UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Posgrados

Secuenciamiento del genoma completo de *Mycobacterium tuberculosis*, variabilidad genética entre aislados de tuberculosis pulmonar y extrapulmonar

Tesis en torno a una hipótesis o problema de investigación y su contrastación

Carolina Susana Hernández Lucero

Daniel Garzón-Chávez, Ph.D.
Director de Trabajo de Titulación

Trabajo de titulación de posgrado presentado como requisito
para la obtención del título de Máster en Microbiología

Quito, 2025

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

COLEGIO DE POSGRADOS

HOJA DE APROBACIÓN DE TRABAJO DE TITULACIÓN

Whole genome sequencing of *Mycobacterium tuberculosis*, genetic variability between isolates from patients with extrapulmonary tuberculosis and pulmonary tuberculosis

Carolina Susana Hernández Lucero

Nombre del Director del Programa:	Patricio Rojas Silva
Título académico:	M.D. PhD.
Director del programa de:	Maestría en Microbiología

Nombre del Decano del colegio Académico:	Carlos Valle
Título académico:	Ph.D.
Decano del Colegio:	COCIBA

Nombre del Decano del Colegio de Posgrados:	Dario Niebieskikwiat
Título académico:	Ph.D.

Quito, 20 de enero del 2025

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombre del estudiante:	Carolina Susana Hernández Lucero
Código de estudiante:	00331875
C.I.:	172356020-5
Lugar y fecha:	Quito, 20 de enero de 2025.

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en http://bit.ly/COPETheses.

UNPUBLISHED DOCUMENT

Note: The following graduation project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on http://bit.ly/COPETheses.

DEDICATORIA

A mi madre, Soledad y mis tíos Susana y Carlos, por ser mi mayor inspiración y mi refugio en cada paso del camino. Gracias por su amor incondicional, su apoyo constante y por enseñarme que los sueños se alcanzan con esfuerzo, perseverancia. Este logro es tanto mío como suyo, porque siempre han creído en mí incluso cuando yo dudaba.

AGRADECIMIENTOS

Quiero expresar mi más profundo agradecimiento a todas las personas e instituciones que hicieron posible la realización de este trabajo. A la Universidad San Francisco de Quito, por la oportunidad invaluable de becarme y permitirme adentrarme en el campo de la microbiología. Al Instituto de Microbiología y al laboratorio de la Facultad de Medicina, por brindarme los recursos necesarios y un entorno académico enriquecedor que fueron fundamentales para llevar a cabo esta investigación.

A Daniel Garzón mi director de tesis por su guía, paciencia y apoyo a lo largo de este proceso. Su experiencia y orientación fueron fundamentales para el desarrollo y la finalización de esta investigación. Mi gratitud también se extiende a Álvaro y Pamela, quienes con su colaboración, discusiones enriquecedoras y apoyo logístico hicieron que este proyecto fuera posible.

A mi familia, quienes han sido mi pilar inquebrantable de apoyo, les expreso mi más profundo agradecimiento por su amor incondicional, su comprensión y su confianza en cada etapa de mi formación académica.

A las nuevas amistades que conseguí en este periodo de mi vida, en especial, a Domenica Arguello, por compartir conmigo largas horas de estudio, por sus valiosos consejos y por todo lo que me ha enseñado, siendo una inspiración y una amiga excepcional en este camino.

RESUMEN

La tuberculosis se manifiesta predominantemente en su forma pulmonar; no obstante, entre el 15 % y el 20 % de los casos afectan otros órganos, presentándose como tuberculosis extrapulmonar, una condición que en ciertos casos está asociada con tasas elevadas de mortalidad. Los mecanismos genéticos que favorecen la diseminación de *Mycobacterium tuberculosis* fuera del pulmón aún no se comprenden completamente. En este estudio, se empleó la secuenciación del genoma completo mediante tecnología Nanopore para comparar cuatro aislamientos extrapulmonares con dos aislamientos pulmonares obtenidos de pacientes atendidos en un hospital de referencia en Ecuador, con el objetivo de identificar variantes genéticas asociadas a la tuberculosis extrapulmonar.

El uso de secuenciación Nanopore permitió la identificación precisa de regiones genómicas complejas y elementos repetitivos debido a la capacidad de generar lecturas largas, una característica esencial para resolver las estructuras genómicas complicadas del patógeno. El análisis reveló 220 SNPs no sinónimos comunes en los aislados extrapulmonares, concentrados principalmente en genes PE/PPE (asociados a la evasión inmune). Además, el gen Rv0986 el cual cumple un rol importante en la capacidad del patógeno para atravesar la barrera hematoencefálica, facilitado su diseminación hacia el sistema nervioso central presento una mutación no sinónima.

Este estudio proporciona información valiosa sobre variantes genéticas relacionadas con la tuberculosis extrapulmonar; sin embargo, se requieren investigaciones adicionales que consideren factores del hospedador, como variantes genéticas y condiciones inmunológicas, para confirmar estas asociaciones y dilucidar su impacto biológico en el desarrollo de la enfermedad.

Palabras clave: Extrapulmonar, *Mycobacterium tuberculosis*, secuenciación del genoma completo, Nanopore, variaciones genéticas.

ABSTRACT

Tuberculosis primarily manifests in its pulmonary form; however, approximately 15–20% of cases extend beyond the lungs, presenting as extrapulmonary tuberculosis. This condition is often associated with significant clinical challenges, including higher mortality rates in certain presentations. The genetic mechanisms facilitating the dissemination of *Mycobacterium tuberculosis* beyond the pulmonary environment remain incompletely understood. In this study, whole-genome sequencing using Nanopore technology was employed to compare four extrapulmonary isolates with two pulmonary isolates obtained from patients treated at a referral hospital in Ecuador. The objective was to identify genetic variants potentially associated with the development of extrapulmonary tuberculosis.

Nanopore sequencing facilitated the precise identification of complex genomic regions and repetitive elements, leveraging its ability to generate long reads, a crucial feature for resolving the intricate genomic structures characteristic of *Mycobacterium tuberculosis*. The analysis revealed 241 non-synonymous SNPs common in extrapulmonary isolates, mainly concentrated in PE/PPE genes (associated with immune evasion). In addition, the Rv0986 gene, which plays an important role in the ability of the pathogen to cross the blood-brain barrier, facilitating its dissemination to the central nervous system, presented a non-synonymous mutation.

This study provides information on genetic variants related to extrapulmonary tuberculosis; however, further research considering host factors, such as genetic variants and immunological conditions, is required to confirm these associations and elucidate their biological impact on disease development.

Key words: Extrapulmonary, *Mycobacterium tuberculosis*, whole genome sequencing, Nanopore, genetic variations.

TABLA DE CONTENIDO

DEDICATORIA **5**
AGRADECIMIENTOS **6**
RESUMEN **7**
ABSTRACT **9**
TABLA DE CONTENIDO **11**
ÍNDICE DE TABLAS **12**
ÍNDICE DE FIGURAS **13**
INTRODUCCIÓN **14**
METODOLOGÍA Y DISEÑO DE LA INVESTIGACIÓN 18
Aislamientos de *Mycobacterium tuberculosis* 18
Extracción de ADN 18
Preparación de las librerías y secuenciación 20
Control de calidad y clasificación taxonómica de los datos de secuenciación 20
Llamada de variantes 20
Ensamblaje del genoma completo y obtención del perfil MIRU-VNTR 21
Determinación del linaje y sublinaje 21
RESULTADOS 22
Características de los aislamientos de *Mycobacterium tuberculosis* 22
Distribución de sublinajes entre aislados extrapulmonares y pulmonares 22
Relación con cepas MIRU publicadas previamente en Ecuador 23
Distribución de variaciones genéticas en aislados de *Mycobacterium tuberculosis* 25
DISCUSIÓN **29**
CONCLUSIONES **34**
REFERENCIAS **35**
ÍNDICE DE ANEXOS 43

ÍNDICE DE TABLAS

Tabla 1. Información clínica de los aislados de *Mycobacterium tuberculosis* usados en este estudio..... 19
Tabla 2. Evaluación de la Calidad del Ensamblaje en Aislados de *Mycobacterium tuberculosis*..... 22
Tabla 3. Distribución de linajes y sublinajes en aislados de *Mycobacterium tuberculosis* identificados con TB-profiler, Fastlin y SNP-IT..... 23
Tabla 4. Variaciones genéticas presentes en cada aislado de *Mycobacterium tuberculosis*..... 25
Tabla 5. SNPs en genes de importancia en la patogenicidad de aislados de *Mycobacterium tuberculosis*..... 27
Tabla 6. SNPs no sinónimos identificados únicamente en aislados extrapulmonares de *Mycobacterium tuberculosis*..... 27
Tabla 7. Variaciones estructurales identificados en aislados de *Mycobacterium tuberculosis* 28

ÍNDICE DE FIGURAS

Figura 1. Árbol de expansión mínima de los aislamientos Mycobacterium tuberculosis de Ecuador usando MIRU-VNTR. 24
Figura 2. Clasificación categórica funcional de SNPs no sinónimos de aislados extrapulmonares 26

INTRODUCCIÓN

La tuberculosis, causada por el bacilo *Mycobacterium tuberculosis*, es la segunda causa de muerte por un único agente infeccioso a nivel mundial. Según la Organización Mundial de la Salud (OMS), en 2023 se reportaron 10.8 millones de casos nuevos de tuberculosis, con una mortalidad de 1.25 millones, incluidos los casos de coinfección con VIH (World Health Organization (WHO), 2024)

Aunque la tuberculosis afecta predominantemente los pulmones, entre el 15% y el 20% de los casos en personas VIH negativas corresponden a formas extrapulmonares, mientras que la proporción se incrementa al 50% en personas VIH positivas, quienes tienen un riesgo mayor de desarrollar tuberculosis extrapulmonar (Das et al., 2013).

La tuberculosis extrapulmonar representa un desafío clínico y epidemiológico significativo debido a su diversidad en la presentación y los órganos afectados. Entre las formas más comunes, la tuberculosis ganglionar es la más prevalente, comprendiendo aproximadamente el 50% de los casos extrapulmonares y afectando principalmente los ganglios linfáticos cervicales. La tuberculosis pleural sigue en frecuencia, representando cerca del 18% de los casos, la tuberculosis osteoarticular 6%, y afecta huesos y articulaciones, la meningitis tuberculosa, aunque menos frecuente del 5% al 10% de los casos, es especialmente preocupante debido a su alta tasa de mortalidad y complicaciones graves (Haulman et al., 2008; Heye et al., 2011; Torok, 2015). Otras formas, como la tuberculosis peritoneal y genitourinaria, completan el cuadro clínico y varían en prevalencia dependiendo de factores como la inmunosupresión y las condiciones socioeconómicas. Estas manifestaciones no solo complican el diagnóstico y tratamiento, sino que también subrayan la necesidad de una mejor comprensión de los mecanismos de diseminación del patógeno (Baykan et al., 2022).

La secuenciación del genoma completo (Whole Genome Sequencing, WGS) se ha convertido en una herramienta fundamental para comprender las variaciones genéticas relacionadas con la patogenicidad, la virulencia y las interacciones huésped-patógeno de *Mycobacterium tuberculosis* (Meehan et al., 2019). Aunque las plataformas de lectura corta, como Illumina, son comúnmente utilizadas, presentan limitaciones significativas para su implementación en entornos con alta prevalencia de tuberculosis, principalmente debido a los elevados costos de los reactivos, la infraestructura requerida y la necesidad de personal altamente capacitado (McNerney et al., 2017).

En este contexto, las tecnologías de secuenciación de lectura larga, como las desarrolladas por Oxford Nanopore Technologies, representan una alternativa prometedora. Dispositivos portátiles como el MinION son más asequibles y accesibles para laboratorios en países en vías de desarrollo. Además, las tecnologías Nanopore son particularmente útiles para el análisis de regiones genómicas complejas y elementos repetitivos, características prominentes en el genoma de *Mycobacterium tuberculosis* (Cervantes et al., 2020).

El análisis de la diversidad genética mediante la secuenciación del genoma completo (WGS, por sus siglas en inglés) ha permitido una clasificación filogenética más precisa y robusta de los linajes que conforman el complejo *Mycobacterium tuberculosis*. Además, esta tecnología mejora la resolución para distinguir cepas que presentan una alta similitud genética. Este aspecto tiene una relevancia clínica significativa, ya que métodos como la electroforesis en gel de polimorfismo de longitud de fragmentos de restricción (RFLP) y los análisis de repeticiones en tándem de unidades intercaladas de micobacterias (MIRU-VNTR) presentan limitaciones para diferenciar entre recaídas y reinfecciones en casos de tuberculosis recurrente. Estas restricciones técnicas pueden comprometer la precisión del diagnóstico, conduciendo a enfoques terapéuticos inadecuados y aumentando la probabilidad de fallos en el tratamiento.

Además, estas metodologías no detectan infecciones mixtas dentro de un mismo hospedador, lo que subestima la complejidad del cuadro clínico y puede contribuir al desarrollo y propagación de tuberculosis multirresistente (Bryant et al., 2013; Guerra-Assunção et al., 2015).

Actualmente, se reconocen siete linajes principales asociados a infecciones humanas: linaje 1 Indo-Oceánico, linaje 2 Beijing, linaje 3 Centroasiático, linaje 4 Euroamericano, linaje 5 Africano Occidental 1, linaje 6 Africano Occidental 2 y linaje 7 Etiopía los cuales comparten aproximadamente el 99.9% de identidad en la secuencia genómica. Los linajes 2 y 4 se destacan por su amplia distribución global. Aunque la diversidad genética de *Mycobacterium tuberculosis* es relativamente baja en comparación con otros patógenos puesto que no realiza transferencia horizontal de genes ni utiliza plásmidos, estudios previos han identificado que la acumulación mutilaciones cromosómicas se dan como resultado de deleciones, duplicaciones, inserciones y polimorfismos de un solo nucleótido (SNP) que permiten analizar la diversidad genética en los diferentes linajes (Chiner-Oms et al., 2022). Estas variaciones podrían influir en cambios fenotípicos, como diferencias en virulencia, capacidad de diseminación, patogenicidad y resistencia a los fármacos (Coscolla & Gagneux, 2014).

Diversos estudios epidemiológicos y experimentos en modelos animales han explorado la posible asociación entre linajes o genotipos específicos y la manifestación de tuberculosis extrapulmonar. No obstante, los hallazgos actuales son inconsistentes y no concluyentes, lo que dificulta atribuir de manera definitiva un papel específico a las variaciones genéticas exclusivas en cada linaje para la presentación extrapulmonar de la enfermedad (Li et al., 2024).

Se ha identificado que las proteinas ESAT-6, HBHA y CFP-10 desempeñan un papel clave en la capacidad de *Mycobacterium tuberculosis* para diseminarse desde el tejido pulmonar hacia el sistema linfático, facilitando su propagación en el huésped (Gao et al., 2004; Smith

et al., 2008). Este proceso está relacionado con la respuesta inmune del huésped, ya que la migración de *Mycobacterium tuberculosis* al sistema linfático contribuye a generar una respuesta inmunitaria temprana mediante la activación de células T, pero la bacteria puede diseminarse hematológicamente y a través del sistema linfático, lo que llevaría a una afectación extrapulmonar (Krishnan et al., 2010).

Por otro lado, se ha demostrado una asociación significativa entre mutaciones en los genes *TIRAP* y *TLR-2* y una mayor susceptibilidad a la meningitis tuberculosa, en contraste con la tuberculosis pulmonar. Estos hallazgos destacan la relevancia de los factores genéticos del hospedador en la predisposición a formas extrapulmonares de tuberculosis (Hawn et al., 2006; Thuong et al., 2007).

La pared celular de *Mycobacterium tuberculosis* desempeña un papel esencial en la fisiopatología de las infecciones extrapulmonares, actuando como una barrera multifuncional que permite la supervivencia del patógeno en condiciones adversas. Entre sus funciones clave, destaca la inhibición de la fusión entre el fagosoma y el lisosoma, un mecanismo que evita la destrucción bacteriana por parte de las células fagocíticas. El éxito evolutivo de *Mycobacterium tuberculosis* radica en su capacidad de persistir dentro de los macrófagos, donde se replica como un patógeno intracelular. Por esta razón, las variaciones en familias de genes como PE/PPE y ESX son de particular interés, ya que estos genes están presentes en la pared celular y podrían estar implicados en mecanismos clave de evasión de la respuesta inmune (Akhter et al., 2012).

Bajo estas circunstancias, la comparación de las variaciones genéticas entre cepas extrapulmonares y pulmonares obtenidas del mismo entorno hospitalario busca determinar si dichas diferencias están relacionadas con el fenotipo extrapulmonar, lo que podría aportar nuevas perspectivas sobre los mecanismos genéticos involucrados en la diseminación.

METODOLOGÍA Y DISEÑO DE LA INVESTIGACIÓN

Aislamientos de *Mycobacterium tuberculosis*

Para el presente estudio se seleccionaron 20 aislamientos de *Mycobacterium tuberculosis* cultivados y mantenidos en el Hospital del Instituto Ecuatoriano de Seguridad Social Quito Sur. Los aislados se obtuvieron de pacientes diagnosticados con tuberculosis en alos años 2021 (n=2), 2022 (n=6), 2023 (n=9) y 2024 (n=3). Este estudio fue aprobado por comité de ética de la Universidad San Francisco de Quito (código 2022-037M).

Extracción de ADN

La extracción de ADN se llevó a cabo utilizando aislados obtenidos en medios de cultivo líquidos MGIT y sólidos Lowenstein-Jensen, siguiendo un protocolo de extracción con fenol-cloroformo basado en el método descrito por (De Almeida et al., 2013), con algunas modificaciones para optimizar la eficiencia del proceso.

Para comenzar, se tomaron varias colonias de los cultivos y se resuspendieron en 1 ml de agua ultrapura. La suspensión bacteriana se sometió a un tratamiento térmico de inactivación mediante incubación a 80°C durante 15 minutos para lisar las células y liberar el ADN. Posteriormente, las muestras se centrifugaron a 14,000 RPM durante 10 minutos para separar los restos celulares. El sobrenadante se descartó, y el sedimento obtenido se utilizó como punto de partida para el procedimiento de extracción.

Se añadió 200 µL de solución de lisozima (10 mg/mL) al sedimento y se incubó durante 1 hora a 37°C. Posteriormente se añadieron 200 µL de solución de proteinasa K (10 mg/mL) y la mezcla se incubó a 55°C durante 1 hora la solución se almacenó a -20°C durante la noche.

El protocolo de extracción con fenol-cloroformo incluyó la adición de volúmenes iguales de fenol y cloroformo al sedimento 400 uL en total, con la finalidad de desnaturalizar

proteínas y otros componentes celulares. Las muestras se agitaron vigorosamente y se centrifugaron a 12000 RPM durante 15 minutos nuevamente para separar las fases. El ADN, localizado en la fase acuosa, se recuperó cuidadosamente y se sometió a una segunda extracción con isopropanol y acetato de sodio para garantizar su pureza.

El ADN extraído fue lavado utilizando etanol absoluto helado y resuspendido en 30 µl de agua buffer TE para su almacenamiento. La pureza fue evaluada mediante espectrofotometría, midiendo las relaciones de absorbancia 260/280 y 230/280 nm con un espectrofotómetro NanoDrop. La concentración se determinó utilizando el equipo Qubit con el kit 1X dsDNA, broad range. Finalmente, la integridad del ADN fue verificada mediante electroforesis en gel de agarosa al 2%. Únicamente el ADN de ocho muestras presento la calidad adecuada para proceder con el secuenciamiento.

Tabla 1. Información clínica de los aislados de *Mycobacterium tuberculosis* usados en este estudio. LCR, liquido cefalorraquídeo; LBA, lavado bronco alveolar; M, masculino; F, femenino

Muestra	Origen	Año	Edad	Sexo	Comorbilidades	Medio de cultivo
EPTB1	Liquido pleural	2021	30	M	0	MGIT
EPTB2	LCR	2022	84	M	0	MGIT
EPTB3	Orina	2022	26	M	0	MGIT
EPTB4	Absceso	2024	71	M	VIH+	Lowenstein-Jensen
PTB1	Esputo	2023	48	F	0	MGIT
PTB2	LBA	2023	62	M	0	MGIT
PTB3	LBA	2023	36	M	0	MIGIT
PTB4	LBA	2024	77	F	0	Lowenstein-Jensen

Preparación de las librerías y secuenciación

La secuenciación se realizó con la plataforma MinION Mk1B (Oxford Nanopore Technologies, Oxford, Reino Unido) con la celda de flujo R10 versión (FLO-MIN114) y el kit Rapid Barcoding (SQK-RBK114.24). Se usaron 20 ng/ul de ADN genómico como material genético de partida para la preparación de librería. Se ejecutó la secuenciación por 48 horas y se usó Guppy v6.4.6 para traducir las señales generadas por el secuenciador en secuencias de ADN que se almacenan en archivos FASTQ.

Control de calidad y clasificación taxonómica de los datos de secuenciación

Los datos de secuenciación en formato FASTQ fueron sometidos a un control de calidad y filtrado utilizando tres herramientas. Inicialmente, Porechop_ABI v0.5.0 fue empleado para eliminar adaptadores de las secuencias. Posteriormente, las lecturas fueron procesadas con NanoFilt v2.8.0 para descartar aquellas con una calidad menor a 10 o con una longitud inferior a 500 pb. Finalmente, las lecturas filtradas fueron clasificadas taxonómicamente mediante Kraken 2 v2.1.3, utilizando la base de datos k2_Mycobacterium (Iñaki & Peter, 2019).

Llamada de variantes

Las lecturas filtradas se procesaron para la identificación de variantes utilizando la herramienta Snippy v4.6.0. Las secuencias se mapearon contra el genoma de referencia de *Mycobacterium tuberculosis* H37Rv (NC_000962.3). Los parámetros de Snippy se configuraron para garantizar la identificación de variantes de alta confianza, estableciendo una calidad mínima de base de 30, una cobertura mínima de lectura de 10, una frecuencia alélica mínima de 0.9 y un 60% de cobertura de mapeo (Gan et al., 2020). Se realizó la búsqueda manual de variantes genéticas en común entre aislados extrapulmonares. Las variantes no sinónimas de importancia se buscaron en Mycobrowser para asignarle su categoría funcional.

Ensamblaje del genoma completo y obtención del perfil MIRU-VNTR

El genoma de cada muestra fue ensamblado *De Novo* utilizando la herramienta Flye v2.9, asegurando la generación de secuencias genómicas completas y continuas. La calidad del ensamblaje se evaluó mediante métricas clave, como el número de contigs, la longitud total del ensamblaje, el N50 y la completitud, calculadas con Quast v5.2.0. Solo los genomas que cumplieron con los criterios de calidad previamente establecidos fueron seleccionados para análisis posteriores.

A partir de estos genomas de alta calidad, se extrajeron las secuencias correspondientes a los 24 loci MIRU-VNTR utilizando la herramienta MIRUreader. Este análisis permitió generar un perfil MIRU único para cada muestra. Los perfiles obtenidos fueron comparados con datos de estudios previos realizados en Ecuador (Castro-Rodriguez et al., 2024; Garzon-Chavez et al., 2020), con el objetivo de identificar si alguna de las cepas analizadas había sido reportada previamente en el país, este análisis se realizó en la aplicación web MIRU-VNTR plus disponible en https://www.miru-vntrplus.org/MIRU/index.faces.

Determinación del linaje y sublinaje

Para la identificación del linaje y sublinaje de cada muestra, se empleó la herramienta bioinformática SNP-IT v1.1, la cual analizó los datos generados a partir del llamado de variantes procesados con Snippy. Adicionalmente, las herramientas Fastlin v0.2.1 y TB-Profiler se utilizaron para analizar directamente las lecturas limpias. Los resultados se compararon para evaluar la concordancia entre la genotipificación basada en lecturas limpias y aquella derivada de datos procesados mediante herramientas de llamado de variantes. Además, TB-Profiler proporcionó información complementaria sobre los perfiles de resistencia a antibióticos.

RESULTADOS

Características de los aislamientos de *Mycobacterium tuberculosis*

Se realizó la extracción de ADN de un total de 20 aislados de *Mycobaterium tuberculosis*. Sin embargo, únicamente ocho de las muestras presentaron la calidad adecuada para proceder con la secuenciación. Entre estas, el ensamblaje del genoma se completó con éxito en seis muestras, cuyos detalles específicos se describen en la Tabla 1. De las muestras con ensamblaje exitoso, cuatro correspondieron a tejidos extrapulmonares y dos a muestras pulmonares, lo que permitió explorar diferencias genómicas según el tipo de muestra.

Tabla 2. Evaluación de la Calidad del Ensamblaje en Aislados de *Mycobacterium tuberculosis*

Muestra	Lecturas totales	Fragmento	N50	Longitud del fragmento	Profundidad	GC%
EPTB1	4441753	1	4441753	4441753	111	65.62
EPTB2	4401091	1	4401091	4401091	44	65.63
EPTB3	4404191	3	4395352	4395352	140	65.6
EPTB4	4411069	1	4411069	4411069	50	65.63
PTB1	3872559	59	119100	361866	30	71.95
PTB2	4402306	1	4402306	4402306	46	65.64
PTB3	890870	109	10504	35841	3	66.94
PTB4	4421037	1	4421037	4421037	81	65.61

Distribución de sublinajes entre aislados extrapulmonares y pulmonares

A partir de las seis muestras analizadas, se identificaron los linajes y sublinajes correspondientes. Los resultados obtenidos fueron consistentes entre todas las herramientas bioinformáticas utilizadas TB-Profiler, fastlin y SNP-IT, confirmando que las seis muestras pertenecen al linaje 4 (Euroamericano). Dentro de los aislados extrapulmonares, se identificaron los sublinajes Haarlem, LAM, X y T. En las muestras pulmonares, los sublinajes detectados fueron LAM y Haarlem como se describe en la tabla 3.

Tabla 3. Distribución de linajes y sublinajes en aislados de *Mycobacterium tuberculosis* identificados con TB-profiler, Fastlin y SNP-IT

Muestra	Linaje	Sublinaje	Spoligotype código octal	MIRU-VNTR Loci set 24
EPTB1	Euroamericano	Haarlem	737733777760731	231214342334445163 34ND732
EPTB2	Euroamericano	X-type	500036436260770	223224342734425143 33ND632
EPTB3	Euroamericano	LAM	777777607760751	134234332224126133 33ND73
EPTB4	Euroamericano	T	471277706660770	224222522324225153 33ND522
PTB2	Euroamericano	Haarlem	771676736660770	223235322534425153 33ND732
PTB4	Euroamericano	LAM	774077607760731	244244332324226133 13ND922

Relación con cepas MIRU publicadas previamente en Ecuador

La distribución de cinco cepas de *Mycobacterium tuberculosis* del presente estudio reveló un alto grado de diversidad genética, evidenciado por la ausencia de agrupamientos en complejos clonales. Esto se debe a que las cepas presentaron más de una variación en los 24 loci analizados mediante MIRU-VNTR, lo que sugiere diferencias genotípicas significativas entre las mismas.

Sin embargo, una excepción fue la muestra EPTB2, que mostró una aparente asociación clonal con una cepa previamente reportada en Guayaquil en el año 2014 (Garzon-Chavez et al., 2020). Esta relación fue identificada debido a la similitud en sus perfiles MIRU-VNTR manteniendo una sola diferencia en el loci MIRU 52 de los 24 loci analizados.

En la Figura 1 se presenta la distribución de las variaciones genotípicas y su comparación con datos previos. Para facilitar la interpretación, se incluyeron únicamente las cepas más relacionadas con nuestras muestras. El gráfico completo, que incluye todos los datos, está disponible como **Figura Suplementaria 1**.

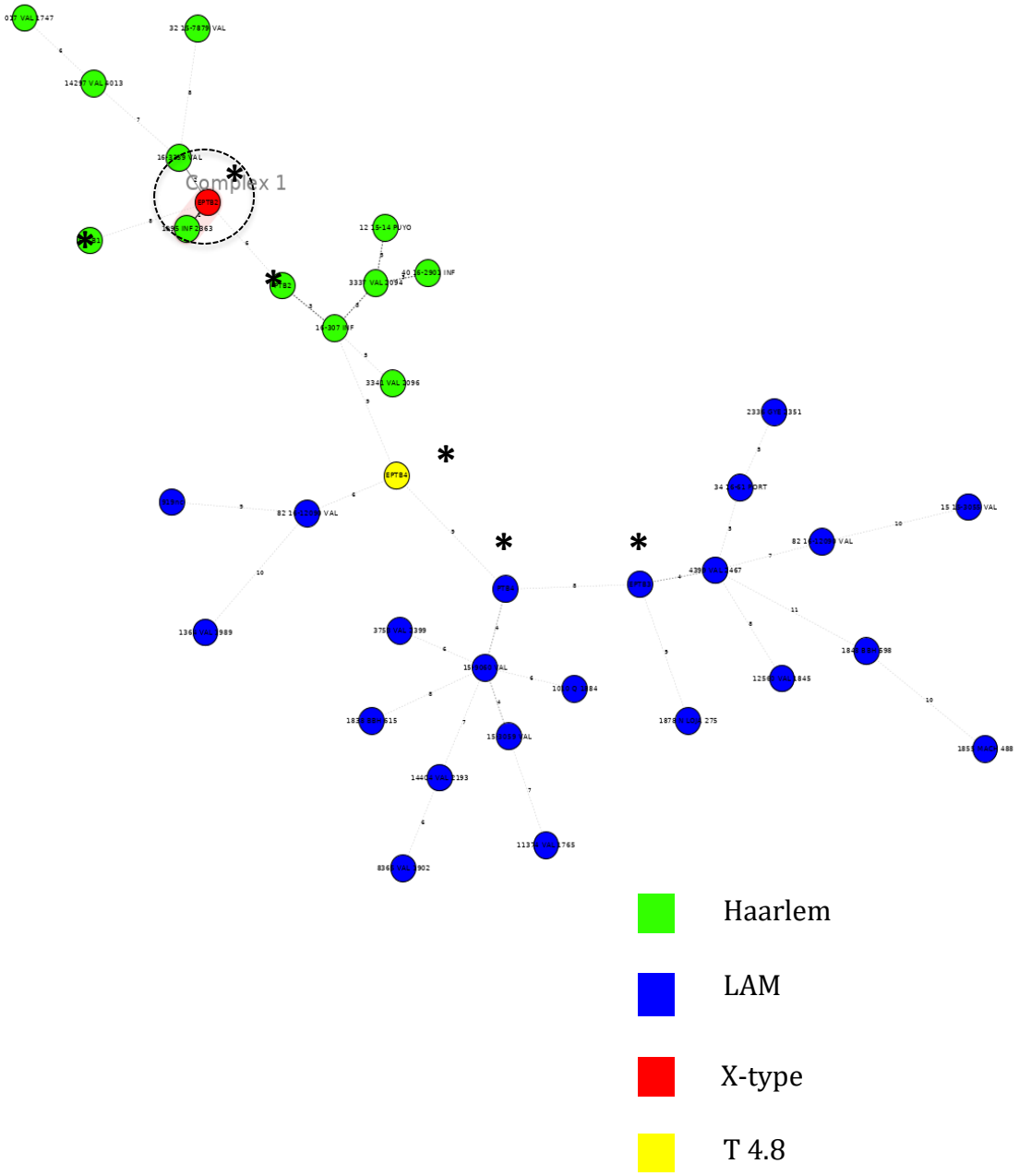

Figura 1. Árbol de expansión mínima de los aislamientos *Mycobacterium tuberculosis* de Ecuador usando MIRU-VNTR. El sublinaje LAM está representado en azul, Haarlem en verde, X en rojo y T en amarillo. Nuestros aislados se identifican con un asterisco y el circulo representa el complejo clonal. Los demás aislados son datos de estudios previos realizados en Ecuador.

Distribución de variaciones genéticas en aislados de *Mycobacterium tuberculosis*

Se identificaron 2,473 variaciones genéticas en los aislados de *Mycobacterium tuberculosis*, y se clasificaron en 2341 polimorfismos de un solo nucleótido SNPs y 132 variantes estructurares, incluyendo deleciones, inserciones, polimorfismos de múltiples nucleótidos y variantes complejas. Los SNPs fueron las variaciones más predominantes. Las muestras EPTB1, EPTB3 y PTB2 destacaron por presentar un mayor número de SNPs, indicando una elevada diversidad genética en comparación con otros aislados. De manera similar las variaciones estructurales se presentaron con mayor frecuencia en aislados extrapulmonares. Estos resultados están descritos en la **Tabla 4.**

Tabla 4. Variaciones genéticas presentes en cada aislado de *Mycobacterium tuberculosis*

	Variaciones genéticas					
Muestras	**SNPs Single Nucleotide Polymorphism**	**MNTs Multi-Nucleotide Polymorphism**	**Deleción**	**Inserción**	**Compleja**	**Total**
EPTB1	534	5	10	6	12	567
EPTB2	339	3	2	3	6	353
EPTB3	449	4	6	3	9	471
EPTB4	270	2	1	9	7	289
PTB1	320	3	3	3	13	342
PTB2	429	3	4	5	10	451
Total	2341	20	26	29	57	2473

Entre los aislados extrapulmonares, se anotaron 988 variaciones, independientemente de su presencia en aislados pulmonares, de las cuales 579 correspondieron a SNPs no sinónimos y 132 variaciones estructurales. Estos resultados detallados se presentan en la **Tabla Suplementaria 1**. Además, 220 SNPs no sinónimos en 61 genes estaban presentes en 1, 2 o 3 de los aislados extrapulmonares. La clasificación categórica funcional de los genes que

presentaron estas mutaciones se encontró principalmente en genes PE-PPE 27%, seguidas por genes relacionados con metabolismo intermediario y respiración con un 25% y genes relacionados a la pared celular con un 23%. Entre ellos destacan *Rv0986, Rv3823c (mmpL8), Rv3408 (vapC47), Rv2830c (vapB22), Rv2931 (ppsA), Rv3892c (PPE69), Rv2946c (pks1), Rv0532 (PE_PGRS6) y Rv2947c (pks15)*. La mayoría de estos genes codifican proteínas de función putativa. Los detalles específicos de estas variaciones se presentan en la Tabla 5.

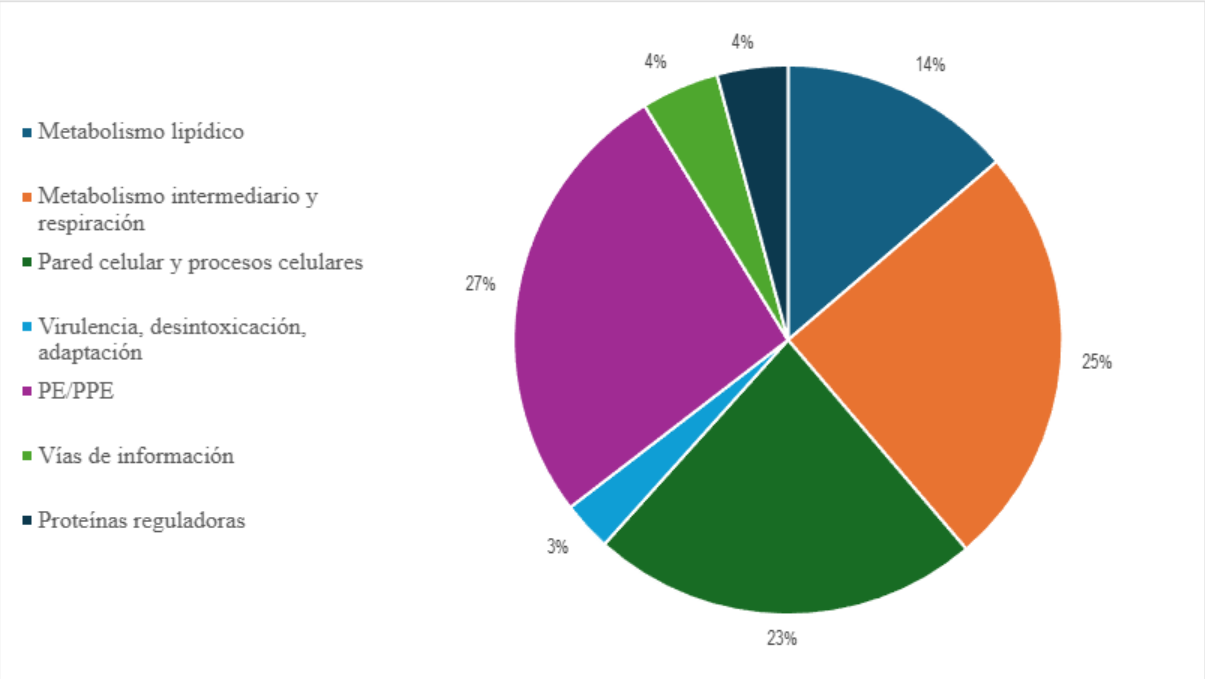

Figura 2. Clasificación categórica funcional de SNPs no sinónimos de aislados extrapulmonares

Tabla 5. SNPs en genes de importancia en la patogenicidad de aislados de *Mycobacterium tuberculosis*

LOCUS	GEN	EFECTO	FUNCION
Rv0986		c.532G>C p.Asp178His	Interviene en la adhesión y la virulencia
Rv3823c	*mmpL8*	c.2533G>T p.Ala845Ser	Interviene en la producción de un glicolípido sulfatado, el sulfolípido-1
Rv3408	*vapC47*	c.137C>T p.Ser46Leu	Sistema toxina antitoxina
Rv2830c	*vapB22*	c.167C>T p.Ala56Val	Sistema toxina antitoxina
Rv2931	*ppsA*	c.4069G>A p.Ala1357Thr	Participa en la biosíntesis de fenoltiocerol y ftiocerol dimicocerosato
Rv3892c	*PPE69*	c.56C>A p.Thr19Lys	Función desconocida
Rv2946c	*pks1*	c.877C>T p.Gln293*	Metabolismo lipídico
Rv2947c	*pks15*	c.998T>C p.Val333Ala	Metabolismo lipídico

También identificamos variaciones que se encontraron en aislados extrapulmonares y no estaban presentes en los aislados pulmonares los cuales se detallan en la **Tabla 6.**

Tabla 6. SNPs no sinónimos identificados únicamente en aislados extrapulmonares de *Mycobacterium tuberculosis*

LOCUS	GEN	EFECTO	FUNCION
Rv1028c	*kdpD*	c.1102C>T p.Pro368Ser	Reguladora en sistema de dos componentes
Rv1662	*pks8*	c.4069G>A p.Ala1357Thr	Metabolismo lipídico
Rv0800	*pepC*	c.416T>G p.Leu139Arg	Metabolismo intermediario y respiración
Rv2741	*PE_PGRS47*	c.811A>G p.Ser271Gly	PE-PPE

En los aislados extrapulmonares analizados, las variaciones estructurales identificadas se localizaron principalmente en genes relacionados con la síntesis y mantenimiento de la pared celular, como aquellos involucrados en la biosíntesis de lípidos y glicolípidos complejos. Estas alteraciones podrían tener implicaciones funcionales significativas, dado el papel central de la pared celular en la virulencia. Adicionalmente, se detectaron variaciones en genes cuya función

aún no ha sido caracterizada, lo que destaca la necesidad de investigaciones futuras para evaluar su posible impacto en la biología del *Mycobacterium tuberculosis*. Los detalles específicos de las variaciones estructurales identificadas, incluyendo su localización genómica, efectos y posibles funciones, se presentan en la **Tabla 7**.

Tabla 7. Variaciones estructurales identificados en aislados de *Mycobacterium tuberculosis*

LOCUS	GEN	EFECTO	FUNCION
Rv0198c	*zmp1*	c.1185_1188delCAAAinsG p.Lys396del	Función desconocida
Rv0405	*pks6*	c.1536_1537delCGinsA p.Gly513fs	Metabolismo lipídico
Rv0305c	*PPE6*	c.2429delT p.Leu810fs	Función desconocida
Rv1198	*esxL*	c.97_98delCGinsTC p.Arg33Ser	Pared celular
Rv1844c	*gnd1*	c.272_277delACGCTG p.Asp91_Ala92del	Participa en la derivación de la hexosa monofosfato

DISCUSIÓN

Este estudio analizó el genoma completo de *Mycobacterium tuberculosis* a partir de cuatro aislados extrapulmonares y dos pulmonares mediante tecnología de lectura larga de Oxford Nanopore, para identificar diferencias genotípicas entre las diferentes afecciones de la enfermedad en Ecuador. En el país no se ha realizado estudios que analicen fenotipos extrapulmonares, los estudios previos se han centrado en analizar variaciones relacionadas a la resistencia a antibióticos (Garzon-Chavez et al., 2020; Morey-León et al., 2022, 2023)

Encontramos que todos los aislados de nuestro estudio pertenecen al linaje 4 Euroamericano, específicamente a los sublinajes LAM, Haarlem, X y T. El linaje 4 es el más prevalente en Latinoamérica, lo que coincide con su alta frecuencia en Ecuador (Joseph & Lindo, 2023; Mokrousov et al., 2016). Así mismo los sublinajes LAM y Haarlem predominan en el país, mientras que el sublinaje X se detecta en solo un 4% (Garzon-Chavez et al., 2020). El sublinaje T se identificó en un 4% de los casos estudiados en la región costera; sin embargo, la información a nivel nacional sobre este sublinaje es limitada (Morey-León et al., 2023). Las principales técnicas moleculares usadas para estudiar la estructura poblacional de *Mycobacterium tuberculosis* en Ecuador son el genotipado mediante MIRU-VNTR y espoligotipado, a pesar de que los últimos 20 años estas técnicas han sido el estándar de oro presentan limitaciones importantes al diferenciar entre cepas estrechamente relacionadas, lo que puede llevar a una subestimación de la diversidad genética en poblaciones con alta homogeneidad, los perfiles MIRU pueden coincidir en cepas de diferentes linajes debido a convergencia evolutiva o limitaciones en los loci seleccionados (Comas et al., 2009).

Al comparar los perfiles MIRU de nuestros aislados con cepas previamente reportadas, observamos que la muestra EPTB2, perteneciente al sublinaje X, formó un complejo clonal con cepas del sublinaje Haarlem esto sucede debido a que estas técnicas de genotipificación

examinan regiones del genoma que son repetitivas o altamente conservadas, es por esto que no pueden detectar variaciones en otras regiones del genoma provocando una identificación errónea a nivel del sublinaje, estas técnicas son útiles para estudios epidemiológicos pero deficientes en análisis filogenéticos. De manera similar (Garzon-Chavez et al., 2020), reportaron dificultades para identificar correctamente cepas del sublinaje X mediante la tipificación MIRU-VNTR. No obstante, la secuenciación del genoma completo permite analizar diferencias a nivel genómico, como los SNPs, lo que proporciona una resolución significativamente mayor para distinguir entre cepas estrechamente relacionadas desde un punto de vista filogenético (Rasoahanitralisoa et al., 2017).

A menudo se considera que la patogénesis de *Mycobacterium tuberculosi*s está influenciada por el linaje, ya que ciertos genes determinantes de virulencia se asocian específicamente con algunos linajes (Reed et al., 2004). Sin embargo, estos genes también pueden estar presentes en otros linajes, lo que sugiere que el linaje por sí solo no explica completamente la patogénesis de la enfermedad. En nuestras muestras no se encontró un sublinaje consenso entre aislados extrapulmonares ni entre los pulmonares, por lo tanto, se decidió analizar las variaciones de todos los aislados en conjunto.

En este contexto, identificamos 220 SNPs no sinónimos en 61 genes anotados en el genoma de *Mycobacterium tuberculosis* H37Rv (NC_000962.3), entre estos destacan *Rv0532* (*PE_PGRS6*) c.716C>G y c.680A>G, *Rv1662 (pks8)* c.4069G>A, *Rv2101 (helZ)* c.1802C>A, *Rv2584c (apt)* c.440C>A, c.364A>G, *Rv3347c (PPE55)*, *Rv3408* (*vapC47)* c.137C>T identificados en nuestros resultados y que se han relacionado a fenotipos extrapulmonares causantes de meningitis tuberculosa por (Faksri et al., 2018), sin embargo, asociación entre estos genes y la diseminación fuera del pulmón aún no se ha comprobado experimentalmente.

Los genes PE/PPE presentaron mayor cantidad de mutaciones en nuestros resultados con relación a los demás genes, *Rv2741 (PE_PGRS47), Rv0532, (PE_PGRS6)* presentaron mutaciones en la mayoría de los resultados extrapulmonares y no estuvieron presentes en los pulmonares. Las proteínas codificadas por los genes PE/PPE, conocidas por su secreción o localización en la superficie celular, han sido propuestas como factores de virulencia clave, ya que participan en la modulación de la respuesta inmune del hospedador. Debido a su relevancia funcional, estas proteínas son consideradas puntos críticos de variaciones genéticas de alto impacto (Worakitchanon et al., 2024). Algunos genes de esta familia están agrupados con genes similares a ESAT-6, una familia de proteínas ampliamente reconocida por su capacidad para estimular potentes respuestas de células T, lo que resalta su relevancia en la interacción hospedador-patógeno (Akhter et al., 2012; Banu et al., 2002). Entre nuestros resultados, se identificó una variación no sinónima en el gen esxL consiste en una deleción e inserción combinada (c.97_98delCGinsTC, p.Arg33Ser). Esta variación podría ser relevante, ya que estudios previos han demostrado que la interrupción de esta proteína confiere una ventaja en el crecimiento de *Mycobacterium tuberculosis* en condiciones in vitro (DeJesus et al., 2017).

En cuanto a genes relacionados a la pared celular identificamos en nuestro estudio una mutación no sinónima c.532G>C p.Asp178His en el gen *Rv0986* que codifica una proteína transportadora ABC (ATP Binding Cassette) (Be et al., 2008) identificaron que este gen es importante para la invasión y supervivencia de *Mycobacterium tuberculosis* en el sistema nervioso central de un modelo murino. De forma similar se identificó un SNP en el gen *Rv3823c, mmpL8* codifica una proteína transmembrana que exporta trehalosa monomicolato (TMM) precursor de los ácidos micólicos que conforman la estructura de la pared celular (Bolla, 2020; Ma et al., 2020) estas propiedades son fundamentales para la adaptación y persistencia de la bacteria en sitios extrapulmonares (Melly & Purdy, 2019). También los genes

Rv2947c (pks15) y *Rv2946c (pks1)* y *Rv2931 (ppsA)* tienen variaciones de importancia ya que participan en la síntesis de lípidos como phthiocerol dimycocerosate (PDIM) esenciales en la conformación de la pared celular micobacteriana y la persistencia dentro de los macrófagos (DeJesus et al., 2017). (Reed et al., 2004) reporto variaciones en genes pks1 y pks15 producen cepas hipervirulentas.

Los múltiples factores de virulencia de *Mycobacterium tuberculosis* ayudan al crecimiento, latencia y diseminación del microorganismo, en este contexto, los genes VapBC involucrados en sistemas de toxina antitoxina regulan el crecimiento bajo estrés, hemos identificado varias mutaciones en estos genes, principalmente *Rv1982 (vapC36)* presente en un aislado extrapulmonar (Arcus et al., 2011).

Entre las variaciones estructurales identificadas en *Mycobacterium tuberculosis*, se observó una mutación en el gen *Rv0198c (zmp1)* que resulta en la pérdida de una lisina en la posición 396 (p.Lys396del). Esta deleción podría impactar la estructura o función de la proteína Zmp1, una metaloproteasa de zinc clave en la virulencia de MTB. La lisina, al ser crucial en interacciones de carga y posiblemente en la unión al sitio activo del zinc, podría alterar la capacidad de Zmp1 para inhibir el inflamasoma (Master et al., 2008). *Rv1844c (gnd1)*, presenta la eliminación de seis nucleótidos que no alteran el marco de lectura, una enzima clave en la generación de NADPH, crucial para contrarrestar el estrés oxidativo y mantener el balance redox, además se ha identificado que la interrupción de este gen proporciona una ventaja de crecimiento para el crecimiento in vitro de H37Rv (DeJesus et al., 2017).

Algunas de las variaciones genéticas identificadas en este estudio podrían estar relacionadas con la capacidad de *Mycobacterium tuberculosis* para diseminarse fuera del tejido pulmonar, particularmente aquellas asociadas con los genes de las familias *PE/PPE, pks*, y *esx*. Estas familias codifican proteínas involucradas en la virulencia, evasión inmune y la interacción

patógeno-hospedador, lo que podría facilitar la invasión a otros tejidos. No obstante, es fundamental analizar un mayor número de muestras representativas de diferentes linajes para evitar sesgos y comprender mejor las características genéticas subyacentes a las manifestaciones extrapulmonares.

Además, resulta relevante explorar las variaciones genéticas en el hospedador que podrían incrementar la susceptibilidad a la tuberculosis extrapulmonar. Estudios previos han sugerido que polimorfismos en genes relacionados con la inmunidad innata y adaptativa, como aquellos que codifican citoquinas (IL-12, IFN-γ) o receptores tipo Toll (TLR), podrían influir en la predisposición a formas extrapulmonares de la enfermedad. Este enfoque podría complementar los hallazgos genómicos del patógeno y proporcionar una visión integral de la interacción huésped-patógeno en contextos extrapulmonares (Caws et al., 2008).

Una limitación significativa de este estudio radica en el uso exclusivo de la secuenciación Nanopore. Aunque esta tecnología permite el análisis de regiones genómicas complejas y elementos repetitivos, característicos del genoma de *Mycobacterium tuberculosis*, presenta una mayor tasa de error en comparación con la secuenciación de lectura corta Illumina, que ofrece mayor precisión y cobertura. No obstante, Illumina también tiene limitaciones, particularmente en la resolución de regiones estructuralmente complejas y de repetición larga. En vista de estas limitaciones, sugerimos la implementación de un enfoque de secuenciación híbrida que combine la precisión de Illumina para corregir los errores de las lecturas largas de Nanopore y la capacidad de Nanopore para resolver regiones difíciles (Di Marco et al., 2023). El uso conjunto de ambas tecnologías permitiría mejorar la detección de variantes genéticas, superando así las limitaciones inherentes a cada tecnología de forma individual.

CONCLUSIONES

En resumen, nuestro estudio ha identificado que los aislados de *Mycobacterium tuberculosis* provenientes de pacientes con tuberculosis extrapulmonar presentan variantes genéticas que podrían estar asociadas con la diseminación del patógeno fuera del pulmón. A través de los análisis de SNPs y variaciones estructurales, observamos una acumulación de mutaciones en genes PE/PPE, los cuales están implicados en la evasión del sistema inmune, aunque la función exacta de la mayoría de estos genes sigue siendo desconocida. Entre las variantes no sinónimas más destacadas, se encuentra una mutación en el gen Rv0986, el cual juega un papel crucial en la capacidad de *Mycobacterium tuberculosis* para sobrevivir en el sitema nervioso central de modelos murinos, lo que podría tener implicaciones cruciales en la patogénesis de la tuberculosis extrapulmonar.

REFERENCIAS

Akhter, Y., Ehebauer, M. T., Mukhopadhyay, S., & Hasnain, S. E. (2012). The PE/PPE multigene family codes for virulence factors and is a possible source of mycobacterial antigenic variation: Perhaps more? *Biochimie*, *94*(1), 110-116. https://doi.org/10.1016/j.biochi.2011.09.026

Arcus, V. L., McKenzie, J. L., Robson, J., & Cook, G. M. (2011). The PIN-domain ribonucleases and the prokaryotic VapBC toxin-antitoxin array. *Protein Engineering Design and Selection*, *24*(1-2), 33-40. https://doi.org/10.1093/protein/gzq081

Banu, S., Honoré, N., Saint-Joanis, B., Philpott, D., Prévost, M., & Cole, S. T. (2002). Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? *Molecular Microbiology*, *44*(1), 9-19. https://doi.org/10.1046/j.1365-2958.2002.02813.x

Baykan, A. H., Sayiner, H. S., Aydin, E., Koc, M., Inan, I., & Erturk, S. M. (2022). Extrapulmonary tuberculosıs: An old but resurgent problem. *Insights into Imaging*, *13*(1), 39. https://doi.org/10.1186/s13244-022-01172-0

Be, N. A., Lamichhane, G., Grosset, J., Tyagi, S., Cheng, Q., Kim, K. S., Bishai, W. R., & Jain, S. K. (2008). Murine Model to Study the Invasion and Survival of *Mycobacterium tuberculosis* in the Central Nervous System. *The Journal of Infectious Diseases*, *198*(10), 1520-1528. https://doi.org/10.1086/592447

Bolla, J. R. (2020). Targeting MmpL3 for anti-tuberculosis drug development. *Biochemical Society Transactions*, *48*(4), 1463-1472. https://doi.org/10.1042/BST20190950

Bryant, J. M., Harris, S. R., Parkhill, J., Dawson, R., Diacon, A. H., Van Helden, P., Pym, A., Mahayiddin, A. A., Chuchottaworn, C., Sanne, I. M., Louw, C., Boeree, M. J., Hoelscher, M., McHugh, T. D., Bateson, A. L. C., Hunt, R. D., Mwaigwisya, S.,

Wright, L., Gillespie, S. H., & Bentley, S. D. (2013). Whole-genome sequencing to establish relapse or re-infection with Mycobacterium tuberculosis: A retrospective observational study. *The Lancet Respiratory Medicine*, *1*(10), 786-792. https://doi.org/10.1016/S2213-2600(13)70231-5

Castro-Rodriguez, B., Franco-Sotomayor, G., Orlando, S. A., & Garcia-Bereguiain, M. Á. (2024). Molecular epidemiology of Mycobacterium tuberculosis in Ecuador: Recent advances and future challenges. *Journal of Clinical Tuberculosis and Other Mycobacterial Diseases*, *37*, 100465. https://doi.org/10.1016/j.jctube.2024.100465

Caws, M., Thwaites, G., Dunstan, S., Hawn, T. R., Thi Ngoc Lan, N., Thuong, N. T. T., Stepniewska, K., Huyen, M. N. T., Bang, N. D., Huu Loc, T., Gagneux, S., Van Soolingen, D., Kremer, K., Van Der Sande, M., Small, P., Thi Hoang Anh, P., Chinh, N. T., Thi Quy, H., Thi Hong Duyen, N., … Farrar, J. (2008). The Influence of Host and Bacterial Genotype on the Development of Disseminated Disease with Mycobacterium tuberculosis. *PLoS Pathogens*, *4*(3), e1000034. https://doi.org/10.1371/journal.ppat.1000034

Cervantes, J., Yokobori, N., & Hong, B.-Y. (2020). Genetic Identification and Drug-Resistance Characterization of Mycobacterium tuberculosis Using a Portable Sequencing Device. A Pilot Study. *Antibiotics*, *9*(9), 548. https://doi.org/10.3390/antibiotics9090548

Chiner-Oms, Á., López, M. G., Moreno-Molina, M., Furió, V., & Comas, I. (2022). Gene evolutionary trajectories in *Mycobacterium tuberculosis* reveal temporal signs of selection. *Proceedings of the National Academy of Sciences*, *119*(17), e2113600119. https://doi.org/10.1073/pnas.2113600119

Comas, I., Homolka, S., Niemann, S., & Gagneux, S. (2009). Genotyping of Genetically Monomorphic Bacteria: DNA Sequencing in Mycobacterium tuberculosis Highlights the Limitations of Current Methodologies. *PLoS ONE*, *4*(11), e7815. https://doi.org/10.1371/journal.pone.0007815

Coscolla, M., & Gagneux, S. (2014). Consequences of genomic diversity in Mycobacterium tuberculosis. *Seminars in Immunology*, *26*(6), 431-444. https://doi.org/10.1016/j.smim.2014.09.012

Das, S., Roychowdhury, T., Kumar, P., Kumar, A., Kalra, P., Singh, J., Singh, S., Prasad, H., & Bhattacharya, A. (2013). Genetic heterogeneity revealed by sequence analysis of Mycobacterium tuberculosis isolates from extra-pulmonary tuberculosis patients. *BMC Genomics*, *14*(1), 404. https://doi.org/10.1186/1471-2164-14-404

De Almeida, I. N., Da Silva Carvalho, W., Rossetti, M. L., Costa, E. R. D., & De Miranda, S. S. (2013). Evaluation of six different DNA extraction methods for detection of Mycobacterium tuberculosis by means of PCR-IS6110: Preliminary study. *BMC Research Notes*, *6*(1), 561. https://doi.org/10.1186/1756-0500-6-561

DeJesus, M. A., Gerrick, E. R., Xu, W., Park, S. W., Long, J. E., Boutte, C. C., Rubin, E. J., Schnappinger, D., Ehrt, S., Fortune, S. M., Sassetti, C. M., & Ioerger, T. R. (2017). Comprehensive Essentiality Analysis of the *Mycobacterium tuberculosis* Genome via Saturating Transposon Mutagenesis. *mBio*, *8*(1), e02133-16. https://doi.org/10.1128/mBio.02133-16

Di Marco, F., Spitaleri, A., Battaglia, S., Batignani, V., Cabibbe, A. M., & Cirillo, D. M. (2023). Advantages of long- and short-reads sequencing for the hybrid investigation of the Mycobacterium tuberculosis genome. *Frontiers in Microbiology*, *14*, 1104456. https://doi.org/10.3389/fmicb.2023.1104456

Faksri, K., Xia, E., Ong, R. T.-H., Tan, J. H., Nonghanphithak, D., Makhao, N., Thamnongdee, N., Thanormchat, A., Phurattanakornkul, A., Rattanarangsee, S., Ratanajaraya, C., Suriyaphol, P., Prammananan, T., Teo, Y.-Y., & Chaiprasert, A. (2018). Comparative whole-genome sequence analysis of Mycobacterium tuberculosis isolated from tuberculous meningitis and pulmonary tuberculosis patients. *Scientific Reports*, *8*(1), 4910. https://doi.org/10.1038/s41598-018-23337-y

Gan, G. L., Nguyen, M. H., Willie, E., Rezaie, M. H., Lee, B., Chauve, C., Libbrecht, M., & Chindelevitch, L. (2020). *Geographic heterogeneity impacts drug resistance predictions in* Mycobacterium tuberculosis. https://doi.org/10.1101/2020.09.17.301226

Gao, L., Guo, S., McLaughlin, B., Morisaki, H., Engel, J. N., & Brown, E. J. (2004). A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion. *Molecular Microbiology*, *53*(6), 1677-1693. https://doi.org/10.1111/j.1365-2958.2004.04261.x

Garzon-Chavez, D., Garcia-Bereguiain, M. A., Mora-Pinargote, C., Granda-Pardo, J. C., Leon-Benitez, M., Franco-Sotomayor, G., Trueba, G., & De Waard, J. H. (2020a). Population structure and genetic diversity of Mycobacterium tuberculosis in Ecuador. *Scientific Reports*, *10*(1), 6237. https://doi.org/10.1038/s41598-020-62824-z

Guerra-Assunção, J. A., Houben, R. M. G. J., Crampin, A. C., Mzembe, T., Mallard, K., Coll, F., Khan, P., Banda, L., Chiwaya, A., Pereira, R. P. A., McNerney, R., Harris, D., Parkhill, J., Clark, T. G., & Glynn, J. R. (2015). Recurrence due to Relapse or Reinfection With *Mycobacterium tuberculosis*: A Whole-Genome Sequencing Approach in a Large, Population-Based Cohort With a High HIV Infection Prevalence

and Active Follow-up. *Journal of Infectious Diseases*, *211*(7), 1154-1163. https://doi.org/10.1093/infdis/jiu574

Haulman, N. J., Hawn, T. R., & Nolan, C. M. (2008). Tuberculosis in Travelers and Immigrants. En *The Travel and Tropical Medicine Manual* (pp. 391-406). Elsevier. https://doi.org/10.1016/B978-141602613-6.10024-2

Hawn, T. R., Dunstan, S. J., Thwaites, G. E., Simmons, C. P., Thuong, N. T., Lan, N. T. N., Quy, H. T., Chau, T. T. H., Hieu, N. T., Rodrigues, S., Janer, M., Zhao, L. P., Hien, T. T., Farrar, J. J., & Aderem, A. (2006). A Polymorphism in Toll-Interleukin 1 Receptor Domain Containing Adaptor Protein Is Associated with Susceptibility to Meningeal Tuberculosis. *The Journal of Infectious Diseases*, *194*(8), 1127-1134. https://doi.org/10.1086/507907

Heye, T., Stoijkovic, M., Kauczor, H.-U., Junghanss, T., & Hosch, W. (2011). Extrapulmonale Tuberkulose: Die radiologische Bildgebung eines fast vergessenen Verwandlungskünstlers. *RöFo - Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*, *183*(11), 1019-1029. https://doi.org/10.1055/s-0031-1273429

Iñaki, & Peter. (2019). *M. tuberculosis bioinformatics training data* (Versión 3) [Dataset]. Zenodo. https://doi.org/10.5281/ZENODO.3960260

Joseph, S. K., & Lindo, J. (2023). The evolutionary history of infectious disease in the ancient AMERICAS and the pathogenic consequences of EUROPEAN contact. *American Journal of Biological Anthropology*, *182*(4), 532-541. https://doi.org/10.1002/ajpa.24595

Krishnan, N., Robertson, B. D., & Thwaites, G. (2010). The mechanisms and consequences of the extra-pulmonary dissemination of Mycobacterium tuberculosis. *Tuberculosis*, *90*(6), 361-366. https://doi.org/10.1016/j.tube.2010.08.005

Li, Y., Li, Y., Liu, Y., Kong, X., Tao, N., Hou, Y., Wang, T., Han, Q., Zhang, Y., Long, F., & Li, H. (2024). Iron-related gene mutations driving global Mycobacterium tuberculosis transmission revealed by whole-genome sequencing. *BMC Genomics*, *25*(1), 249. https://doi.org/10.1186/s12864-024-10152-1

Ma, S., Huang, Y., Xie, F., Gong, Z., Zhang, Y., Stojkoska, A., & Xie, J. (2020). Transport mechanism of *Mycobacterium tuberculosis* MmpL/S family proteins and implications in pharmaceutical targeting. *Biological Chemistry*, *401*(3), 331-348. https://doi.org/10.1515/hsz-2019-0326

Master, S. S., Rampini, S. K., Davis, A. S., Keller, C., Ehlers, S., Springer, B., Timmins, G. S., Sander, P., & Deretic, V. (2008). Mycobacterium tuberculosis Prevents Inflammasome Activation. *Cell Host & Microbe*, *3*(4), 224-232. https://doi.org/10.1016/j.chom.2008.03.003

McNerney, R., Clark, T. G., Campino, S., Rodrigues, C., Dolinger, D., Smith, L., Cabibbe, A. M., Dheda, K., & Schito, M. (2017). Removing the bottleneck in whole genome sequencing of Mycobacterium tuberculosis for rapid drug resistance analysis: A call to action. *International Journal of Infectious Diseases*, *56*, 130-135. https://doi.org/10.1016/j.ijid.2016.11.422

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. *Nature Reviews Microbiology*, *17*(9), 533-545. 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), 70. https://doi.org/10.3390/microorganisms7030070

Mokrousov, I., Vyazovaya, A., Iwamoto, T., Skiba, Y., Pole, I., Zhdanova, S., Arikawa, K., Sinkov, V., Umpeleva, T., Valcheva, V., Alvarez Figueroa, M., Ranka, R., Jansone, I., Ogarkov, O., Zhuravlev, V., & Narvskaya, O. (2016). Latin-American-Mediterranean lineage of Mycobacterium tuberculosis: Human traces across pathogen's phylogeography. *Molecular Phylogenetics and Evolution*, *99*, 133-143. https://doi.org/10.1016/j.ympev.2016.03.020

Morey-León, G., Andrade-Molina, D., Fernández-Cadena, J. C., & Berná, L. (2022). Comparative genomics of drug-resistant strains of Mycobacterium tuberculosis in Ecuador. *BMC Genomics*, *23*(1), 844. https://doi.org/10.1186/s12864-022-09042-1

Morey-León, G., Mejía-Ponce, P. M., Granda Pardo, J. C., Muñoz-Mawyin, K., Fernández-Cadena, J. C., García-Moreira, E., Andrade-Molina, D., Licona-Cassani, C., & Berná, L. (2023). A precision overview of genomic resistance screening in Ecuadorian isolates of Mycobacterium tuberculosis using web-based bioinformatics tools. *PLOS ONE*, *18*(12), e0294670. https://doi.org/10.1371/journal.pone.0294670

Rasoahanitralisoa, R., Rakotosamimanana, N., Stucki, D., Sola, C., Gagneux, S., & Rasolofo Razanamparany, V. (2017). Evaluation of spoligotyping, SNPs and customised MIRU-VNTR combination for genotyping Mycobacterium tuberculosis clinical isolates in Madagascar. *PLOS ONE*, *12*(10), e0186088. https://doi.org/10.1371/journal.pone.0186088

Reed, M. B., Domenech, P., Manca, C., Su, H., Barczak, A. K., Kreiswirth, B. N., Kaplan, G., & Barry, C. E. (2004). A glycolipid of hypervirulent tuberculosis strains that inhibits

the innate immune response. *Nature*, *431*(7004), 84-87. https://doi.org/10.1038/nature02837

Smith, J., Manoranjan, J., Pan, M., Bohsali, A., Xu, J., Liu, J., McDonald, K. L., Szyk, A., LaRonde-LeBlanc, N., & Gao, L.-Y. (2008). Evidence for Pore Formation in Host Cell Membranes by ESX-1-Secreted ESAT-6 and Its Role in *Mycobacterium marinum* Escape from the Vacuole. *Infection and Immunity*, *76*(12), 5478-5487. https://doi.org/10.1128/IAI.00614-08

Thuong, N. T. T., Hawn, T. R., Thwaites, G. E., Chau, T. T. H., Lan, N. T. N., Quy, H. T., Hieu, N. T., Aderem, A., Hien, T. T., Farrar, J. J., & Dunstan, S. J. (2007). A polymorphism in human TLR2 is associated with increased susceptibility to tuberculous meningitis. *Genes & Immunity*, *8*(5), 422-428. https://doi.org/10.1038/sj.gene.6364405

Torok, M. E. (2015). Tuberculous meningitis: Advances in diagnosis and treatment. *British Medical Bulletin*, *113*(1), 117-131. https://doi.org/10.1093/bmb/ldv003

Worakitchanon, W., Yanai, H., Piboonsiri, P., Miyahara, R., Nedsuwan, S., Imsanguan, W., Chaiyasirinroje, B., Sawaengdee, W., Wattanapokayakit, S., Wichukchinda, N., Omae, Y., Palittapongarnpim, P., Tokunaga, K., Mahasirimongkol, S., & Fujimoto, A. (2024). Comprehensive analysis of Mycobacterium tuberculosis genomes reveals genetic variations in bacterial virulence. *Cell Host & Microbe*, *32*(11), 1972-1987.e6. https://doi.org/10.1016/j.chom.2024.10.004

World Health Organization (WHO). (2024). *Global Tuberculosis Report 2024*. World Health Organization. https://iris.who.int/bitstream/handle/10665/379339/9789240101531-eng.pdf?sequence=1

ÍNDICE DE ANEXOS

Figura suplementaria 1. Árbol de expansión mínima de los aislamientos *Mycobacterium tuberculosis* y datos previos de Ecuador usando MIRU-VNTR…..……....… 45

Tabla suplementaria1. Variaciones genéticas comunes en aislados extrapulmonares de *Mycobacterium Tuberculosis* comparados con aislamientos pulmonares……....… 46

Figura suplementaria 1. Árbol de expansión mínima de los aislamientos *Mycobacterium tuberculosis* y datos previos de Ecuador usando MIRU-VNTR.

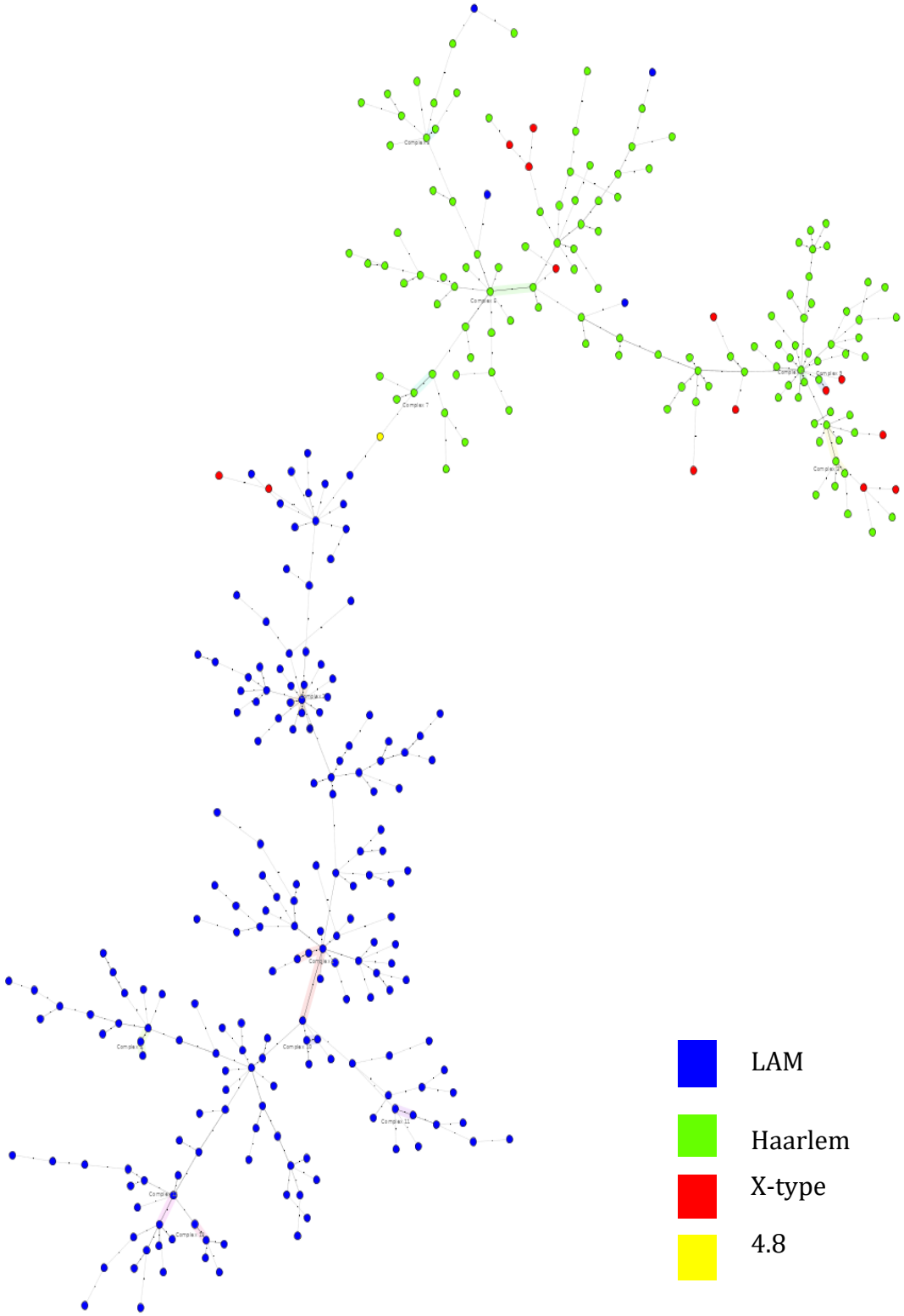

Tabla suplementaria 1. Variaciones genéticas comunes en aislados extrapulmonares de *Mycobacterium Tuberculosis* comparados con aislamientos pulmonares.

SNPs encontrados en la mayoría de aislados extrapulmonares de *Mycobacterium tuberculosis* al comparar con aislados pulmonares					
LOCUS	**GEN**	**EFECTO**	**PRODUCTO**	**Total, EP aislados**	**Total, P aislados**
Rv2090		missense_variant c.1074C>G p.Phe358Leu	5'-3' exonuclease	3	1
Rv1378c		missense_variant c.109C>T p.Arg37Trp	hypothetical protein	4	0
Rv1028c	kdpD	missense_variant c.1102C>T p.Pro368Ser	sensor protein KdpD	3	0
Rv1554	frdC	missense_variant c.118A>G p.Met40Val	fumarate reductase membrane anchor subunit	3	2
Rv3630		missense_variant c.118G>A p.Ala40Thr	integral membrane protein	3	2
Rv3737		missense_variant c.119A>G p.Asp40Gly	transmembrane protein	4	1
Rv3331	sugI	missense_variant c.1268C>T p.Pro423Leu	sugar-transport integral membrane protein SugI	3	1
Rv0507	mmpL2	missense_variant c.1277G>A p.Arg426His	transmembrane transport protein MmpL2	3	1
Rv1313c		missense_variant c.1298T>G p.Leu433Arg	insertion sequence element IS1557 transposase	3	1
Rv0064		missense_variant c.1370G>A p.Gly457Asp	transmembrane protein	3	1
Rv3408	vapC47	missense_variant c.137C>T p.Ser46Leu	ribonuclease VapC47	3	1
Rv1364c		missense_variant c.1394C>A p.Ala465Glu	sigma factor regulatory protein	3	1
Rv3245c	mtrB	missense_variant c.1549A>C p.Met517Leu	two component sensory histidine kinase MtrB	3	1
Rv3151	nuoG	missense_variant c.1810A>G p.Thr604Ala	NADH-quinone oxidoreductase subunit G	3	2
Rv2931	ppsA	missense_variant c.1872C>G p.Asp624Glu	phthiocerol synthesis polyketide synthase type I PpsA	3	2
Rv2377c		missense_variant c.206T>C p.Val69Ala	hypothetical protein	3	1
Rv2807		missense_variant c.215A>T p.Glu72Val	hypothetical protein	3	0
Rv0064		missense_variant c.2197A>G p.Asn733Asp	transmembrane protein	3	2
Rv2512c		missense_variant c.25A>G p.Thr9Ala	insertion sequence element IS1081 transposase	3	1
Rv2627c		missense_variant c.310A>G p.Arg104Gly	hypothetical protein	4	1
Rv1459c		missense_variant c.337A>G p.Lys113Glu	alpha-(1->6)-mannopyranosyltransferase	3	2
Rv2337c		missense_variant c.356G>T p.Gly119Val	hypothetical protein	4	1
Rv1618	tesB1	missense_variant c.362A>T p.His121Leu	acyl-CoA thioesterase II	3	1
Rv1604	impA	missense_variant c.371C>A p.Pro124Gln	inositol-monophosphatase ImpA	3	2
Rv2458	mmuM	missense_variant c.374A>G p.Tyr125Cys	homocysteine S-methyltransferase MmuM	4	1
Rv0192		missense_variant c.379T>C p.Ser127Pro	hypothetical protein	3	1
Rv2982c	gpdA2	missense_variant c.397C>G p.Pro133Ala	glycerol-3-phosphate dehydrogenase	3	1
Rv1662	pks8	missense_variant c.4069G>A p.Ala1357Thr	polyketide synthase	3	0
Rv1374c		missense_variant c.406A>G p.Thr136Ala	hypothetical protein	3	2
Rv0159c	PE3	missense_variant c.40A>G p.Thr14Ala	PE family protein PE3	3	1
Rv0800	pepC	missense_variant c.416T>G p.Leu139Arg	M18 family aminopeptidase	3	0
Rv2398c	cysW	missense_variant c.422G>C p.Gly141Ala	sulfate ABC transporter permease CysW	3	2
Rv3213c		missense_variant c.430A>G p.Lys144Glu	SOJ/ParA-like protein	3	2

Rv2688c		missense_variant c.466C>A p.Pro156Thr	antibiotic ABC transporter ATP-binding protein	3	1
Rv0966c		missense_variant c.524T>C p.Val175Ala	hypothetical protein	3	1
Rv0259c		missense_variant c.545C>T p.Ala182Val	hypothetical protein	3	1
Rv1462		missense_variant c.547A>G p.Asn183Asp	hypothetical protein	4	1
Rv0962c	lprP	missense_variant c.557C>T p.Pro186Leu	lipoprotein LprP	3	1
Rv3892c	PPE69	missense_variant c.56C>A p.Thr19Lys	PPE family protein PPE69	3	1
Rv0746	PE_PGRS9	missense_variant c.572A>G p.Glu191Gly	PE-PGRS family protein PE_PGRS9	3	2
Rv1300	hemK	missense_variant c.580C>T p.Arg194Cys	release factor glutamine methyltransferase	3	1
Rv1602	hisH	missense_variant c.602G>A p.Ser201Asn	imidazole glycerol phosphate synthase subunit HisH	3	2
Rv3093c		missense_variant c.630C>G p.Cys210Trp	oxidoreductase	3	1
Rv3711c	dnaQ	missense_variant c.631G>T p.Val211Leu	DNA polymerase III subunit epsilon	3	2
Rv2999	lppY	missense_variant c.634A>T p.Met212Leu	lipoprotein LppY	3	1
Rv0068		missense_variant c.640G>C p.Val214Leu	oxidoreductase	3	1
Rv0532	PE_PGRS6	missense_variant c.680A>G p.Asp227Gly	PE-PGRS family protein PE_PGRS6	4	1
Rv1644	tsnR	missense_variant c.695T>C p.Leu232Pro	23S rRNA methyltransferase TsnR	3	2
Rv3827c		missense_variant c.754A>G p.Thr252Ala	transposase	3	1
Rv2083		missense_variant c.766C>G p.Leu256Val	hypothetical protein	3	1
Rv2178c	aroG	missense_variant c.795C>A p.Asp265Glu	phospho-2-dehydro-3-deoxyheptonate aldolase AroG	3	2
Rv2741	PE_PGRS47	missense_variant c.811A>G p.Ser271Gly	PE-PGRS family protein PE_PGRS47	3	0
Rv2436	rbsK	missense_variant c.845T>C p.Val282Ala	ribokinase RbsK	3	1
Rv2037c		missense_variant c.935G>A p.Cys312Tyr	transmembrane protein	3	1
Rv0218		missense_variant c.946T>C p.Cys316Arg	transmembrane protein	3	1
Rv2125		missense_variant c.97G>A p.Gly33Ser	hypothetical protein	3	2
Rv3776		missense_variant c.985A>G p.Met329Val	hypothetical protein	3	2
Rv2947c	pks15	missense_variant c.998T>C p.Val333Ala	polyketide synthase	3	2
Rv0986		missense_variant c.532G>C p.Asp178His	adhesion component ABC transporter ATP-binding protein	1	0
Variantes genéticas estructurales encontradas comúnmente entre los aislados extrapulmonares en comparación con aislados pulmonares de *Mycobacterium Tuberculosis*					
Rv0198c	zmp1	disruptive_inframe_deletion&synonymous_variant c.1185_1188delCAAAinsG p.Lys396del	zinc metalloprotease	1	0
Rv0405	pks6	frameshift_variant&synonymous_variant c.1536_1537delCGinsA p.Gly513fs	membrane bound polyketide synthase	1	0
Rv1452c	PE_PGRS28	missense_variant c.1097_1098delATinsCC p.Asp366Ala	PE-PGRS family protein PE_PGRS28	1	1
Rv1452c	PE_PGRS28	missense_variant c.1247_1250delGCCCinsCCAA p.GlyPro416AlaAsn	PE-PGRS family protein PE_PGRS28	1	0
Rv1452c	PE_PGRS28	missense_variant c.1294_1296delCATinsAAC p.His432Asn	PE-PGRS family protein PE_PGRS28	1	0

Rv0980c	PE_PGRS18	missense_variant c.213_217delCCAGGinsTCAGC p.Glu73Gln	PE-PGRS family protein PE_PGRS18	1	0
Rv0279c	PE_PGRS4	missense_variant c.229_230delGTinsAC p.Val77Thr	PE-PGRS family protein PE_PGRS4	1	0
Rv2931	ppsA	missense_variant c.2407_2412delGCCCGGinsACTCGC p.Ala803Thr	phthiocerol synthesis polyketide synthase type I PpsA	1	1
Rv0978c	PE_PGRS17	missense_variant c.316_328delGCACCCGTTCAGTinsACGCCGACTGAGG p.AlaProValGlnSer106ThrProThrGluAla	PE-PGRS family protein PE_PGRS17	1	0
Rv0095c		missense_variant c.383_386delGCCGinsCCCT p.GlyArg128AlaLeu	hypothetical protein	1	0
Rv0406c		missense_variant c.4204_4205delCGinsAC p.Arg1402Thr	beta lactamase-like protein	1	0
Rv1361c	PPE19	missense_variant c.489_493delCGAGGinsGGCGA p.GluAla164AlaThr	PPE family protein PPE19	1	0
Rv1361c	PPE19	missense_variant c.574_575delATinsTC p.Ile192Ser	PPE family protein PPE19	1	0
Rv3077		missense_variant c.924_928delTGTGGinsGGTGC p.Gly310Arg	hydrolase	4	2
Rv1198	esxL	missense_variant c.97_98delCGinsTC p.Arg33Ser	ESAT-6 like protein EsxL	1	0
Rv3590c	PE_PGRS58	conservative_inframe_deletion c.1534_1542delGCCGGCGGC p.Ala512_Gly514del	PE-PGRS family protein PE_PGRS58	1	0
Rv1847		disruptive_inframe_deletion c.176_181delTCGAGA p.Ile59_Glu60del	esterase	1	0
Rv1844c	gnd1	disruptive_inframe_deletion c.272_277delACGCTG p.Asp91_Ala92del	6-phosphogluconate dehydrogenase	1	0
Rv2100		disruptive_inframe_deletion c.806_814delGCGCGCTGG p.Gly269_Leu271del	hypothetical protein	1	0
Rv1725c		frameshift_variant c.125delG p.Gly42fs	hypothetical protein	1	0
Rv0670	end	frameshift_variant c.174delC p.Tyr59fs	endonuclease IV	1	0
Rv0305c	PPE6	frameshift_variant c.2429delT p.Leu810fs	PPE family protein PPE6	2	2
Rv3911	sigM	frameshift_variant c.478delC p.Arg160fs	ECF RNA polymerase sigma factor SigM	1	0
Rv1115		frameshift_variant c.74_78delCGCTG p.Thr25fs	hypothetical protein	1	0
Rv2113		conservative_inframe_insertion c.1069_1071dupCGC p.Arg357dup	integral membrane protein	1	0
Rv0977	PE_PGRS16	frameshift_variant c.1969dupG p.Ala657fs	PE-PGRS family protein PE_PGRS16	1	0
Rv0354c	PPE7	frameshift_variant c.374dupG p.Ile126fs	PPE family protein PPE7	1	0
Rv1575		frameshift_variant c.389dupG p.Val131fs	phage protein	1	0
Rv3897c		frameshift_variant c.496_497insCCCCG p.Gly166fs	hypothetical protein	1	0

Rv0392c	ndhA	frameshift_variant c.851dupC p.Leu285fs	NADH dehydrogenase NdhA	1	0
Rv3284		frameshift_variant c.9_10dupGC p.Pro4fs	hypothetical protein	1	0
Rv3514	PE_PGRS57	missense_variant c.1960_1961delAAinsGG p.Lys654Gly	PE-PGRS family protein PE_PGRS57	1	0
Rv2931	ppsA	missense_variant c.2630_2631delGCinsAT p.Arg877His	phthiocerol synthesis polyketide synthase type I PpsA	3	1
Rv3514	PE_PGRS57	missense_variant c.3507_3509delTGAinsCAC p.Asp1170Thr	PE-PGRS family protein PE_PGRS57	1	0
Rv0978c	PE_PGRS17	missense_variant c.363_364delGAinsCC p.Asn122His	PE-PGRS family protein PE_PGRS17	1	0
Rv3099c		missense_variant c.636_637delGTinsTC p.Ser213Pro	hypothetical protein	1	0
Rv0279c	PE_PGRS4	missense_variant c.76_77delGCinsAG p.Ala26Ser	PE-PGRS family protein PE_PGRS4	1	0
Rv3844		missense_variant c.92_93delATinsTG p.Asp31Val	transposase	1	0