. E., & Stamatakis, A. (2010). How many bootstrap replicates are necessary? *Journal of Computational Biology*, *17*(3), 337–354. <https://doi.org/10.1089/cmb.2009.0179>
- Paneque Ramos E, Rojas Rodríguez LY, Pérez Loyola M. La Tuberculosis a través de la Historia: un enemigo de la humanidad. Rev haban cienc méd [Internet] 2018 [citado];17(3):[353-363]. Disponible en:  
<http://www.revhabanera.sld.cu/index.php/rhab/article/view/2058>
- Pensar, J., Puranen, S., Arnold, B., MacAlasdair, N., Kuronen, J., Tonkin-Hill, G., Pesonen, M., Xu, Y., Sipola, A., Sánchez-Busó, L., Lees, J. A., Chewapreecha, C., Bentley, S. D., Harris, S. R., Parkhill, J., Croucher, N. J., & Corander, J. (2019). Genome-wide epistasis and co-selection study using mutual information. *Nucleic Acids Research*, *47*(18), E112.  
<https://doi.org/10.1093/NAR/GKZ656>.
- Plomp, K. A., Roberts, C. A., Elton, S., & Bentley, G. R. (Eds.). (2022). *Palaeopathology and Evolutionary Medicine: An Integrated Approach*. Oxford University Press.
- Pyle MM. Relative numbers of resistant tubercle bacilli in sputa of patients before and during treatment with streptomycin. Proc Staff Meet Mayo Clin. 1947 Oct 15;22(21):465-73.



- Ramaswamy, S., & Musser, J. M. (1998). *Molecular genetic basis of antimicrobial agent resistance in Mycobacterium tuberculosis : 1998 update.*
- Reichmuth, M. L., Hömke, R., Zürcher, K., Sander, P., Avihingsanon, A., Collantes, J., ... & Keller, P. M. (2020). Natural polymorphisms in Mycobacterium tuberculosis conferring resistance to delamanid in drug-naive patients. *Antimicrobial agents and chemotherapy*, 64(11), e00513-20.
- Rempel, S., Gati, C., Nijland, M., Thangaratnarajah, C., Karyolaimos, A., de Gier, J. W., Guskov, A., & Slotboom, D. J. (2020). A mycobacterial ABC transporter mediates the uptake of hydrophilic compounds. *Nature*, 580(7803), 409–412.  
<https://doi.org/10.1038/s41586-020-2072-8>
- Rothschild, B. M., Martin, L. D., Lev, G., Bercovier, H., Kahila Bar-Gal, G., Greenblatt, C., Donoghue, H., Spigelman, M., & Brittain, D. (2001). Mycobacterium tuberculosis Complex DNA from an Extinct Bison Dated 17,000 Years before the Present. In *Origins of Tuberculosis in North America • CID (Vol. 33)*.  
<https://academic.oup.com/cid/article/33/3/305/277074>
- Sackton, T. B., & Hartl, D. L. (2016). Genotypic Context and Epistasis in Individuals and Populations. In *Cell (Vol. 166, Issue 2, pp. 279–287)*. Cell Press.  
<https://doi.org/10.1016/j.cell.2016.06.047>
- Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
- Seung, K. J., Keshavjee, S., & Rich, M. L. (2015). Multidrug-resistant tuberculosis and extensively drug-resistant tuberculosis. *Cold Spring Harbor Perspectives in Medicine*, 5(9).  
<https://doi.org/10.1101/cshperspect.a017863>
- Sharma, A. K., Arora, D., Singh, L. K., Gangwal, A., Sajid, A., Molle, V., Singh, Y., & Nandicoori, V. K. (2016). Serine/threonine protein phosphatase pstp of mycobacterium tuberculosis is necessary for accurate cell division and survival of pathogen. *Journal of Biological Chemistry*, 291(46), 24215–24230. <https://doi.org/10.1074/jbc.M116.754531>
- Skwark, M. J., Croucher, N. J., Puranen, S., Chewapreecha, C., Pesonen, M., Xu, Y. Y., Turner, P., Harris, S. R., Beres, S. B., Musser, J. M., Parkhill, J., Bentley, S. D., Aurell, E., & Corander, J. (2017). Interacting networks of resistance, virulence and core machinery

- genes identified by genome-wide epistasis analysis. *PLoS Genetics*, 13(2).  
<https://doi.org/10.1371/journal.pgen.1006508>
- Sulis, G., & Pai, M. (2020). Isoniazid-resistant tuberculosis: A problem we can no longer ignore. In *PLoS Medicine* (Vol. 17, Issue 1). Public Library of Science.  
<https://doi.org/10.1371/JOURNAL.PMED.1003023>
- Tameris, M. D., Hatherill, M., Landry, B. S., Scriba, T. J., Snowden, M. A., Lockhart, S., Shea, J. E., McClain, J. B., Hussey, G. D., Hanekom, W. A., Mahomed, H., & McShane, H. (2013). Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: A randomised, placebo-controlled phase 2b trial. *The Lancet*, 381(9871), 1021–1028. [https://doi.org/10.1016/S0140-6736\(13\)60177-4](https://doi.org/10.1016/S0140-6736(13)60177-4)
- Tornheim, J. A., & Dooley, K. E. (2019). The Global Landscape of Tuberculosis Therapeutics. *Annu. Rev. Med*, 70, 105–125. <https://doi.org/10.1146/annurev-med-040717>
- Trauner, A., Borrell, S., Reither, K., & Gagneux, S. (2014). Evolution of drug resistance in tuberculosis: Recent progress and implications for diagnosis and therapy. *Drugs*, 74(10), 1063–1072. <https://doi.org/10.1007/s40265-014-0248-y>
- Tukey, J. (1977). *Exploratory Data Analysis*. Mass.: Addison-Wesley
- Vågene, Å. J., Honap, T. P., Harkins, K. M., Rosenberg, M. S., Giffin, K., Cárdenas-Arroyo, F., ... & Bos, K. I. (2022). Geographically dispersed zoonotic tuberculosis in pre-contact South American human populations. *Nature communications*, 13(1), 1-12.
- Walker, T. M., Ip, C. L. C., Harrell, R. H., Evans, J. T., Kapatai, G., Dediccoat, M. J., Eyre, D. W., Wilson, D. J., Hawkey, P. M., Crook, D. W., Parkhill, J., Harris, D., Walker, A. S., Bowden, R., Monk, P., Smith, E. G., & Peto, T. E. A. (2013). Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: A retrospective observational study. *The Lancet Infectious Diseases*, 13(2), 137–146. [https://doi.org/10.1016/S1473-3099\(12\)70277-3](https://doi.org/10.1016/S1473-3099(12)70277-3)
- Wei, W. H., Hemani, G., & Haley, C. S. (2014). Detecting epistasis in human complex traits. In *Nature Reviews Genetics* (Vol. 15, Issue 11, pp. 722–733). Nature Publishing Group.  
<https://doi.org/10.1038/nrg3747>
- WHO. (2020). *Global tuberculosis report 2020*. World Health Organization.
- WHO. (2021). *GLOBAL TUBERCULOSIS REPORT 2021*. World Health Organization.  
<http://apps.who.int/bookorders>.

Yew, W. W., Cynamon, M., & Zhang, Y. (2011). Emerging drugs for the treatment of tuberculosis. In *Expert Opinion on Emerging Drugs* (Vol. 16, Issue 1, pp. 1–21). <https://doi.org/10.1517/14728214.2011.521497>

Zhang, Y. (2007). Advances in the treatment of tuberculosis. In *Clinical Pharmacology and Therapeutics* (Vol. 82, Issue 5, pp. 595–600). <https://doi.org/10.1038/sj.clpt.6100362>

**CAPÍTULO 11. ANEXO I**

**ARTÍCULO PRESENTADO COMO REQUISITO**

**An epistatic network describes *oppA* and *glgB* as relevant genes for *Mycobacterium tuberculosis*.**



# An Epistatic Network Describes *oppA* and *glgB* as Relevant Genes for *Mycobacterium tuberculosis*

Ali-Berenice Posada-Reyes<sup>1,2\*</sup>, Yalbi I. Balderas-Martínez<sup>3</sup>, Santiago Ávila-Ríos<sup>3</sup>, Pablo Vinuesa<sup>4</sup> and Salvador Fonseca-Coronado<sup>2\*</sup>

<sup>1</sup>Posgrado en Ciencias Biológicas, UNAM, Mexico, Mexico, <sup>2</sup>Facultad de Estudios Superiores Cuautitlán, UNAM, Estado de Mexico, Mexico, <sup>3</sup>Instituto Nacional de Enfermedades Respiratorias "Ismael Cosío Villegas", Ciudad de Mexico, Mexico, <sup>4</sup>Centro de Ciencias Genómicas, UNAM, Cuernavaca, Mexico

*Mycobacterium tuberculosis* is an acid-fast bacterium that causes tuberculosis worldwide. The role of epistatic interactions among different loci of the *M. tuberculosis* genome under selective pressure may be crucial for understanding the disease and the molecular basis of antibiotic resistance acquisition. Here, we analyzed polymorphic loci interactions by applying a model-free method for epistasis detection, SpydrPick, on a pan-genome-wide alignment created from a set of 254 complete reference genomes. By means of the analysis of an epistatic network created with the detected epistatic interactions, we found that *glgB* ( $\alpha$ -1,4-glucan branching enzyme) and *oppA* (oligopeptide-binding protein) are putative targets of co-selection in *M. tuberculosis* as they were associated in the network with *M. tuberculosis* genes related to virulence, pathogenesis, transport system modulators of the immune response, and antibiotic resistance. In addition, our work unveiled potential pharmacological applications for genotypic antibiotic resistance inherent to the mutations of *glgB* and *oppA* as they epistatically interact with *fprA* and *embC*, two genes recently included as antibiotic-resistant genes in the catalog of the World Health Organization. Our findings showed that this approach allows the identification of relevant epistatic interactions that may lead to a better understanding of *M. tuberculosis* by deciphering the complex interactions of molecules involved in its metabolism, virulence, and pathogenesis and that may be applied to different bacterial populations.

**Keywords:** *glgB*, *oppA*, epistatic network, co-selection, *Mycobacterium tuberculosis*, tuberculosis

## OPEN ACCESS

### Edited by:

Alberto Jesus Martin,  
Universidad Mayor, Chile

### Reviewed by:

Ramandeep Singh,  
Translational Health Science and  
Technology Institute (THSTI), India  
Verlaine Joy Timms,  
The University of Newcastle, Australia

### \*Correspondence:

Ali-Berenice Posada-Reyes  
berenicaposadareyes@  
yahoo.com.mx  
Salvador Fonseca-Coronado  
fonsecacoronado@yahoo.com

### Specialty section:

This article was submitted to  
Biological Modeling and Simulation,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 16 January 2022

**Accepted:** 11 March 2022

**Published:** 31 May 2022

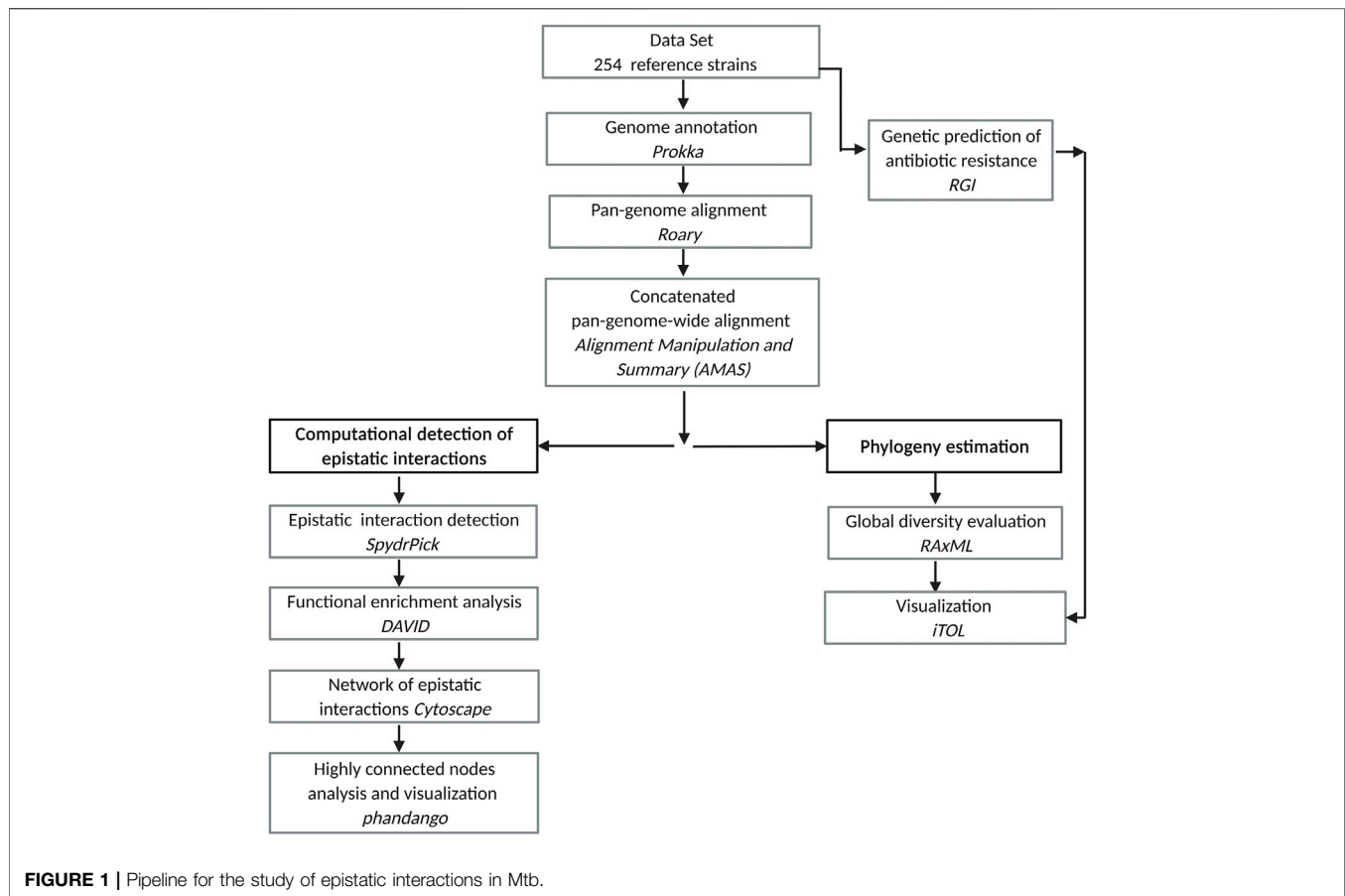
### Citation:

Posada-Reyes A-B,  
Balderas-Martínez YI, Ávila-Ríos S,  
Vinuesa P and Fonseca-Coronado S  
(2022) An Epistatic Network Describes  
*oppA* and *glgB* as Relevant Genes for  
*Mycobacterium tuberculosis*.  
Front. Mol. Biosci. 9:856212.  
doi: 10.3389/fmolb.2022.856212

## 1 INTRODUCTION

In humans, tuberculosis (TB) is a chronic and highly contagious disease that causes more than 10 million human infections and 1.8 million deaths worldwide every year. The constant arrival of drug-resistant strains complicates its control and eradication (Gupta et al., 2018). This disease is mainly caused by members of the *Mycobacterium tuberculosis* complex (MTBC) (Coscolla and Gagneux, 2014) via aerosolized bacteria released by patients with TB (Lerner et al., 2015).

*Mycobacterium tuberculosis* (Mtb) lineages L1–L4 and L7 form a large group of human-adapted strains responsible for the vast majority of global human TB cases, whereas *Mycobacterium africanum* lineages (L5 and L6), which are restricted to humans from West Africa, are phylogenetically linked with the eighth lineage, which comprises various animal-adapted strains (Gonzalo-Asensio et al., 2014).



The first complete genome sequence of Mtb was described in 1998 (Cole et al., 1998). Since then, whole-genome sequencing (WGS) has been applied to a wide range of clinical scenarios, with the potential to revolutionize TB diagnosis, outbreak investigation, development of drugs and vaccines, and to assist in understanding the evolution and pathogenicity of MTBC (Satta et al., 2018). The increase in genomic data in this new era of big data can be considered a great opportunity to continue with the epidemiological surveillance of Mtb associated with the evaluation of genotypic antibiotic resistance. Moreover, it may allow us to unveil new genes with characteristics that lead us to a better understanding of TB.

Recent advances in the scale and diversity of population genomic data for Mtb provide the potential for revealing whole-genome genetic patterns. Statistical methods combined with recent advances in computational structural biology have identified the polymorphic loci (positions inside a genome) under the strongest co-evolutionary pressures or epistatic interactions (Skwark et al., 2017). Such epistatic interactions describe a functional relationship between genes or polymorphic loci (Sackton and Hartl, 2016). Studies of interactions between mutations in Mtb that result in resistance to diverse drugs have suggested that epistasis may be related to multidrug resistance (Trauner et al., 2014; Kavvas et al., 2018). However, the role of epistatic interactions among many regions of the genome under selection in Mtb remains unknown, and further study will contribute to improving our knowledge of TB.

In this study, we analyzed polymorphic loci interactions for epistatic detection in a set of 254 complete reference genomes from Mtb by the use of the model-free method, SpydrPick (Pensar et al., 2019). SpydrPick is based on calculating the mutual information between two polymorphic loci. This well-annotated reference collection integrates genome annotation, gene characterization, and a sequence variation report with a high certainty of genomic location. First, a pan-genome was created using Roary (Page et al., 2015). Then, using AMAS, a pan-genome-wide alignment was obtained by concatenating individual gene alignments. This pan-genome-wide alignment was the input for SpydrPick.

The application of the method to this data set allowed us to reconstruct an epistatic network. The analysis of this network revealed two putative targets of co-selection (*glgB* and *oppA*) associated with Mtb genes related to virulence, pathogenesis, transport system modulators of the immune response, and antibiotic resistance. This work may have relevant applications in the characterization of new genes involved in the worldwide problem of Mtb drug resistance (WHO, 2021).

## 2 MATERIALS AND METHODS

An overview of our approach is depicted in **Figure 1**. The steps are described in the following subsections.

## 2.1 Data Set

We gathered 254 reference strains of Mtb from the NCBI Refseq database that was available as of 4 November 2020. The list of strains is provided as **Supplementary Data S1**.

## 2.2 Creating Pan-Genome-Wide Alignment

Following the strategy of Pensar et al. (2019), we created a pan-genome-wide alignment of the 254 strains. First, we employed Prokka (Seemann, 2014) to annotate genes and features of interest in the set of strains. This genome annotation (GFF3 format) was the input to create a pan-genome of the strains with Roary (Page et al., 2015).

This tool extracts the gene sequences from the input and then identifies clusters to obtain gene alignments. Roary considers two categories of genes: core and accessory. A gene is considered “core” if it is in at least a certain percentage of strains (isolates) defined by the user. In our study, we followed the approach of Pensar et al. (2019), who set this percentage on 95% strains. The output of Roary is a set of files with individual gene alignments, with one file per gene. These files are concatenated in a matrix using the Alignment Manipulation and Summary (AMAS) tool (Borowiec, 2016). This matrix is formed by gene 1 joined on the right with gene 2 and so on with the rest of the genes [see the example “A: concatenation” from Figure 1 in Borowiec (2016)]. Thus, the columns of the output matrix are the genes, and the number of rows is the number of strains used to generate the pan-genome-wide alignment (254 in this case).

## 2.3 Global Diversity Evaluation

The pan-genome-wide alignment was evaluated for global diversity by estimating a phylogeny using RAxML Next Generation (Kozlov et al., 2019). A standard nonparametric bootstrap of 1,000 replicates was performed. Phylogenies were visualized using iTOL v. 6.4.1 (Letunic and Bork, 2021).

## 2.4 Genetic Prediction of Antibiotic Resistance

We predicted a resistome for the 254 strains using the Resistance Gene Identifier (RGI) tool v. 5.1.1 (Alcock et al., 2020). RGI uses the Comprehensive Antibiotic Resistance Database (CARD) as reference data. Using the output of RGI, we annotated strains for two genotypic characterizations of antibiotic resistance: multidrug-resistant (MDR) strains for those strains with genes resistant to isoniazid and rifampicin and extensively drug-resistant (XDR) strains if they have genes resistant to isoniazid, rifampicin, fluoroquinolone, and at least one of the following three antibiotics: kanamycin, amikacin, or capreomycin. These annotations were incorporated for visualization into the phylogeny displayed by iTOL.

## 2.5 Computational Detection of Epistatic Interactions

### 2.5.1 Epistatic Interaction Detection

We utilized SpydrPick (Pensar et al., 2019) to detect the epistatic interactions in the pan-genome-wide alignment. SpydrPick is a

model-free method whose computational efficiency enables analysis at the scale of pan-genomes of bacteria. This method facilitates the detection of targets of co-selection related to virulence and antibiotic resistance. The potential of this method is the detection of epistatic interactions in the absence of phenotypic data.

The approach of SpydrPick is based on calculating the mutual information (MI) between two polymorphic loci. MI is an information-theoretic measure of the amount of information that one random variable,  $X$ , contains about another random variable,  $Y$ . MI is also defined as the reduction in uncertainty in  $X$  after observing  $Y$ ; in other words, MI manifests the reduction in uncertainty of  $X$  due to the knowledge of  $Y$  (Cover and Thomas, 2006). MI gives a measure of association or correlation between  $X$  and  $Y$  (Chanda et al., 2020); if the two variables,  $X$  and  $Y$ , are independent, then the MI is zero. MI is formally defined as follows:

$$MI(X, Y) = \sum_{x \in (X)} \sum_{y \in (Y)} p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right), \quad (1)$$

where  $p(x, y)$  is the joint probability and  $p(x)$  and  $p(y)$  are the marginal probabilities of  $X$  and  $Y$ . MI has been successfully used for detecting co-selection in bacterial population genomics at a genome-wide scale. Another relevant feature introduced by SpydrPick’s approach is the correction for the population structure. This is applied by a sequence reweighting strategy based on how different are the sequences in the pan-genome-wide alignment (Pensar et al., 2019).

SpydrPick detects direct and indirect interactions between loci. A direct interaction occurs between two positions ( $P_1 \rightarrow P_2$ ), whereas an indirect interaction occurs when the two positions ( $P_1$  and  $P_2$ ) are also linked through a third position ( $P_1 \rightarrow P_3 \rightarrow P_2$ ). In the case of indirect interactions ( $P_1 \rightarrow P_2$ ), SpydrPick removes the interaction if the MI is not larger than the other two interactions ( $P_1 \rightarrow P_3$  and  $P_3 \rightarrow P_2$ ).

In addition, SpydrPick performs an analysis to detect outlier interactions. A first criterion to filter outliers is that the distance (bp) between the positions of polymorphic loci must be greater than a linkage disequilibrium (LD) parameter. In this case, a strong LD refers to a close genetic distance between two nucleotide positions. Due to a strong LD hiding a prospective signal of shared co-evolutionary selection pressure, SpydrPick filters out pairs of positions with strong LD to select outlier interactions. According to the SpydrPick’s documentation (<https://github.com/santeripuranen/SpydrPick>), for bacterial genomes, the typical values of the LD are in the 500–20,000 bp range, and the default approach to filtering out strong LD pairs is using a simple distance-based cut-off (20,000 in our case). The second criterion is that the MI must be greater than a threshold obtained from Tukey’s outlier test  $Q_3 + 1.5 \times (Q_3 - Q_1)$  (Tukey, 1977).

The output of SpydrPick is a table of epistatic interactions that includes the pair of positions of two interacting polymorphic loci in the pan-genome-wide alignment, the genome distance between the two positions, the type of interaction (direct/indirect), and the MI score. When SpydrPick detects outliers,

they are reported in another table, including three additional fields: the MI score without gaps, the gap effect, and if the outlier is considered an extreme outlier ( $MI > Q_3 + 3 \times (Q_3 - Q_1)$ ). From the input alignment, SpydPick categorizes any character different from A, C, G, and T as a gap. Gaps are considered in the default MI calculation, so  $X$  and  $Y$  have an outcome space of five categories. As the gaps may not be informative, SpydPick calculates for each pair of positions in the outliers another MI score considering only those strains without gaps in either of the two positions. This MI score is named mutual information without gaps (MI\_wo\_gaps). Using the MI score without gaps, the gap effect is calculated as  $(1 - MI\_wo\_gaps/MI) \times 100$  to quantify the positive or negative effect on the MI by discarding strains with gaps in the two positions.

Comparing MI scores without gaps in a meaningful way is difficult due to the fact that the set of strains without gaps in the two positions varies between pairs of positions (Pensar et al., 2019). However, a high value of the gap effect for a given pair of positions may indicate a gap-driven interaction, and a manual analysis of the pair should be required. Thus, following the analysis performed by Pensar et al. (2019), we used the default MI, leaving the analysis of the MI\_wo\_gaps for a future in-depth study.

The loci of epistatic interaction were annotated with gene id and gene name. Gene names were obtained from the partitions generated by AMAS using an R script. Afterward, using another R script (<https://github.com/biotb/epitb-net>) and the R Biomart library (Drost and Paszkowski, 2017), we retrieved the ENTREZ gene id by searching the gene name in the GFF file of the Mtb H37Rv reference genome (GCF\_000195955.2).

If there is no gene name detected by Roary during the pan-genome creation, then Roary gives a unique generic name formed by the prefix group and a consecutive number. These generic names also appear in partitions of AMAS; however, no ENTREZ id could be associated with these generic names as these names did not exist in the reference genome GFF file.

On the other hand, Prokka was indicated by a numeric suffix different annotation for the same gene, such as *carB\_1* and *carB\_2* (carbamoyl-phosphate synthase large chain). These names were also not found in the reference genome GFF file. In these cases, we eliminated the numeric suffix to find the gene name in the reference genome file. For example, we were able to find the gene id 886,253 for *carB*.

### 2.5.2 Functional Enrichment Analysis

We used the database for annotation, visualization, and integrated discovery (DAVID) v6.8 (Huang et al., 2009) to obtain a functional annotation of Gene Ontology (GO) terms and KEGG pathways of the genes participating in the epistatic interactions. Specifically, we used the DAVID Web Service Python Script (Jiao et al., 2012) to generate a chart report.

### 2.5.3 Network of Epistatic Interactions

The set of epistatic interactions can be seen as a model of complex epistatic relations that may be analyzed and displayed as a network. Here, we used Cytoscape (Kohl et al., 2011) to study our set of epistatic interactions. This tool has been utilized for

studying diverse types of genetic networks. Cytoscape includes an Analyze Network Tool that calculates several network parameters, such as node degree and betweenness centrality. Another useful tool of Cytoscape is the set of layout algorithms based on the yFile Layout Algorithm App. These algorithms visually organized a network by aligning and rotating groups of nodes.

### 2.5.4 Highly Connected Nodes Analysis and Visualization

We focused on the most highly connected genes (the highest degree) for analyzing our epistatic network. Functional characterization of these genes was performed by literature curation and showing enriched GO terms for genes interacting with them. In addition, we used the R package SeqinR (Gouy et al., 1985) to upload the pan-genome-wide alignment and extract the allele distribution at loci involved in their epistatic interactions. We used the interactive web tool Phandango, which is used to visualize phylogenetic trees and associated genomic information (Hadfield et al., 2017), to show the estimated phylogeny and allele distribution of loci.

## 3 RESULTS AND DISCUSSION

### 3.1 Pan-Genome-Wide Alignment

A total of 6,205 individual genes were aligned by Roary, including 3,659 core genes. After concatenating all individual genes with AMAS, a pan-genome-wide alignment of 6,751,593 bp was obtained.

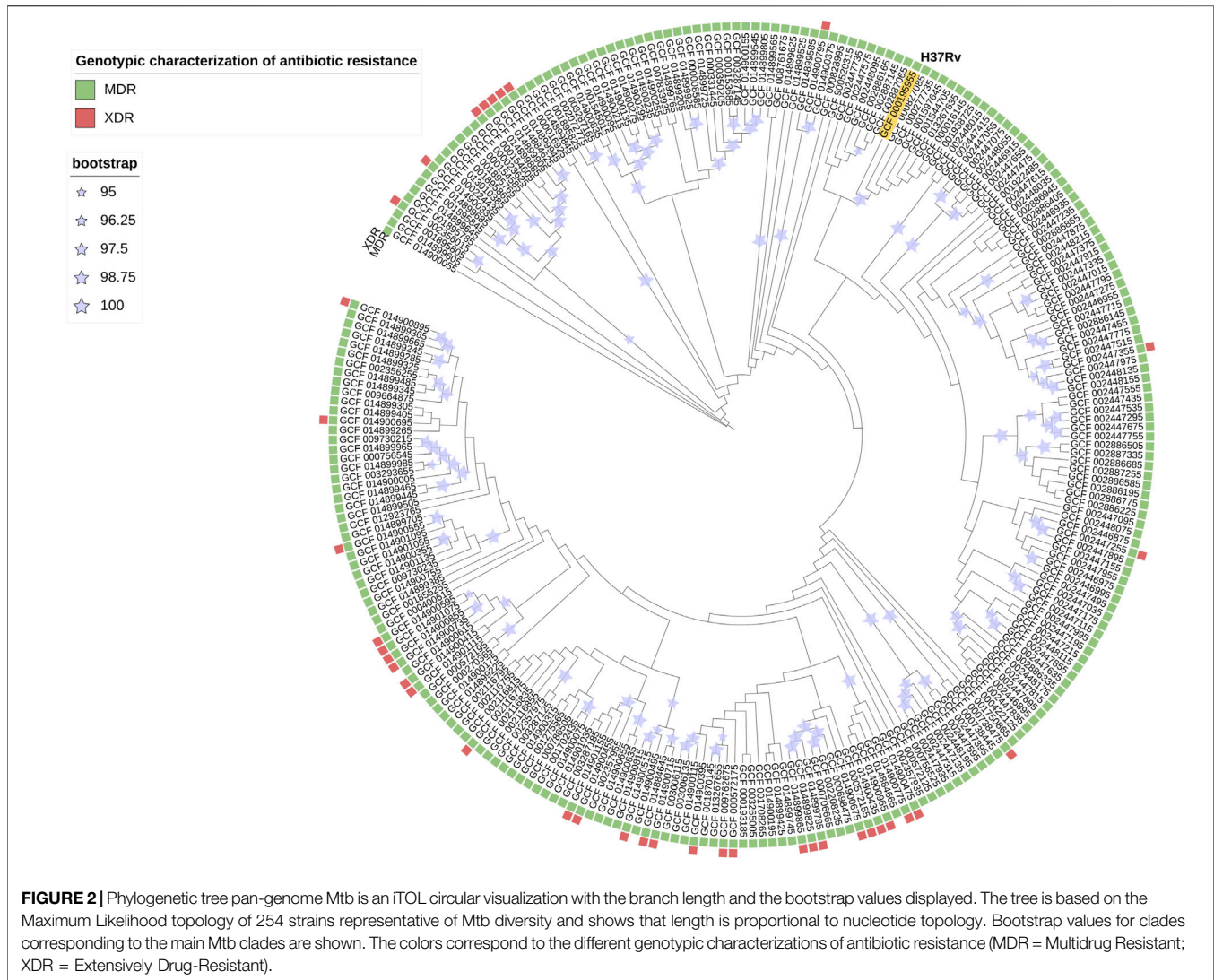
### 3.2 Estimated Phylogeny and Antibiotic Resistance Prediction

Based on Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), which are theoretical information criteria to penalize complex models, we selected the estimated phylogeny using a GTR model with four free rates (GTR-R4-FO). The comparison of models is provided in **Supplementary Table S1**. Convergence using the extended majority rule (MRE) criterion (Pattengale et al., 2010) with a 3% cutoff for the bootstrapping was reached after 400 trees.

The prediction of antibiotic resistance by RGI reported that 100% of the 254 strains were MDR (**Figure 2**) and, within this, 15% were XDR. This result indicates that bacterial strains, perhaps currently circulating, present a high level of resistance to first-line treatments, hindering the successful response to treatment and facilitating the dissemination of strains with drug resistance mutations. Thus, detecting epistatic interactions to elucidate polymorphic loci under the strongest co-evolutionary pressure is of utmost importance for molecular surveillance with bioinformatic tools that help us characterize them promptly. Currently, it is reported that 3.4% of the new TB patients and 20% of the patients with a history of previous treatment for TB were diagnosed with MDR TB worldwide (WHO, 2021).

The phylogenetic tree (**Figure 2**) shows the nucleotide diversity of Mtb (254 strain collection). In this study, there is





heterogeneity of submitters, 100% of the strains present genes linked to antibiotic resistance and with respect to the H37Rv strain (reference), and most of the strains present greater genetic diversity.

### 3.3 Detected Epistatic Interactions

SpydrPick detected 10,573 outlier epistatic interactions (5,484 directed and 5,089 indirect). These interactions describe polymorphic loci under the strongest co-evolutionary pressure. A table with the complete list of outliers is provided as **Supplementary Data S2**. This table includes the fields described in **subsection 2.5.1**, that is, the pair of positions of the two interacting polymorphic loci in the pan-genome-wide alignment, genome distance between the two positions, type of interaction (direct/indirect), MI score, MI score without gaps, gap effect, and whether the outlier interaction is considered an extreme outlier.

After gene annotation, we generated a new table of epistatic interactions that excluded the generic gene names given by Roary.

The new table of outliers included 1,940 epistatic interactions among 107 unique genes. From this set of genes, we only found 70 in the reference genome GFF file, and they were associated with their ENTREZ id. Filtering only those interactions, including these 70 genes, we obtained a final table with 890 outlier interactions. The remaining interactions that were not considered in our study will be included in a future analysis.

The final table of outlier interactions includes the two positions of the two interacting polymorphic loci in the pan-genome-wide alignment, gene ENTREZ id and gene name for each position, distance between the two positions (bp), type of interaction (direct = 1, indirect = 0), MI score of the interaction, and if the interaction outlier is an extreme outlier (yes = 1, no = 0). This final table of outliers (**Supplementary Data S3**) was used for enrichment analysis, network reconstruction, and analyses.

SpydrPick was able to find long-distance interactions surpassing the two million bp (**Table 1**). This fact confirms that our study has a whole-genome scale. However, because

**TABLE 1** | Statistics of the distance between positions of the two interacting polymorphic loci and statistics of the MI scores, both for the final outlier interactions.

Statistic of the final outlier interactions	Values
Minimum distance (bp)	20,870
Maximum distance (bp)	2,328,291
Median distance (bp)	721,980
Mean distance (bp)	846,454
Minimum MI score	0.4130
Maximum MI score	0.5020
Median MI score	0.4202
Mean MI score	0.4509

we identified epistatic interactions from a pan-genome-wide alignment constructed by concatenating individual gene alignments, the positions are not straightforward whole-genome loci. The minimum distance (bp) between the positions of the two interacting polymorphic loci nearly surpassed the LD criterion of 20,000 bp. The mean of the distance between polymorphic loci in the outlier interactions was 846,454 bp; considering that it is greater than the median (721,980 bp), there may be a slight skewness to distances lower than the mean. On the other hand, the range of MI scores was short, from 0.4130 to 0.5020 (Table 1). The MI scores might show a skewness to low values as the mean (0.4509) was higher than the median (0.4202).

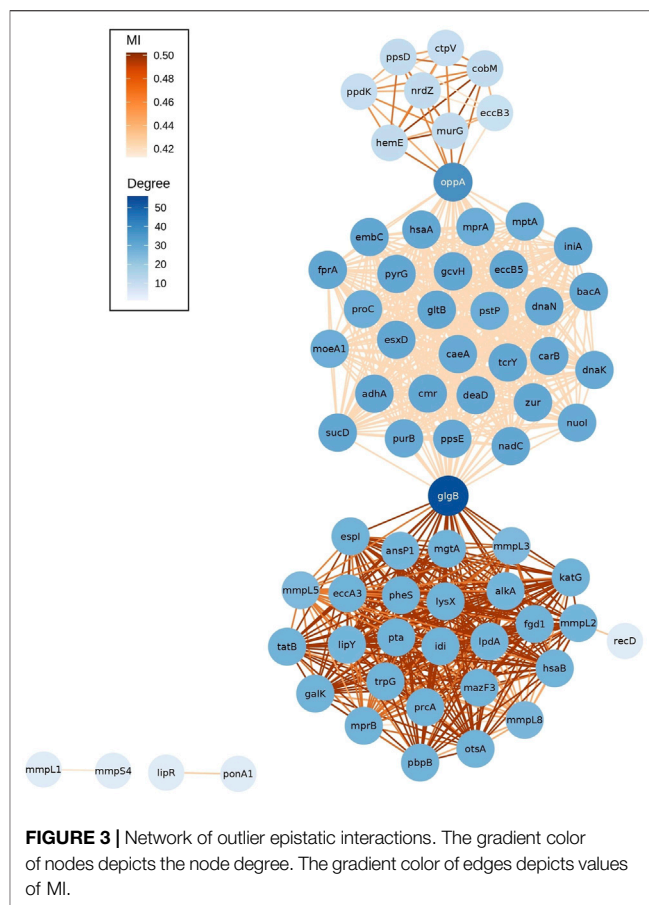
All loci in the 890 interactions were found in the described single-nucleotide polymorphisms (SNPs) when we used the pan-genome-wide alignment with the tool SNP-sites v. 2.5.1, which can rapidly identify SNPs from a multi-FASTA alignment (Page et al., 2016). This additional step was developed to identify polymorphisms involved in the detected epistatic interactions.

### 3.4 Epistatic Network Analysis

#### 3.4.1 *glgB* and *oppA* as Putative Targets of Co-selection

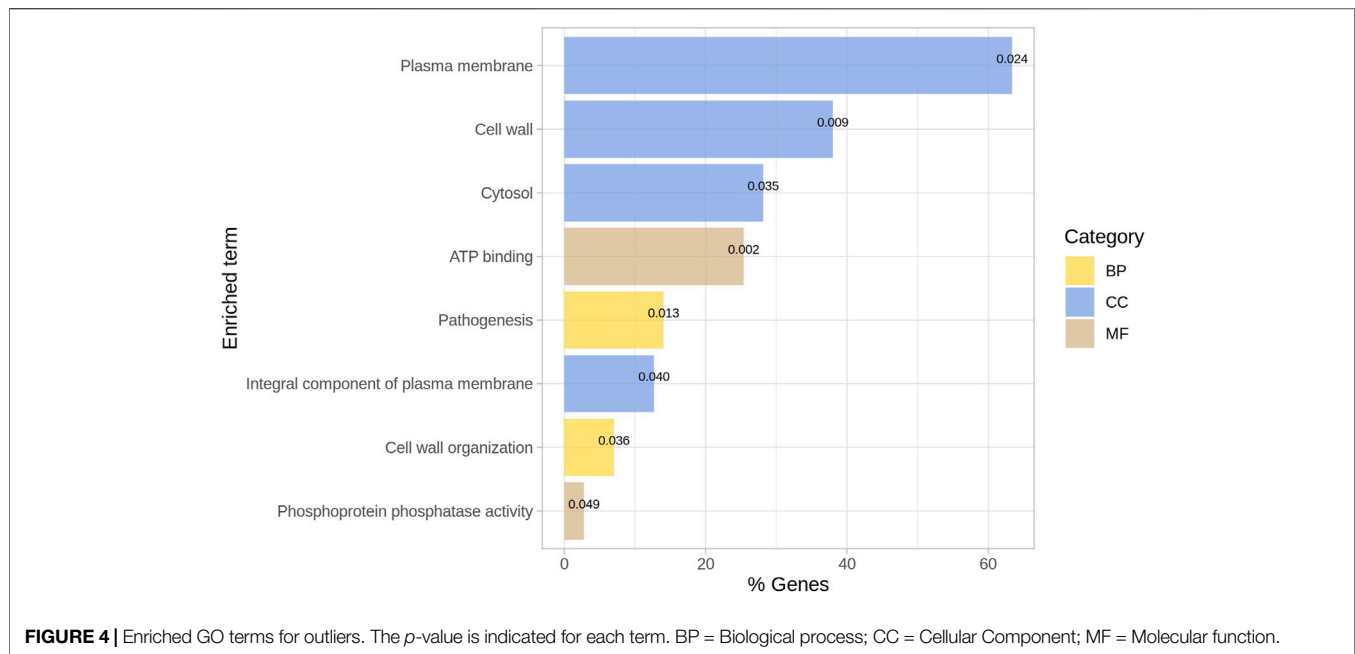
The network of epistatic interactions was analyzed to figure out those genes with a high node degree (the number of edges), that is, a high level of connectivity of the gene with other genes. The most highly connected genes were *glgB* (ENTREZ:886,893, degree = 56), a  $\alpha$ -1,4-glucan branching enzyme (GlgB), and *oppA* (ENTREZ:886,985, degree = 37), an oligopeptide-binding protein (OppA) (Figure 3).

These two genes also have the highest value of betweenness centrality (*glgB* = 0.493, *oppA* = 0.219). Betweenness centrality is higher for those nodes that join subnets (communities) than those located inside the subnets. Here, we observed three subnets connected by these two genes. An interesting pattern is that each subnet has a different distribution of values of MI (see gradient color of edges in Figure 3). The subnet at the bottom has higher MI values (median MI = 0.502) than the other two, the top subnet has a median MI of 0.463, and the subnet at the middle has lower values (median MI = 0.420). A further study is required to elucidate the cause of this pattern. In addition, future analysis will be required to identify the patterns associated with the isolated subnets (*mmpL1-mmpS4* and *lipR-ponA1*).

**FIGURE 3** | Network of outlier epistatic interactions. The gradient color of nodes depicts the node degree. The gradient color of edges depicts values of MI.

Thus, we consider these two genes as relevant putative targets of co-selection because they may be associated with several genes related to potential pharmacological applications. The GlgB enzyme (encoded by Rv1326c) is the key enzyme involved in the biosynthesis of  $\alpha$ -glucan, which plays a significant role in the virulence and pathogenesis of Mtb. Recently, enzymes that participate in the biosynthesis of trehalose have gained major attention as drug targets, especially in Mtb (Dkhar et al., 2015), as capsular polysaccharides of bacteria have been found to modulate the host immune response. The importance of the metabolism of GlgB has been described (De Smet et al., 2000), but the epistatic interactions with other genes remain unknown.

On the other hand, the gene *oppA* (oligopeptide-binding protein) works as a substrate-binding protein for the oligopeptide transport system (Opp), which is responsible for peptide importation. The Opp system is an ATP-binding cassette transporter. This helps in peptide absorption, giving pathogens the essential nutrients as a source of carbon, nitrogen, and amino acids. The Opp system affects many cellular processes, including internalization of quorum-sensing peptides, biofilm production, cell surface modification, and antibiotic resistance (Hopfe et al., 2011). The relevance of the characterization of the peptide transporter system has been described by Dasgupta et al. (2010). Previous studies uncovered the novel observation that this peptide transporter modulates the innate immune response of macrophages infected (Cassio Barreto de Oliveira and Balan,



2020) with Mtb, but the epistatic interactions of *oppA* with other loci remain unknown.

A bacterium is able to adapt its response to host conditions, such as intracellular residence in phagocytic cells, oxidative stress, hypoxia, and carbon and nitrogen source. For this reason, evaluating interactions by bioinformatics experiments is necessary for the identification of new epistatic interactions in genes that have been previously reported in databases, such as the catalog of the WHO, or for the understanding of the epistatic interactions in Mtb before the development of new therapies.

### 3.4.2 Enriched GO Terms for the Epistatic Network

The list of ENTREZ ids of the genes of the network was used to perform a functional enrichment analysis with DAVID. From the DAVID chart report, we only considered those terms as relevant with *p*-value < 0.05 (see **Supplementary Data S4** for details of the functional enrichment analysis). Biological processes of pathogenesis (GO:0009405) and cell wall organization (GO:0071555) were enriched in a subset of genes (**Figure 4**). The cell wall (GO:0005618), plasma membrane (GO:0005886), cytosol (GO:0005829), and integral components of the plasma membrane (GO:0005887) were the more abundant cellular components; in this case, 63% of the genes are in the plasma membrane. Regarding molecular functions, we obtained enrichment for ATP binding (GO:0005524) and phosphoprotein phosphatase activity (GO:0004721) for some genes.

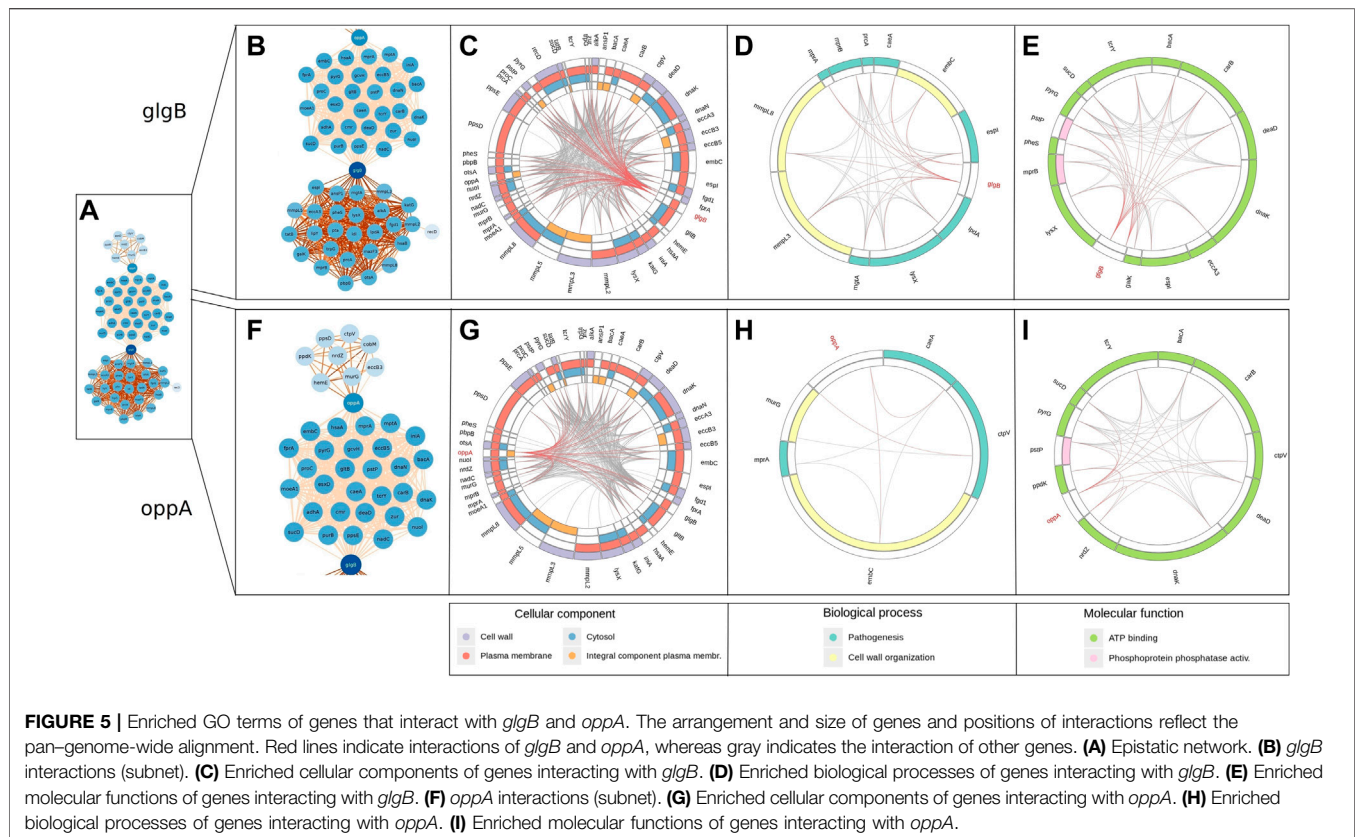
Recently, the biomarkers of Mtb that regulate immune response have been identified to potentially develop drugs for TB. It has been previously described that the functionality of cellular components was associated with infection and verified the regulation of these cellular components as relevant regulators of the immune response in the host (Li et al., 2020). Thus, describing the genes involved in cellular components is crucial

for understanding the interactions of bacteria with host molecules that regulate immune response.

In recent studies, the relevance of the structure and biogenesis-related genes of Mtb encoding glycoconjugates has been confirmed, with particular emphasis on the molecules across the different layers of the cell envelope (Angala et al., 2014). In addition, it has been previously stressed that ATP production is crucial for antibiotic resistance in bacteria (Black et al., 2014).

We show the enriched GO terms of genes interacting with *glgB* (**Figure 5B**) and *oppA* (**Figure 5F**) using circular layouts. In addition, **Supplementary Data S5** also contains in table format the genes interacting with *glgB*, their product, and enriched GO terms; the same information is provided for *oppA* in **Supplementary Data S6**. The layouts were generated using the start and end positions of genes reported in partitions generated by AMAS, so the arrangement and size of genes in the layout and positions of interactions reflect the pan-genome-wide alignment. We have highlighted the interactions of *glgB* and *oppA* in red to distinguish them from the interaction of other genes (shown in gray).

Both *glgB* and *oppA* epistatically interact with genes enriched with the biological processes of pathogenesis and cell wall organization (**Figures 5D,H**). One of these genes is *embC*, which codifies for an arabinosyltransferase involved in the biosynthesis of a major component of the mycobacterial cell wall lipoarabinomannan (LAM). The characteristic manose-capped LAM of Mtb acts as a pathogen-associated molecular pattern (PAMP), modulating the activation of phagocytic cells to control the strength of the host inflammatory immune response, while representing one of the main components in the cell wall organization. In addition, it has been described that *embC* is expressed as part of a polycistronic mRNA controlled by a promoting region differentially expressed depending on the stationary or hypoxia-induced persistence phase of the bacilli,

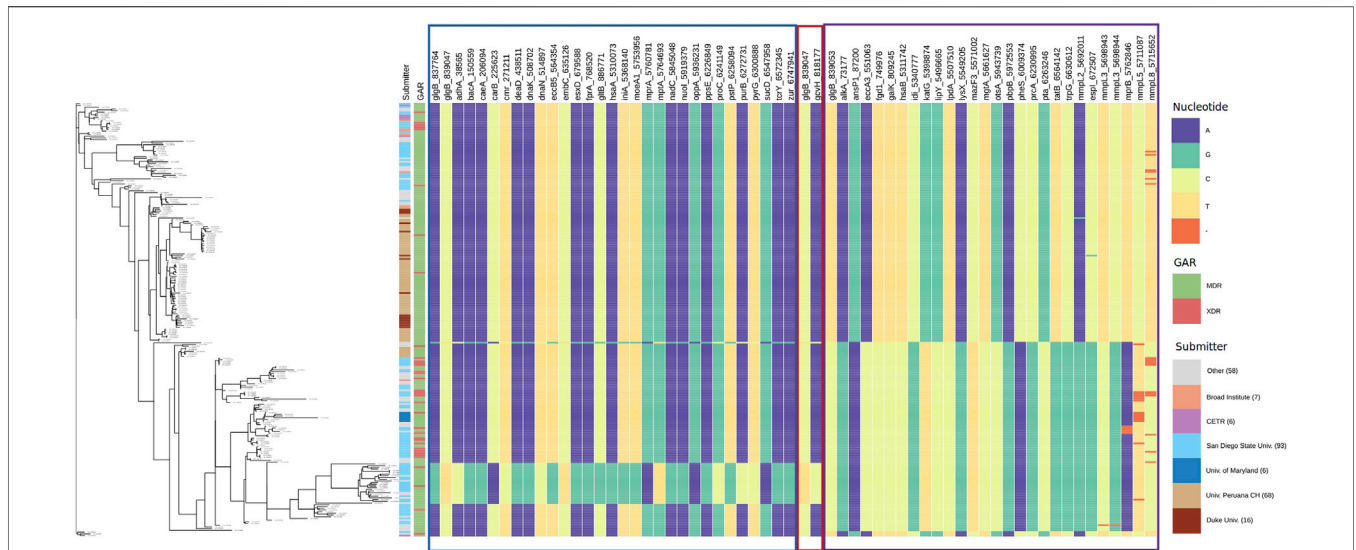


highlighting the important role of this protein in the biological functions of Mtb and the complex interaction involved in cell wall regulation. Whether or not a direct interaction of *embC* with *glgB* and/or *oppA* exists remains an exciting question to be addressed (Goude et al., 2008).

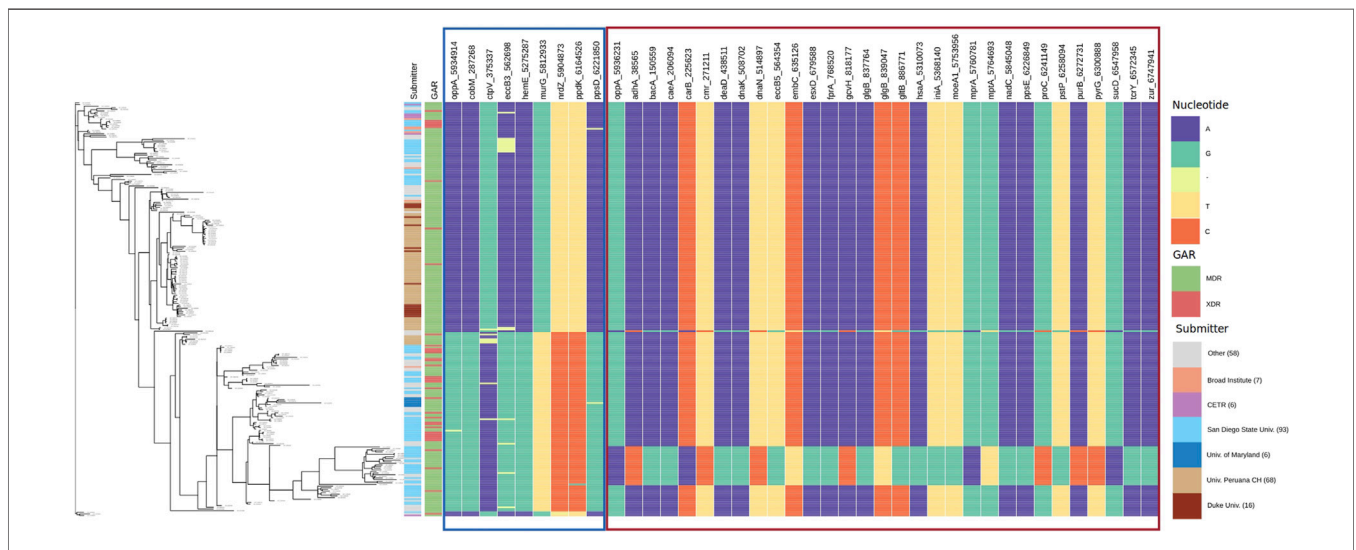
About molecular functions, *oppA* and *glgB* interact with several genes enriched with ATP binding and with three genes enriched with phosphoprotein phosphatase activity (Figures 5E, I). From these genes, *bacA* is another gene found to be interacting at the highest scores with both *glgB* and *oppA* in the network, and *bacA* encodes for a protein of the type IV family of ABC transporter-type exporters; despite the structure, their function as an importer of multi-solute hydrophilic compounds, such as vitamin B12, bleomycin, and aminoglycosides, has been demonstrated due to a large occluded water-filled cavity that spans across the whole lipid membrane. In addition, it has also been demonstrated that this transporter is implicated in the maintenance of chronic infection in murine models by mediating the transport of a molecule that can directly or indirectly modulate the proinflammatory host response. Despite having different structures, BacA and OppA shared their ability to transport a wide range of substrates; in particular, the shared capacity of import peptides related to the innate immune response suggests a complex regulation and interaction of these transporters, guaranteeing the need to carry out studies at the level of gene regulation and function in the near future (Domenech et al., 2009; Cassio Barreto de Oliveira and Balan, 2020; Rempel et al., 2020).

The enriched GO terms that may be related to those associated with ATP synthase in mycobacteria are of particular interest because they contribute to efficient ATP production, and this enzyme has been validated as a target for potential pharmacological applications. In addition, mycobacterial ATP synthase and its characteristics may provide information on adaptations of bacterial energy metabolism. Mtb can survive in human macrophages for an extended time. For Mtb and other pathogenic mycobacteria strains, the blocking of ATP hydrolysis is relevant as it may represent an adaptation to its internal and external human phagosomes, where ATP, once produced, must not be used (Lu et al., 2014). Thus, the importance of epistatic interactions associated with ATP production in pathogenic bacteria may face exceptional challenges as a variety of pathogens need to deal with low energy conditions, such as low oxygen tensions or nutrient limitation inside the host.

Regarding gene interactions enriched with phosphoprotein phosphatase activity, both *oppA* and *glgB* established epistatic interaction with the *pstP* gene, which encodes the Serine/Threonine Protein Phosphatase PstP of Mtb. Signal sensing and transduction *via* phosphorylation and dephosphorylation of specific target proteins are essential for the survival of both eukaryotic and prokaryotic organisms. In the case of Mtb, 11 serine/threonine protein kinases have been described, but only the serine/threonine phosphatase, PstP, has been identified, highlighting the central role of this protein in the control of vital processes as a negative regulator of kinase activity and global serine and threonine phosphorylation (Iswhayudi et al., 2019).



**FIGURE 6** | Allele distribution at the loci of genes interacting with *glgB*. Labels of the column indicate gene name and locus. Estimated phylogeny is included on the left. Interacting loci are organized in sections with borders. The border color corresponds to the color of the sections in **Supplementary Data S7**. GAR = Genotypic Antibiotic Resistance.



**FIGURE 7** | Allele distribution at the loci of genes interacting with *oppA*. Labels of the column indicate gene name and locus. Estimated phylogeny is included on the left. Interacting loci are organized in sections with borders. The border color corresponds to the color of the sections in **Supplementary Data S8**. GAR = Genotypic Antibiotic Resistance.

Recently, other functions of PstP as a regulator of cell wall synthesis and cell division by dephosphorylation of key substrates implicated in both pathways have been described (Sharma et al., 2016). PstP is co-transcribed in an operon with genes involved in peptidoglycan synthesis, with protein kinases PknA and PknB that regulate cell growth and cell division and with *fhaA* and *fhaB*, which encode phosphothreonine recognition proteins that also regulate cell growth and cell division. The involvement of *pstP* with elements necessary for cell wall biosynthesis and their strict dependence on Mn<sup>2+</sup> for function suggests that the interaction

encountered by the computational approach could serve as a starting point for initiating investigations into the molecular interactions that regulate these common processes.

PstP is present as a transmembrane phosphatase and contains a 240–amino acid intracellular catalytic domain, tethered *via* a single transmembrane helix to the 196-amino acid-long extracellular domain (Boitel et al., 2003); it remains to be elucidated if during infection and activation of the innate immune responses (e.g., respiratory burst activation), the degraded bacteria retain the phosphatase activity in the

**TABLE 2** | Antibiotic-resistant genes reported by the WHO catalogue are found in the epistatic network.

Gene	Antibiotic-resistant gene (drug)
<i>glgB</i> and <i>oppA</i>	<i>fprA</i> (AMI)
<i>glgB</i> and <i>oppA</i>	<i>fprA</i> (CAP)
<i>glgB</i> and <i>oppA</i>	<i>embC</i> (EMB)
<i>glgB</i>	<i>katG</i> (INH)
<i>glgB</i>	<i>fgd1</i> (DLM)
<i>glgB</i>	<i>mmpL5</i> (BDQ)
<i>glgB</i>	<i>mmpL5</i> (CFZ)

<sup>a</sup>AMI = amikacin; BDQ = bedaquiline; CAP = capreomycin; CFZ = clofazimine; DLM = delamanid; EMB = ethambutol; INH = isoniazid WHO (2021).

membrane fragments, and these could contribute to the dephosphorylation of the signaling pathways of the innate system, contributing to the evasion of the immune response.

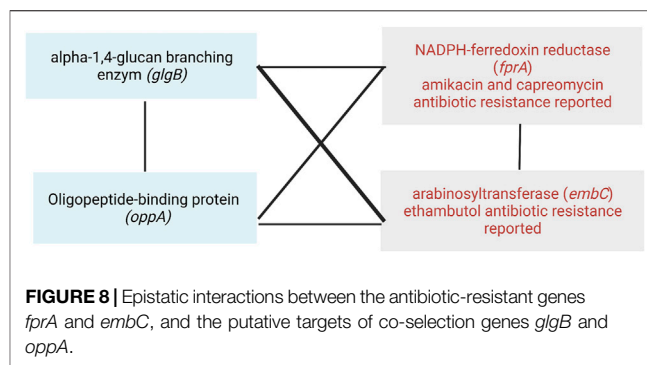
Four cellular components were enriched for genes interacting with *glgB* (Figure 5C) and *oppA* (Figure 5G). The majority of genes encode proteins in the plasma membrane. For example, four genes of the group of mycobacterial membrane protein large (MmpL), *mmpL2*, *mmpL3*, *mmpL5*, and *mmpL8* have epistatic interactions with *glgB* and *oppA*. MmpL proteins export cell envelop components (such as virulence-associated lipids and siderophores) to the periplasmic space, contributing at a high level to the persistence of Mtb in the host (Melly and Purdy, 2019). A further study will be required to investigate the fine regulation between the import and export systems of genes identified under epistatic interaction by our approach in order to establish their relevance and biological implications.

### 3.4.3 Allele Distribution at Loci of Genes Interacting With *glgB* and *oppA*

To observe the patterns of alleles of the interacting polymorphic loci, we show the allele distribution at interacting loci with the loci of *glgB* and *oppA* using Phandango (Hadfield et al., 2017). SpydrPick detected that three loci of *glgB* (837,764, 839,047, and 839,053) interact with 57 polymorphic loci of 56 genes. For *oppA*, two polymorphic loci (5,934,914 and 5,936,231) were found interacting with 38 loci of 37 genes. Tables with loci and genes are available in the **Supplementary Datas S7, S8**.

Interacting loci and gene names are displayed as labels of columns in Figure 6 for *glgB* and in Figure 7 for *oppA*. Interacting loci are organized in sections with borders. Each section includes the interactions for each interacting locus. The border color for each section corresponds to the color of the sections in **Supplementary Data S7, S8**. In Figure 6, the first section starts with the loci 837,764 and 839,047 of *glgB* (*glgB*\_837764 and *glgB*\_839047) followed by the 29 polymorphic loci that interact with them, that is, these two loci epistatically interact with each one of the 29 loci.

By observing the allele distribution of pairs of loci, we confirm that SpydrPick is able to detect, using the MI score, predictable patterns of alleles at the two loci. See, for example, the second section in Figure 6, which only depicts the allele distribution of the locus 839,047 of *glgB* (*glgB*\_839047) and the interacting locus 818,177 of *gcvH* (*gcvH*\_818177). It can be noticed that when there



is a C in the locus *glgB*\_839047, there is an A in the locus *gcvH*\_818,177, and when there is a T in the locus *glgB*\_839047, there is a C in the locus *gcvH*\_818,177. This predictability is quantitatively depicted by the value of  $MI = 0.4201$ . The third section in Figure 6 exposes the allele distribution of interactions between the locus 839,053 of *glgB* (*glgB*\_839053) and 27 loci of 26 genes (two loci of the gene *mmpL3* interact with the locus *glgB*\_839053). In this section, we observe gaps (-) in the allele distribution of the interacting locus 5,762,846 of *mprB* (two component histidine-protein kinase/phosphatase MprB,  $MI = 0.4743$ ), locus 5,711,087 of *mmpL5* (transmembrane transport protein MmpL5  $MI = 0.4670$ ) and locus 5,715,652 of *mmpL8* (integral membrane transport protein MmpL8  $MI = 0.4435$ ).

In Figure 7, we present the allele distribution of loci interacting with loci 5,934,914 and 5,936,231 of the gene *oppA*. The first section contains the interactions with the locus *oppA*\_5,934,914. We notice the presence of gaps in the positions *ctpV*\_375337 ( $MI = 0.4516$ ) and *eccB3*\_562,698 ( $MI = 0.417$ ). Figure 7 also presents well-defined patterns of allele distribution between interacting loci.

The first column in both figures indicates the submitter institution (we included institutions with less than six submitted strains in the category other), and the second column points to genotypic antibiotic resistance (GAR) predicted with RGI. Estimated phylogeny is included on the left to show the diversity of the strain collection that we analyzed. For example, a clade at the bottom of the tree stands out due to its change of nucleotide in relation to the rest of the strains.

The application of this methodology also allowed the identification within the network of various loci in genes associated with resistance. Recently, the WHO published the first catalog of resistance-associated genetic variants for predicting relevant resistance phenotypes based on more than 38,000 WGS phenotyped isolates (WHO, 2021). This has allowed the identification of multiple positions associated with resistance and their classification into five groups. We use this recent classification to find antibiotic-resistant genes in the epistatic network (Table 2).

In our network, an interaction was found between *glgB* and *katG*; mutations conferring monoresistance to isoniazid (INH) are common due to INH having been in clinical use since the 1950s. Nevertheless, INH resistance testing is only recently included in some specialized cartridges (e.g., Xpert MTB/XDR) and is not routinely available in such a way that if INH resistance

is not detected, patients are treated as pan-susceptible, which represents a high risk of treatment failure and a greater propensity to acquire further resistance (Sulis and Pai, 2020).

From the antibiotic-resistance genes that we found in the WHO catalog, we observed that *fprA* (resistant to amikacin and capreomycin) and *embC* (resistant to ethambutol) interact with both genes *glgB* and *oppA* (Table 2). Moreover, they interact between them, forming a clique of four genes (Figure 8). A clique depicts a network where all nodes are fully connected to each other, creating a strong interaction mechanism. This kind of epistatic interactions motivated us to visualize future studies to test new experimental hypotheses to elucidate their biological and pharmacological explanations, and the MI score seems to be a very successful approach to drive so.

## 4 CONCLUSION

Here, we have presented the reconstruction and analysis of an epistatic network for Mtb from a pan-genome-wide alignment by using the model-free method SpydrPick. Our approach allowed us identifying new epistatic interactions with implications in virulence, pathogenesis, transport system modulators of the immune response, and genotypic antibiotic resistance. By the analysis of the epistatic network, we identified *glgB* and *oppA* as putative targets of co-selection. These two genes epistatically interact with *fprA* and *embC*, two antibiotic-resistant genes reported in the catalog of the WHO, as resistant to ethambutol (*embC*) and amikacin and capreomycin (*fprA*). Our results highlight the importance of implementing computational approaches to elucidate new genes associated to putative epistatic interactions in Mtb.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding authors. R scripts are available at <https://github.com/biotb/epitb-net>.

## REFERENCES

- Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., et al. (2020). Card 2020: Antibiotic Resistance Surveillance with the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res.* 48, D517–D525. doi:10.1093/nar/gkz935
- Angala, S. K., Belardinelli, J. M., Huc-Claustre, E., Wheat, W. H., and Jackson, M. (2014). The Cell Envelope Glycoconjugates of *Mycobacterium Tuberculosis*. *Crit. Rev. Biochem. Mol. Biol.* 49, 361–399. doi:10.3109/10409238.2014.925420
- Black, P. A., Warren, R. M., Louw, G. E., van Helden, P. D., Victor, T. C., and Kana, B. D. (2014). Energy Metabolism and Drug Efflux in *Mycobacterium Tuberculosis*. *Antimicrob. Agents Chemother.* 58, 2491–2503. doi:10.1128/aac.02293-13
- Boitel, B., Ortiz-Lombardía, M., Durán, R., Pompeo, F., Cole, S. T., Cerveñansky, C., et al. (2003). PknB Kinase Activity Is Regulated by Phosphorylation in Two Thr Residues and Dephosphorylation by PstP, the Cognate Phospho-Ser/thr Phosphatase, in *mycobacterium Tuberculosis*. *Mol. Microbiol.* 49, 1493–1508. doi:10.1046/j.1365-2958.2003.03657.x

## AUTHOR CONTRIBUTIONS

ABPR and SFC conceptualized, designed, and conducted the investigation. ABPR analyzed the data and generated the epistatic network. YBM, SAR, PV, and SFC supervised the research. All authors wrote the manuscript. ABPR and SFC take responsibility for the manuscript. All authors read and approved the final content.

## FUNDING

This research was partially supported by the Programa Interno de Apoyo para Proyectos de Investigación, FESC, UNAM (PIAPI 2018).

## ACKNOWLEDGMENTS

ABPR is a doctoral student from the *Programa de Doctorado en Ciencias Biológicas, Universidad Nacional Autónoma de México* (UNAM) and received a fellowship from CONACyT (631936). PV gratefully acknowledges support by CONACyT A1-S-11242 and PAPIIT-UNAM IN209321. ABPR gratefully acknowledge the *Programa de Ingeniería Genómica, CCG-UNAM*, for the support during her stay (February–December 2019). We acknowledge the support of Miguel Ángel Pérez León of the *Laboratorio de cómputo de alto rendimiento, Coordinación del departamento de Matemáticas* and of the *Lansbiodyt (Laboratorio Nacional de Soluciones Biomiméticas para Diagnóstico y Terapia), Facultad de Ciencias, UNAM*. Figures 1, 8 were created with BioRender.com.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.856212/full#supplementary-material>

- Borowiec, M. L. (2016). Amas: a Fast Tool for Alignment Manipulation and Computing of Summary Statistics. *PeerJ* 4, e1660. doi:10.7717/peerj.1660
- Cassio Barreto de Oliveira, M., and Balan, A. (2020). The Atp-Binding Cassette (Abc) Transport Systems in *Mycobacterium Tuberculosis*: Structure, Function, and Possible Targets for Therapeutics. *Biology* 9, 443. doi:10.3390/biology9120443
- Chanda, P., Costa, E., Hu, J., Sukumar, S., Van Hemert, J., and Walia, R. (2020). Information Theory in Computational Biology: Where We Stand Today. *Entropy* 22, 627. doi:10.3390/e22060627
- Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., et al. (1998). Deciphering the Biology of *mycobacterium Tuberculosis* from the Complete Genome Sequence. *Nature* 393, 537–544. doi:10.1038/31159
- Coscolla, M., and Gagneux, S. (2014). Consequences of Genomic Diversity in *mycobacterium Tuberculosis*. *Semin. Immunol.* 26, 431–444. doi:10.1016/j.smim.2014.09.012
- Cover, T. M., and Thomas, J. (2006). *Elements of Information Theory*. New Jersey: Wiley-Interscience.
- Dasgupta, A., Sureka, K., Mitra, D., Saha, B., Sanyal, S., Das, A. K., et al. (2010). An Oligopeptide Transporter of *Mycobacterium Tuberculosis* Regulates Cytokine

- Release and Apoptosis of Infected Macrophages. *PLoS one* 5, e12225. doi:10.1371/journal.pone.0012225
- De Smet, K. A. L., Weston, A., Brown, I. N., Young, D. B., and Robertson, B. D. (2000). Three Pathways for Trehalose Biosynthesis in Mycobacteria. *Microbiology* 146, 199–208. doi:10.1099/00221287-146-1-199
- Dkhar, H. K., Gopalsamy, A., Loharch, S., Kaur, A., Bhutani, I., Saminathan, K., et al. (2015). Discovery of *Mycobacterium tuberculosis*  $\alpha$ -1,4-Glucan Branching Enzyme (GlgB) Inhibitors by Structure- and Ligand-Based Virtual Screening. *J. Biol. Chem.* 290, 76–89. doi:10.1074/jbc.m114.589200
- Domenech, P., Kobayashi, H., LeVier, K., Walker, G. C., and Barry, C. E. (2009). BacA, an ABC Transporter Involved in Maintenance of Chronic Murine Infections with *Mycobacterium tuberculosis*. *J. Bacteriol.* 191, 477–485. doi:10.1128/JB.01132-08
- Drost, H. G., and Paszkowski, J. (2017). Biomart: Genomic Data Retrieval with R. *Bioinformatics* 33, 1216–1217. doi:10.1093/bioinformatics/btw821
- Gonzalo-Asensio, J., Malaga, W., Pawlik, A., Astarie-Dequeker, C., Passemar, C., Moreau, F., et al. (2014). Evolutionary History of Tuberculosis Shaped by Conserved Mutations in the Phopr Virulence Regulator. *Proc. Natl. Acad. Sci. U.S.A.* 111, 11491–11496. doi:10.1073/pnas.1406693111
- Goude, R., Amin, A. G., Chatterjee, D., and Parish, T. (2008). The Critical Role of EmcB in mycobacterium Tuberculosis. *J. Bacteriol.* 190, 4335–4341. doi:10.1128/jb.01825-07
- Gouy, M., Gautier, C., Attimonelli, M., Lanave, C., and Paola, G. d. (1985). Acnuc – A Portable Retrieval System For Nucleic Acid Sequence Databases: Logical And Physical Designs And Usage. *Bioinformatics* 1, 167–172. doi:10.1093/bioinformatics/1.3.167
- Gupta, N., Kumar, R., and Agrawal, B. (2018). New Players in Immunity to Tuberculosis: the Host Microbiome, Lung Epithelium, and Innate Immune Cells. *Front. Immunol.* 9, 709. doi:10.3389/fimmu.2018.00709
- Hadfield, J., Croucher, N. J., Goater, R. J., Abudahab, K., Aanensen, D. M., and Harris, S. R. (2017). Phandango: an Interactive Viewer for Bacterial Population Genomics. *Bioinformatics* 34, 292–293. doi:10.1093/bioinformatics/btx610
- Hopfe, M., Dahlmann, T., and Henrich, B. (2011). In Mycoplasma Hominis the Oppa-Mediated Cytoadhesion Depends on its AtPase Activity. *BMC Microbiol.* 11, 185–189. doi:10.1186/1471-2180-11-185
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and Integrative Analysis of Large Gene Lists Using David Bioinformatics Resources. *Nat. Protoc.* 4, 44–57. doi:10.1038/nprot.2008.211
- Iswahyudi, Mukamolova, G. V., Straatman-Iwanowska, A. A., Allcock, N., Ajuh, P., Turapov, O., et al. (2019). Mycobacterial Phosphatase PstP Regulates Global Serine Threonine Phosphorylation and Cell Division. *Sci. Rep.* 9, 8337–8411. doi:10.1038/s41598-019-44841-9. <https://www.nature.com/articles/s41598-019-44841-9#citeas>.
- Jiao, X., Sherman, B. T., Huang, D. W., Stephens, R., Baseler, M. W., Lane, H. C., et al. (2012). DAVID-WS: a Stateful Web Service to Facilitate Gene/Protein List Analysis. *Bioinformatics* 28, 1805–1806. doi:10.1093/bioinformatics/bts251
- Kavvas, E. S., Catoiu, E., Mih, N., Yurkovich, J. T., Seif, Y., Dillon, N., et al. (2018). Machine Learning and Structural Analysis of mycobacterium Tuberculosis Pan-Genome Identifies Genetic Signatures of Antibiotic Resistance. *Nat. Commun.* 9, 4306–4309. doi:10.1038/s41467-018-06634-y
- Kohl, M., Wiese, S., and Warscheid, B. (2011). “Cytoscape: Software for Visualization and Analysis of Biological Networks,” in *Data Mining in Proteomics* (Springer), 291–303. doi:10.1007/978-1-60761-987-1\_18
- Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a Fast, Scalable and User-Friendly Tool for Maximum Likelihood Phylogenetic Inference. *Bioinformatics* 35, 4453–4455. doi:10.1093/bioinformatics/btz305
- Lerner, T. R., Borel, S., and Gutierrez, M. G. (2015). The Innate Immune Response in Human Tuberculosis. *Cell Microbiol* 17, 1277–1285. doi:10.1111/cmi.12480
- Letunic, I., and Bork, P. (2021). Interactive Tree of Life (iTOL) V5: an Online Tool for Phylogenetic Tree Display and Annotation. *Nucleic Acids Res.* 49, W293–W296. doi:10.1093/nar/gkab301
- Li, L., Lv, J., He, Y., and Wang, Z. (2020). Gene Network in Pulmonary Tuberculosis Based on Bioinformatic Analysis. *BMC Infect. Dis.* 20, 612–711. doi:10.1186/s12879-020-05335-6
- Lu, P., Lill, H., and Bald, D. (2014). Atp Synthase in Mycobacteria: Special Features and Implications for a Function as Drug Target. *Biochim. Biophys. Acta (Bba) - Bioenerg.* 1837, 1208–1218. doi:10.1016/j.bbabo.2014.01.022
- Melly, G., and Purdy, G. (2019). Mmpl Proteins in Physiology and Pathogenesis of m. tuberculosis. *Microorganisms* 7, 70. doi:10.3390/microorganisms7030070
- Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: Rapid Large-Scale Prokaryote pan Genome Analysis. *Bioinformatics* 31, 3691–3693. doi:10.1093/bioinformatics/btv421
- Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, J. A., et al. (2016). Snp-sites: Rapid Efficient Extraction of Snps from Multi-Fasta Alignments. *Microb. Genomics* 2, e000056. doi:10.1099/mgen.0.000056
- Pattengall, N. D., Alipour, M., Bininda-Emonds, O. R. P., Moret, B. M. E., and Stamatakis, A. (2010). How many Bootstrap Replicates Are Necessary? *J. Comput. Biol.* 17, 337–354. doi:10.1089/cmb.2009.0179
- Pensar, J., Puranen, S., Arnold, B., MacAlasdair, N., Kuronen, J., Tonkin-Hill, G., et al. (2019). Genome-Wide Epistasis and Co-Selection Study Using Mutual Information. *Nucleic Acids Res.* 47, e112. doi:10.1093/nar/gkz656
- Rempel, S., Gati, C., Nijland, M., Thangaratnarajah, C., Karyolaimos, A., de Gier, J. W., et al. (2020). A Mycobacterial Abc Transporter Mediates the Uptake of Hydrophilic Compounds. *Nature* 580, 409–412. doi:10.1038/s41586-020-2072-8
- Sackton, T. B., and Hartl, D. L. (2016). Genotypic Context and Epistasis in Individuals and Populations. *Cell* 166, 279–287. doi:10.1016/j.cell.2016.06.047
- Satta, G., Lipman, M., Smith, G. P., Arnold, C., Kon, O. M., and McHugh, T. D. (2018). *Mycobacterium tuberculosis* and Whole-Genome Sequencing: How Close Are We to Unleashing its Full Potential? *Clin. Microbiol. Infect.* 24, 604–609. doi:10.1016/j.cmi.2017.10.030
- Seemann, T. (2014). Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* 30, 2068–2069. doi:10.1093/bioinformatics/btu153
- Sharma, A. K., Arora, D., Singh, L. K., Gangwal, A., Sajid, A., Molle, V., et al. (2016). Serine/threonine Protein Phosphatase PstP of mycobacterium Tuberculosis Is Necessary for Accurate Cell Division and Survival of Pathogen. *J. Biol. Chem.* 291, 24215–24230. doi:10.1074/jbc.m116.754531
- Skwark, M. J., Croucher, N. J., Puranen, S., Chewapreecha, C., Pesonen, M., Xu, Y. Y., et al. (2017). Interacting Networks of Resistance, Virulence and Core Machinery Genes Identified by Genome-wide Epistasis Analysis. *Plos Genet.* 13, e1006508. doi:10.1371/journal.pgen.1006508
- Sulis, G., and Pai, M. (2020). Isoniazid-resistant Tuberculosis: A Problem We Can No Longer Ignore. *Plos Med.* 17, e1003023. doi:10.1371/journal.pmed.1003023
- Trauner, A., Borrell, S., Reither, K., and Gagneux, S. (2014). Evolution of Drug Resistance in Tuberculosis: Recent Progress and Implications for Diagnosis and Therapy. *Drugs* 74, 1063–1072. doi:10.1007/s40265-014-0248-y
- Tukey, J. (1977). *Exploratory Data Analysis*. Reading, Massachusetts: Addison-Wesley.
- WHO (2021). *Catalogue of Mutations in mycobacterium Tuberculosis Complex and Their Association with Drug Resistance*. Geneva: World Health Organization.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Posada-Reyes, Balderas-Martínez, Ávila-Ríos, Vinuesa and Fonseca-Coronado. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.