



UNIVERSIDAD MICHOACANA
DE SAN NICOLÁS DE HIDALGO
Cuna de héroes, crisol de pensadores



**UNIVERSIDAD MICHOACANA DE
SAN NICOLÁS DE HIDALGO**

**PROGRAMA INSTITUCIONAL DE DOCTORADO
EN CIENCIAS BIOLÓGICAS
OPCIÓN:
BIOTECNOLOGÍA MOLECULAR**

**Análisis genómico comparativo entre cepas
pulmonares y extrapulmonares de *Mycobacterium
tuberculosis***

TESIS

**Que para obtener el grado de:
DOCTORA EN CIENCIAS BIOLÓGICAS**

PRESENTA:

M. en C. S. Andrea Monserrat Negrete Paz

Directora:

D. en C. Ma. Soledad Vázquez Garcidueñas

Co-director:

D. en C. Gerardo Vázquez Marrufo

Morelia, Michoacán

Agosto 2022

El presente trabajo se desarrolló en el Laboratorio de Genética Molecular Microbiana de la Facultad de Ciencias Médicas y Biológicas “Dr Ignacio Chávez” de la Universidad Michoacana de San Nicolás de Hidalgo, bajo la dirección de la D. en C. Ma. Soledad Vázquez Garcidueñas y la co-dirección del D. en C. Gerardo Vázquez Marrufo de la Facultad de Medicina y Veterinaria y Zootecnia.

La realización de este trabajo fue posible gracias a la beca para estudios de doctorado otorgada por el Consejo Nacional de Ciencia y Tecnología con número de expediente 664711.

COMITÉ TUTORAL:

D. en C. Alejandra Ochoa Zarzosa

D. en C. Joel Edmundo López Meza

D. en C. Octavio Martínez de la Vega

ÍNDICE GENERAL

	Página
RESUMEN	1
SUMMARY.....	3
INTRODUCCIÓN GENERAL.....	5
JUSTIFICACIÓN	26
HIPÓTESIS.....	27
OBJETIVO GENERAL	27
OBJETIVOS ESPECÍFICOS	27
RESULTADOS	28
CAPÍTULO 1: Whole-genome comparative analysis at the lineage/sublineage level discloses relationships between <i>Mycobacterium tuberculosis</i> genotype and clinical phenotype.	28
CAPÍTULO II: Caracterización genómica y fenotípica de 10 cepas de <i>M. tuberculosis</i> aisladas en el estado de Michoacán, México	47
CAPÍTULO III: Identificación de polimorfismos asociados a los fenotipos pulmonar y extrapulmonar de cepas de <i>Mycobacterium tuberculosis</i>.....	89
CAPÍTULO IV: Identificación de genes asociados a los fenotipos pulmonar y extrapulmonar de cepas de <i>Mycobacterium tuberculosis</i>.....	104
CONCLUSIONES GENERALES	127
DISCUSIÓN GENERAL	123
CAPÍTULO V.....	128
ANEXOS.....	138

ÍNDICE DE FIGURAS

Página

1. Representación esquemática de los posibles escenarios tras la exposición a <i>M. tuberculosis</i>.	6
2. Secuenciación del genoma de <i>M. tuberculosis</i>	15
3. Flujo de trabajo estándar para el análisis genómico de cepas del complejo <i>M. tuberculosis</i>.	16
4. Incidencia por año de casos de TB pulmonar y extrapulmonar en Michoacán	49
5. Clasificación filogenética de las cepas MYC	63
6. Rearreglos genómicos en las cepas MYC	67
7. Filogenia de las cepas de <i>M. tuberculosis</i> aisladas en México	70
8. Disrupción génica ocasionada por la inserción de IS6110 en los genomas de cepas de <i>M. tuberculosis</i> aisladas en México	71
9. Comparación del contenido de genes basado en la clasificación funcional asignada por RAST para las cepas MYC.	74
10. Morfología colonial de las cepas MYC en agar BBL Seven H11 adicionado con sangre humana	75
11. Formación de cuerdas microscópicas en las cepas MYC	76
12. Diagrama de Venn de las microvariantes identificadas en las cepas de <i>M. tuberculosis</i>	94
13. Construcción filogenética realizada por Roary para las cepas de estudio a partir del pangenoma	112
14. Distribución de los genes que conforman el pangenoma	112
15. Curvas representativas del contenido de genes en el pangenoma	113

ÍNDICE DE TABLAS

	Página
Tabla 1. Características epidemiológicas de las cepas de estudio.	53
Tabla 2. Características del ADN extraído de las cepas seleccionadas.....	61
Tabla 3. Rendimiento de secuenciación por cepa.....	61
Tabla 4. Asignación de espoligotipo, SIT y familia por Spotyping.....	62
Tabla 5. Comparativo de la asignación de familias por MTBSeq y Spotyping.	64
Tabla 6. Resistencia genotípica a fármacos de primera línea.....	65
Tabla 7. Características de las regiones exclusivas identificadas en los genomas de las cepas MYC.....	66
Tabla 8. Microvariantes por cepa identificadas mediante el mapeo contra el genoma de referencia.....	68
Tabla 9. Secuencias de inserción IS6110 encontradas en los genomas de las cepas MYC.	69
Tabla 10. Características de los ensamblados obtenidos para cada una de las cepas MYC.	72
Tabla 11. Resultados de la anotación en Prokka de los genomas de las cepas MYC.	73
Tabla 12. Resistencia fenotípica a antibióticos de primera línea en las cepas MYC.	76
Tabla 13. Mutaciones asociadas al fenotipo extrapulmonar.....	95
Tabla 14. Mutaciones asociadas al fenotipo pulmonar.....	96
Tabla 15. Genes asociados a los fenotipos PTB y EPTB mediante el análisis del genoma accesorio.	115

RESUMEN

La tuberculosis (TB) es uno de los padecimientos más antiguos de la humanidad que en la actualidad todavía constituye una de las enfermedades transmisibles con mayor tasa de mortalidad. Esta enfermedad afecta a grandes grupos de población, particularmente de áreas marginadas y grupos vulnerables donde predominan la pobreza, desnutrición y el hacinamiento. De acuerdo con la información disponible en el componente de TB del Sistema Nacional de Vigilancia Epidemiológica, en nuestro país se registraron 20,075 mil casos nuevos de tuberculosis en 2021, de los cuales 16,008 fueron casos de TB pulmonar y 455 casos fueron de TB meníngea, una de las formas más severas de la enfermedad. Cabe destacar que en dicho año se observó una disminución en el número de casos nuevos reportados de tuberculosis en comparación con los años 2020 y 2019 a nivel nacional y mundial; sin embargo, esta tendencia no fue un reflejo de la disminución de contagios, sino de las interrupciones en el acceso a los servicios de diagnóstico y tratamiento debido al inicio de la pandemia por COVID-19. No obstante que Michoacán se encuentra por debajo de la media nacional de casos de TB, en los últimos años se ha observado una tendencia de incremento en la tasa de incidencia de tuberculosis pulmonar (PTB) y extrapulmonar (EPTB), fenómeno observable en diversas regiones alrededor del mundo. Este aumento en el número de casos de TB extrapulmonar en los últimos años puede asociarse a diversos factores que incluyen las características inmunogenéticas de la población, sus entornos ecológicos y socioculturales y el fondo genómico de las cepas del patógeno circulantes. A la fecha no existe un consenso sobre los determinantes genéticos que contribuyen a las formas extrapulmonares de la TB. Es por esto que el objetivo del presente trabajo fue buscar determinantes genéticos asociados a la presentación clínica de la TB. Con tal fin se desarrolló una estrategia original de análisis comparativo para analizar cepas de *Mycobacterium tuberculosis* (MTB) de una manera global, accediendo a una gran cantidad de información genómica y metadatos asociados para la comparación de cepas pulmonares y extrapulmonares de MTB. En la caracterización genómica y fenotípica realizada se incluyeron 10 cepas causantes de TB pulmonar y extrapulmonar aisladas en el estado de Michoacán, así como 490 cepas obtenidas de bases de datos públicas

aisladas de siete países diferentes. Para la búsqueda de variaciones genómicas asociadas al fenotipo pulmonar y extrapulmonar, se realizaron alineamientos de las cepas de interés contra el genoma de referencia (H37Rv), así como la construcción del pangenoma. Adicionalmente, se realizó la clasificación filogenómica de las cepas a partir de la asignación de polimorfismos de un solo nucleótido (SNPs), la técnica de mayor poder de discriminación para medir diversidad genética en MTB. Los resultados obtenidos mostraron una asociación de la TB pulmonar con los sublinajes 4.3.1, 4.3.3, 4.5 Asian African 2 y Europe/Russia/B0, los cuales comparten la presencia de dos polimorfismos en el gen *pe_pgrs10* y SNPs característicos en la secuencia de los genes *pe_pgrs4*, *pe_pgrs7*, *pe_pgrs53*, así como con la delección de los genes *hspR*, *plcD*, *pe_pgrs5*, *pe_pgrs25*, *Rv1759c*, *pe_pgrs57*, *Rv817c*, *Rv2818c*, *Rv2816c*, *Rv3740c* y *Rv2550c*. Por otro lado, el fenotipo EPTB se asoció con los sublinajes 1.2.1, 1.1.1, 4.1.2.1, CAS, y Ancestral Beijing, además de la presencia de un SNP exclusivo en el gen *ppe57*, varios SNPs en la secuencia de los genes *ppe54*, *pe_pgrs18*, *ppe60* y *Rv1760*, la presencia del gen *aftC* en el genoma y la delección de los genes *aceA*, *plcA*, *esxR*, *Rv2275* y *ppe50*. Los hallazgos obtenidos en este trabajo permiten concluir que la aparición de un fenotipo clínico de TB no está dada por un elemento genético exclusivo o único, sino por un conjunto de variaciones que pueden desencadenar diferencias en la respuesta inmune del hospedero.

Palabras clave: tuberculosis, secuenciación, resistencia, extrapulmonares, mycobacterium.

SUMMARY

Tuberculosis (TB) is one of the oldest diseases affecting mankind, however, it remains one of the transmissible diseases with the highest mortality rate. This disease affects large population groups, particularly in marginalized areas and vulnerable groups where poverty, malnutrition and overcrowding predominate. According to the information available in the TB component of the National Epidemiological Surveillance System, in our country 20,075 thousand new cases of tuberculosis were registered in 2021, of which 16,008 belong to cases of pulmonary TB and 455 cases belong to meningeal TB, one of the most severe forms of the disease. Notably, a decrease in the number of new reported cases of tuberculosis has been observed compared to the years 2020 and 2019 at the national and global levels, however, this is not because the population does not get sick, it is because of the interruptions in access to diagnostic and treatment services since the start of the COVID-19 pandemic. Even though, Michoacán is below the national average of incidence for TB cases, in recent years there has been a tendency to increase the incidence rate of pulmonary tuberculosis (PTB) and extrapulmonary tuberculosis (EPTB), a phenomenon that can be observed in many regions around the world. This increase in the number of cases of extrapulmonary TB in recent years may be due to various factors: the immunogenic characteristics of the population, its ecological and sociocultural environments, and the genomic background of the circulating strains. To date, there is no consensus on the genetic determinants that contribute to extrapulmonary forms of tuberculosis. Therefore, the present work implements a previously unused genomic analysis strategy that allows the analysis of *Mycobacterium tuberculosis* (MTB) strains in a global way with which an amount of information never explored in the comparison of pulmonary and extrapulmonary strains is accessed. To achieve this objective, a genomic and phenotypic characterization of 10 strains causing pulmonary and extrapulmonary TB in Michoacán was realized, the sequencing of these strains was necessary, because in the databases the Latin American region is not represented in terms of TB extrapulmonary strains. Using these data and a set of 490 strains obtained from public databases that were isolated from 7 different countries, the search of the genomic variants associated with the pulmonary and extrapulmonary phenotype was

performed, using alignments against the reference genome and the construction of the pangenome. Additionally, the phylogenetic classification of the isolates was carried out based on the assignment of single nucleotide polymorphisms, the most discriminating technique to measure genetic diversity in MTB. In the case of pulmonary TB, it was associated with sublineages 4.3.1, 4.3.3, 4.5 Asian African 2 and Europe/Russia/B0, with the presence of 2 polymorphisms in the *pe_pgrs10* gene, with the presence of certain SNPs in the sequence of genes *pe_pgrs4*, *pe_pgrs7*, *pe_pgrs53* and with the deletion of genes *hspR*, *plcD* *pe_pgrs5*, *pe_pgrs25*, *Rv1759c*, *pe_pgrs57*, *Rv817c*, *Rv2818c*, *Rv2816c*, *Rv3740c* and *Rv2550c*. On the other hand, the EPTB phenotype was associated with the 1.2.1, 1.1.1, 4.1.2.1, CAS, and Ancestral Beijing sublineages, in addition to the presence of an exclusive SNP in the *ppe57* gene, several SNPs in the sequence of the *ppe54*, *pe_pgrs18*, *ppe60* and *Rv1760* genes, the presence of the *aftC* gene in the genome and the deletion of the *aceA*, *plcA*, *esxR*, *Rv2275* and *ppe50* genes.

The findings obtained in this work allow us to conclude that the appearance of a clinical phenotype of TB is not given by an exclusive or unique genetic element, but by a set of variations that can trigger differences in the immune response.

INTRODUCCIÓN GENERAL

La tuberculosis (TB) es un problema de salud pública global, siendo considerada una emergencia mundial por la Organización Mundial de la Salud (OMS). En 2020, a nivel mundial se registraron cerca de 10 millones de nuevos casos y aproximadamente 1.6 millones de fallecimientos por TB. De estos, 289,000 casos fueron notificados en América Latina. En el año 2020, México ocupó el tercer lugar en incidencia de casos de TB en América Latina, solo por debajo de Brasil y Perú. A pesar de las políticas de salud implementadas para el diagnóstico y tratamiento oportuno en el país, la TB sigue siendo un problema social y sanitario, subestimado debido a las dificultades en el diagnóstico y el escaso reporte de casos (Yanez-Lema et al., 2021).

Agente causal

El principal agente causal de la TB humana es la actinobacteria *Mycobacterium tuberculosis*, perteneciente al denominado Complejo *Mycobacterium tuberculosis* (MTBC). Las especies de MTBC adaptadas a los humanos son *M. tuberculosis* sensu stricto y *M. africanum*, éstas se dividen en nueve linajes filogenéticos: L1 o Indo-Oceánico; L2, o de Asia oriental; L3, o indio del este de África; L4, o euroamericano; L5 o *M. africanum* de África occidental 1; L6 o *M. africanum* de África occidental 2; L7, o Etiopía (Firdessa et al., 2013; Coscolla y Gagneux, 2014); L8 o *M. tuberculosis* de los Grandes Lagos africanos (Ngabonziza et al., 2020); y el recientemente descrito L9 o *M. africanum* (Coscolla et al., 2021). Todas estas especies comparten más del 99% de identidad en su secuencia de nucleótidos; sin embargo, es posible observar grandes diferencias fenotípicas.

Patogénesis

El encuentro entre las especies del MTB y el hospedero conduce a una respuesta inmune compleja y multifacética que da como resultado cuatro posibles escenarios:

1) el establecimiento de una infección latente, 2) enfermedad tuberculosa activa, 3) enfermedad subclínica o 4) la eliminación completa del patógeno (Figura 1) (De Martino et al., 2019). Después de la infección, la mayoría de los individuos desarrollan una infección latente (LTBI), la cual es definida por la presencia de un test de tuberculina y/o ensayo de interferón gamma positivo, pero sin la presencia de síntomas clínicos detectables. Se estima que alrededor del 5-15% de los individuos con LTBI desarrollarán enfermedad activa (Moller et al., 2018). En estos casos, la bacteria se replica y surgen los síntomas clínicos de la enfermedad, los cuales incluyen: tos persistente y con sangre, fiebre, sudores nocturnos, pérdida de peso y dolor en el pecho.

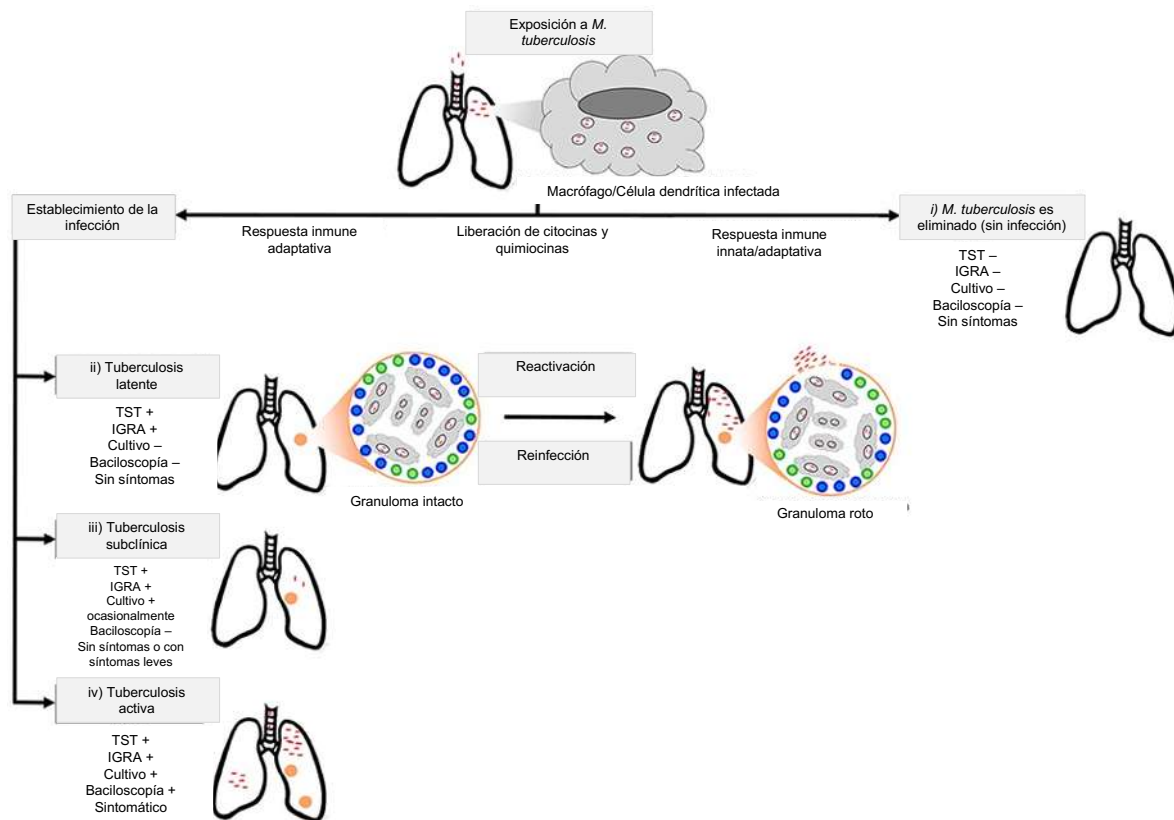


Figura 1. Representación esquemática de los posibles escenarios tras la exposición a MTB. La bacteria ingresa al sistema respiratorio del hospedero mediante la inhalación de gotas de saliva contaminadas que después son fagocitadas por macrófagos y células dendríticas en los pulmones para dar paso a cuatro posibles resoluciones: i) la bacteria es inmediatamente eliminada por el sistema inmune, ii) la bacteria es contenida en los granulomas formados por células de la respuesta inmune adaptativa (incluidas las células T y B) y la infección no progresa a TB activa. Aunque esta contención puede durar toda la vida, bajo condiciones específicas los granulomas pueden romperse ocasionando la diseminación de MTB (reactivación), lo que resulta en la enfermedad activa, iii) enfermedad subclínica caracterizada por síntomas y periodos de contagio intermitentes, y iv) enfermedad tuberculosa activa (Modificado de Moller et al., 2018). Clave: TST, test de tuberculina; IGRA, interferón gamma.

El agente infeccioso ingresa a los pulmones mediante la inhalación de gotas de saliva que contienen al bacilo. La habilidad de la bacteria para invadir las células del hospedero está determinada tanto por su virulencia como por una adecuada respuesta inmune del hospedero. El bacilo utiliza diversos ligandos de superficie para interactuar con los receptores del hospedero que incluyen receptores del complemento, manosa, proteína surfactante A, CD14, receptores “scavenger” de macrófagos y DC-SIGN. Los bacilos inhalados son transportados a los alvéolos pulmonares y fagocitados por macrófagos alveolares, debido a que MTB es un patógeno intracelular es capaz de evadir la respuesta inmune y persistir dentro de los macrófagos por un periodo largo de tiempo. Se sabe que determinadas proteínas presentes en la pared celular micobacteriana, particularmente las glicoproteínas, están asociadas con la evasión de la respuesta inmune. Por ejemplo, se ha demostrado que SodC, una glicoproteína que se encuentra en la pared celular de MTB, mejora la sobrevivencia dentro de los macrófagos (Liao et al., 2013). La eliminación bacteriana depende de la respuesta inmunitaria innata y adaptativa del hospedero en donde los macrófagos juegan un papel crucial. Tras la internalización en los macrófagos, la mayoría de los bacilos mueren, mientras que las bacterias sobrevivientes inhiben la maduración del fagosoma y se replican dentro de ellos, esta replicación conduce a la disrupción de los macrófagos y al desarrollo de la enfermedad tuberculosa activa. Sin embargo, en la mayoría de las personas infectadas los macrófagos alveolares reclutan células dendríticas y otros macrófagos al sitio de la lesión. Las células reclutadas fagocitan, pero rara vez destruyen a los bacilos.

La respuesta inmune adaptativa contra la infección con MTB está mediada principalmente por células T (Correa-Macedo et al., 2019), y se activa aproximadamente 2-3 semanas después de la infección. Se cree que otra glicoproteína, denominada Apa, controla la inmunidad de las células T a través de los receptores de manosa que se encuentran en la superficie de los macrófagos del hospedero (Marimani et al., 2018). Una vez que las células T específicas de antígeno alcanzan el sitio de la infección, se multiplican dentro de las lesiones primarias (Cooper, 2009) e inducen la producción de citocinas proinflamatorias.

Estas citocinas estimulan a los macrófagos para la destrucción de los bacilos intracelulares. Esta etapa particular de la infección se denomina fase latente, en la que se inactiva la replicación bacteriana. Sin embargo, la reactivación de MTB (Tundup et al., 2014; Schubert et al., 2015) puede ocurrir meses o años después cuando el sistema inmunológico del hospedero se ve comprometido. Es importante destacar que hay una serie de factores bacterianos y del hospedero que juegan un papel clave en la reactivación de la infección, además de diversos mecanismos mediante los cuales MTB reprograma los procesos epigenéticos del hospedero para evadir la respuesta inmune después de la infección. En conjunto, estos mecanismos permiten que el patógeno infecte, sobreviva y se replique dentro de las células del hospedero, evitando una respuesta inmune significativa del mismo (Marimani et al., 2018).

Determinantes bacterianos y del hospedero en la TB humana

La progresión de la enfermedad hacia los diferentes posibles escenarios está determinada por diversos factores ambientales y genéticos, tanto del hospedero como de MTB. Numerosos estudios epidemiológicos y experimentales han revelado la presencia de dichos determinantes (Malik et al., 2005; Pareek et al., 2013; Chiner-Oms et al., 2019). Un polimorfismo genético corresponde a varias formas alternativas (alelos) para un mismo gen, que aparecen con una frecuencia $\geq 1\%$ en la población. El tipo más común de polimorfismo es el de un solo nucleótido o SNP (*Single Nucleotide Polymorphism*). Un SNP es una forma de mutación puntual exitosa, desde el punto de vista evolutivo, la cual permanece en una parte significativa de la población de una especie, contribuyendo a la diversidad genética. Se sabe que diferentes polimorfismos genéticos pueden modular la respuesta inmune frente a la infección por MTB (Narendran et al., 2016). Estos se han identificado en antígenos humanos leucocitarios (HLA), receptores tipo Toll, receptores de vitamina D, citocinas con sus receptores y muchos otros componentes funcionales de la inmunidad (Rosain et al., 2019). Además, la susceptibilidad mendeliana a la enfermedad micobacteriana (MSMD) se ha identificado como una condición clínica en la que se presenta susceptibilidad

selectiva a micobacterias poco virulentas en ausencia de una inmunodeficiencia. Esta se caracteriza por un déficit completo del receptor 1 del interferón-gamma que lleva a la alteración de la inmunidad mediada por interferón-gamma y, en consecuencia, a infecciones graves y a menudo letales (Maulén y Fuentes, 2018).

Además de los factores del hospedero, la virulencia bacteriana constituye el otro factor importante a la hora de evaluar el riesgo de infección por MTB. La virulencia no se limita simplemente a la cepa bacteriana o la carga en la secreción respiratoria, sino que tiene en cuenta la expresión diferencial de genes en las diferentes fases de la infección. MTB carece de factores de virulencia clásicos como las toxinas y su capacidad de escape inmunológico depende de la modulación del metabolismo de lípidos, proteínas transportadoras de metales, proteasas y proteínas que inhiben los efectos antimicrobianos de los macrófagos, entre otros (De Martino, 2019).

Un estudio clave en la identificación de genes esenciales para la infección por MTB *in vivo* se realizó en un modelo de ratón. Mediante la hibridación de sondas complementarias de secuencias mutadas por inserción de transposones (TraSH) se obtuvo una primera estimación de la imagen global de los factores de virulencia de MTB, identificándose 194 genes requeridos específicamente para el crecimiento *in vivo* de la bacteria en los ratones (Sasseti y Rubin, 2003). Un estudio similar en macrófagos derivados de precursores de médula ósea de ratones C57BL/6 identificó 126 genes de virulencia, algunos de los cuales resultaron esenciales exclusivamente en el modelo de macrófagos (Rengarajan et al., 2005). En un estudio más reciente que utilizó la secuenciación del genoma completo, se encontraron más de 400 genes adicionales como esenciales *in vivo* (Zhang et al., 2013). Estos datos combinados resaltan el hecho de que MTB utiliza una gran combinación de factores de virulencia en los diferentes estados de la infección, por lo que aún es necesario estudiar genes individuales con más detalle para identificar su papel y funciones exactas en el proceso de patogénesis. Varios genes identificados como esenciales para la virulencia de la bacteria, codifican para proteínas involucradas en el metabolismo básico de este microorganismo, lo que

hace que la definición de un factor de virulencia en MTB sea una tarea complicada (Madacki et al., 2019).

Modelos de infección *in vitro* e *in vivo* han mostrado que diferentes genotipos de MTB inducen diferentes patrones en la respuesta inmune (Reed et al., 2004; Manca et al., 2005). Además, estudios epidemiológicos han mostrado que algunos genotipos podrían estar asociados con la presentación de la TB; sin embargo, los resultados han sido inconsistentes. Por ejemplo, se ha sugerido una asociación entre polimorfismos en el gen *p/c* presente en el cromosoma de MTB y la EPTB (Kong et al., 2003; Yang et al., 2005). Aunque estos antecedentes presentan varias desventajas, como el tratarse de estudios retrospectivos o utilizar un número pequeño de muestra, y no se ha tomado en cuenta el fondo genético o la susceptibilidad inmunológica del paciente.

Diagnóstico y tratamiento de la tuberculosis

El diagnóstico rutinario de la TB consiste en la baciloscopía de una muestra de esputo mediante la tinción de Ziehl-Neelsen. Esta técnica posee varias limitaciones como la incapacidad de detectar pacientes paucibacilares en personas infectadas con VIH y cepas drogo resistentes (OMS, 2021). Aunque esta técnica permite identificar solo alrededor 70-80% de los casos pulmonares positivos, sigue siendo la herramienta de diagnóstico más utilizada debido a su bajo costo y rápida ejecución. Al igual que con la mayoría de las infecciones microbianas, la TB puede ser diagnosticada mediante el uso de pruebas inmunológicas como el ensayo *in vitro* de liberación de interferón-gamma (IGRA) o la prueba cutánea *in vivo* de la tuberculina (TST), aunque ninguno de ellos puede discernir si se trata de una TB activa o de la enfermedad latente (Marimani et al., 2018). La OMS recomienda el uso de la mayor cantidad de herramientas diagnóstico que incluyen IGRA, TST, radiografía de tórax, baciloscopía y cultivo, cuando otras técnicas moleculares de mayor especificidad como Xpert MTB/RIF no están disponibles. El método Xpert MTB/RIF es una prueba de amplificación de ácidos nucleicos totalmente automatizada que emplea un cartucho para diagnosticar la TB y la resistencia a la

rifampicina. Este método automatizado purifica el ADN, lo concentra y amplifica mediante una prueba de reacción en cadena de la polimerasa en tiempo real, identificando secuencias específicas del genoma de MTB. Los resultados se obtienen a partir de muestras de esputo sin procesar en menos de 2 h, con empleo de tiempo mínimo por parte del personal técnico (PAHO, 2021).

Los fármacos de primera línea utilizados para el tratamiento de la TB incluyen rifampicina, isoniazida, etambutol, pirazinamida y estreptomina. Los de segunda línea comprenden fluoroquinolonas, kanamicina, capreomicina, amikacina, viomicina, etionamida, ácido para-aminosalicílico, linezolida y clofazimina, además de los fármacos de uso reciente como bedaquilina, delamanida y protamanida (Khawbung et al., 2021). Los perfiles de resistencia a antibióticos se definen con base en la resistencia e incluyen a las cepas: susceptibles (sin resistencia), drogo resistentes (resistentes a por lo menos un antibiótico), multidrogo resistentes o MDR (resistentes a rifampicina e isoniazida), de resistencia extendida o XDR (resistentes a isoniazida, rifampicina, fluoroquinolonas) y al menos un agente inyectable (amikacina, capreomicina, kanamicina), pre-XDR (resistentes a isoniazida y rifampicina y también a una fluoroquinolona (FQ) o a un inyectable de segunda línea; pero no a ambos (OMS, 2021) y cepas totalmente resistentes a antibióticos o TDR en las cuales se observa resistencia a todos los fármacos de primera y segunda línea. Estos aislados constituyen una amenaza grave para el control mundial de la TB (Khawbung et al., 2021). Presentan cierta variación morfológica cuando esta es determinada a nivel celular y molecular. Se pueden encontrar bacilos redondos, ovalados e incluso formas ramificadas simples y contienen una pared celular significativamente más delgada que las cepas MDR (Velayati et al., 2013).

De acuerdo con la OMS, el tratamiento de la TB debe comenzar con la administración de los fármacos de primera línea (isoniazida, rifampicina, pirazinamida y etambutol) por un periodo de dos meses, seguido de 4 a 6 meses de rifampicina e isoniazida. Cuando se presenta resistencia a rifampicina se incorpora un régimen terapéutico que incluye bedaquilina, delamanida, linezolida y clofazimina por 9 a 12 semanas (Conradie et al., 2020). La TB MDR es

principalmente tratada con flouoroquinolonas (ciprofloxacina, ofloxacina, levofloxacina, y moxifloxacina). Este grupo de fármacos son de especial importancia debido a que los resultados de curación después del tratamiento son mejores en comparación con otros grupos de fármacos, aunque puede presentarse resistencia cruzada ya que comparten el mismo blanco genético, el gen *gyrA*.

Tuberculosis extrapulmonar

Existen dos tipos de manifestación clínica de la TB, la tuberculosis pulmonar (PTB) y la extrapulmonar (EPTB). La EPTB se desarrolla cuando la bacteria infecta otro órgano además de los pulmones, que puede incluir a la pleura, ganglios linfáticos, tracto genitourinario, piel, huesos, articulaciones, meninges, entre otros. Un paciente con PTB y EPTB es clasificado clínicamente como TB pulmonar. Por ejemplo, la TB miliar es considerada como PTB debido a que típicamente se presentan lesiones en los pulmones. Por otro lado, la linfadenopatía tuberculosa sin evidencia radiográfica de anormalidades en los pulmones, constituye un caso de EPTB (Singh et al., 2018).

Se considera que cerca del 20% de los casos de TB a nivel mundial corresponden a EPTB (OMS, 2021), aunque en pacientes VIH positivos el porcentaje se eleva hasta más del 50% de los casos. La PTB representa la mayoría de los casos y es la principal forma transmisible de la enfermedad; sin embargo, la EPTB también contribuye a la carga de la enfermedad y no recibe una atención específica en las estrategias internacionales de control. Además, se considera que, en contraste con los PTB, la tasa de casos de EPTB no ha disminuido, al contrario, se ha observado un incremento en la proporción de EPTB en países como Estados Unidos, Australia y Alemania (Ben Ayed et al., 2018).

Factores de riesgo para EPTB

La infección por VIH es el mayor factor de riesgo para el desarrollo de EPTB en personas con TB latente, debido a la presencia de una respuesta inmune deficiente. La asociación entre EPTB y VIH es muy probablemente debido a la deficiencia de

células T CD4+ y a la deficiencia en la producción de citocinas proinflamatorias en pacientes infectados con VIH. Diversos estudios han demostrado que otras comorbilidades, como la diabetes, también influyen en la presentación y progresión de la enfermedad hacia EPTB, como el realizado por Magee et al. (2016), en donde se demuestra que el riesgo de muerte por TB es mayor en pacientes diabéticos. Por otro lado, se ha reportado que tanto la presencia de una enfermedad hepática como un tratamiento antituberculoso previo incrementan el riesgo de enfermedad diseminada o EPTB. Adicionalmente, se sabe de múltiples genes del hospedero que están involucrados en la susceptibilidad a la diseminación de MTB. Por ejemplo, el polimorfismo T597C en el receptor TLR2 se ha asociado a TB meníngea y a un incremento en la severidad de los síntomas neurológicos. Otros polimorfismos en los genes que codifican para el receptor de interleucina 1 e interleucina 1 β se han asociado al desarrollo de TB pleural (Ben Ayed et al., 2018). Sin embargo, los factores del hospedero no explican completamente por qué la TB se manifiesta de forma PTB o EPTB, por lo que también se ha evaluado la posible asociación de cepas de genotipos particulares de MTB.

Las diferencias en la severidad de la infección entre genotipos particulares de MTB se han caracterizado en modelos de infección *in vitro* en macrófagos e *in vivo* en animales. Dichas diferencias se han definido por la velocidad de crecimiento bacteriano en células u órganos, la muerte de células o animales infectados, y las diferencias en la histopatología de tejidos en los animales infectados (Dormans et al., 2004). Se ha observado que las cepas más virulentas crecen más rápido, causan más daño pulmonar y mayor mortalidad, además de tener una mayor capacidad de transmisión que las cepas atenuadas o de baja virulencia. Como se mencionó anteriormente, se han reportado diferencias entre cepas clínicas con respecto a la respuesta inflamatoria que desencadenan, la gravedad de la enfermedad, la tasa de transmisión y la presentación de la enfermedad. Además, los resultados de estudios clínicos y epidemiológicos sugieren que las cepas pertenecientes al Linaje 5 y 6 son metabólicamente diferentes, crecen más lentamente y son menos virulentas que los otros linajes del CMTB adaptados a humanos (Diarra et al., 2018). De manera similar, se ha demostrado que al menos

ciertos grupos de cepas de linaje 2 y 4 son más virulentos y presentan una mayor capacidad de transmisión de persona a persona; mientras que las cepas pertenecientes al linaje Beijing tienen más probabilidad de progresar a enfermedad tuberculosa activa y se han asociado con EPTB, resistencia a múltiples fármacos, fracaso del tratamiento y recaída (Caws et al., 2008; Thwaites et al., 2008; Parwatti et al., 2010).

Todo lo anterior sugiere que, además de los factores ambientales y del hospedero, la diversidad genética de las cepas de MTB contribuye al resultado variable de la infección y a la presentación clínica de la TB. Sin embargo, los factores genéticos específicos de MTB que determinan la variación en la presentación clínica de la enfermedad siguen siendo en gran parte desconocidos, y las diferencias entre los linajes de MTB no explican el espectro completo de variación fenotípica observada. La secuenciación del genoma completo ofrece la posibilidad de estudiar la diversidad genómica de MTB con la mayor resolución posible, a nivel de cambios de nucleótidos individuales a lo largo de todo el genoma, incluyendo también las diferencias dentro de los linajes. En contraste con la gran cantidad de estudios realizados para identificar factores del hospedero que influyen en el desarrollo de EPTB, son muy pocos los que utilizan secuenciación del genoma completo para identificar determinantes genéticos presentes en las cepas bacterianas que puedan desencadenar en la aparición de EPTB.

Secuenciación genómica de MTB

Los enfoques de secuenciación del genoma completo utilizan plataformas de secuenciación de ADN para reconstruir la secuencia cromosómica completa de un organismo. El genoma relativamente pequeño de aproximadamente 4.4 megabases que poseen los miembros de MTB se acopla bien a estos enfoques de secuenciación. Esto, aunado al incremento en los últimos años en la rapidez, confiabilidad y bajo costo de las técnicas de secuenciación, han permitido la secuenciación de un gran número de cepas, mejorando el diagnóstico, tratamiento y vigilancia epidemiológica de la TB (Figura 2).

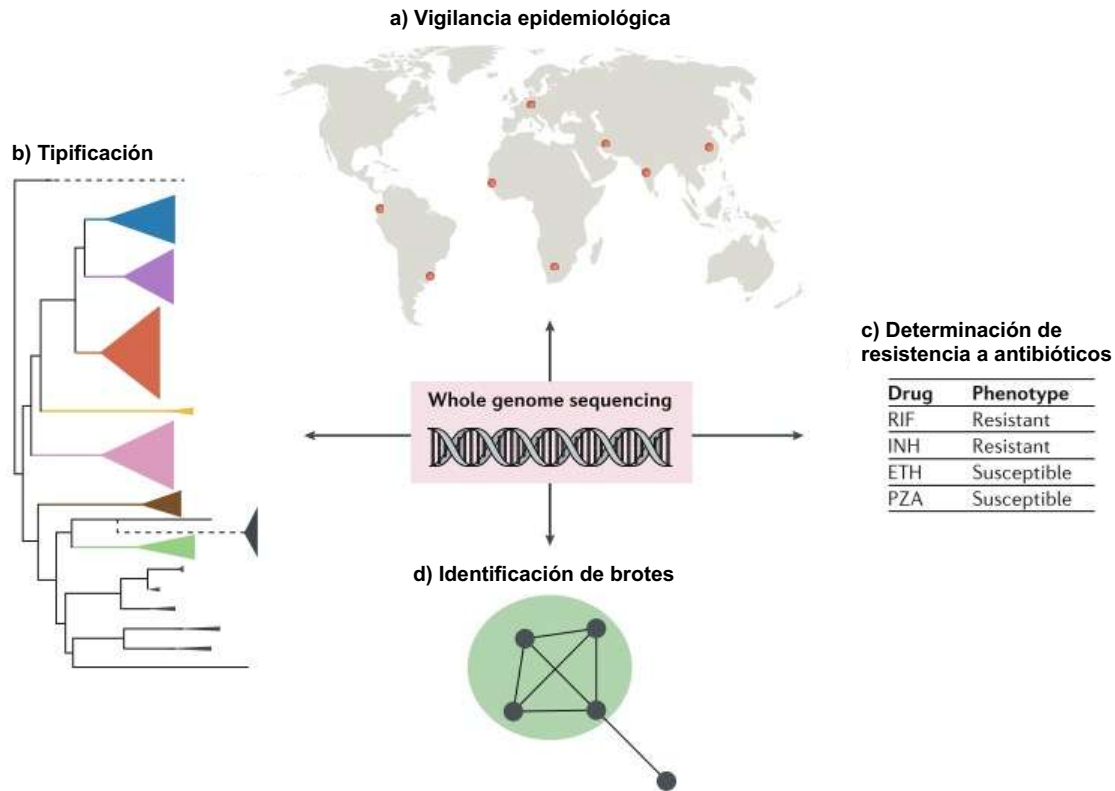


Figura 2. Secuenciación del genoma de MTB. Las principales aplicaciones de la secuenciación en la salud pública incluyen vigilancia epidemiológica (a), clasificación filogenética (b), determinación de patrones de resistencia a antibióticos basados en la presencia de SNPs (c), e identificación de clusters de transmisión y nuevos brotes (d). ETH, etambutol; INH, isoniazida; PZA, pirazinamida; RIF, rifampicina. Modificado de Meehan et al., 2019.

A partir de los datos de secuenciación es posible llevar a cabo una tipificación “*in silico*” de las cepas analizadas, predecir perfiles de resistencia a antibióticos y monitorear la aparición de droga resistencia. El flujo de trabajo estándar para el análisis genómico de cepas de MTB se muestra en la figura 3, este comienza con el aislamiento y cultivo de la cepa en medio sólido (Lowenstein-Jensen) o líquido (MGIT), seguido de la extracción de ADN genómico, preparación de la biblioteca y secuenciación mediante plataformas como Illumina. Después del trabajo experimental de laboratorio, se lleva a cabo el análisis bioinformático que incluye una gran cantidad de pasos clave, como el control de calidad de las lecturas de secuenciación y el posterior mapeo contra el genoma de la cepa de referencia (H37Rv) para la detección de variantes genómicas como SNPs, inserciones y deleciones (*indels*). Finalmente, con base en las variantes encontradas, se realiza

la asignación de linajes y sublinajes, así como la predicción de resistencia a antibióticos e identificación de clusters de transmisión.

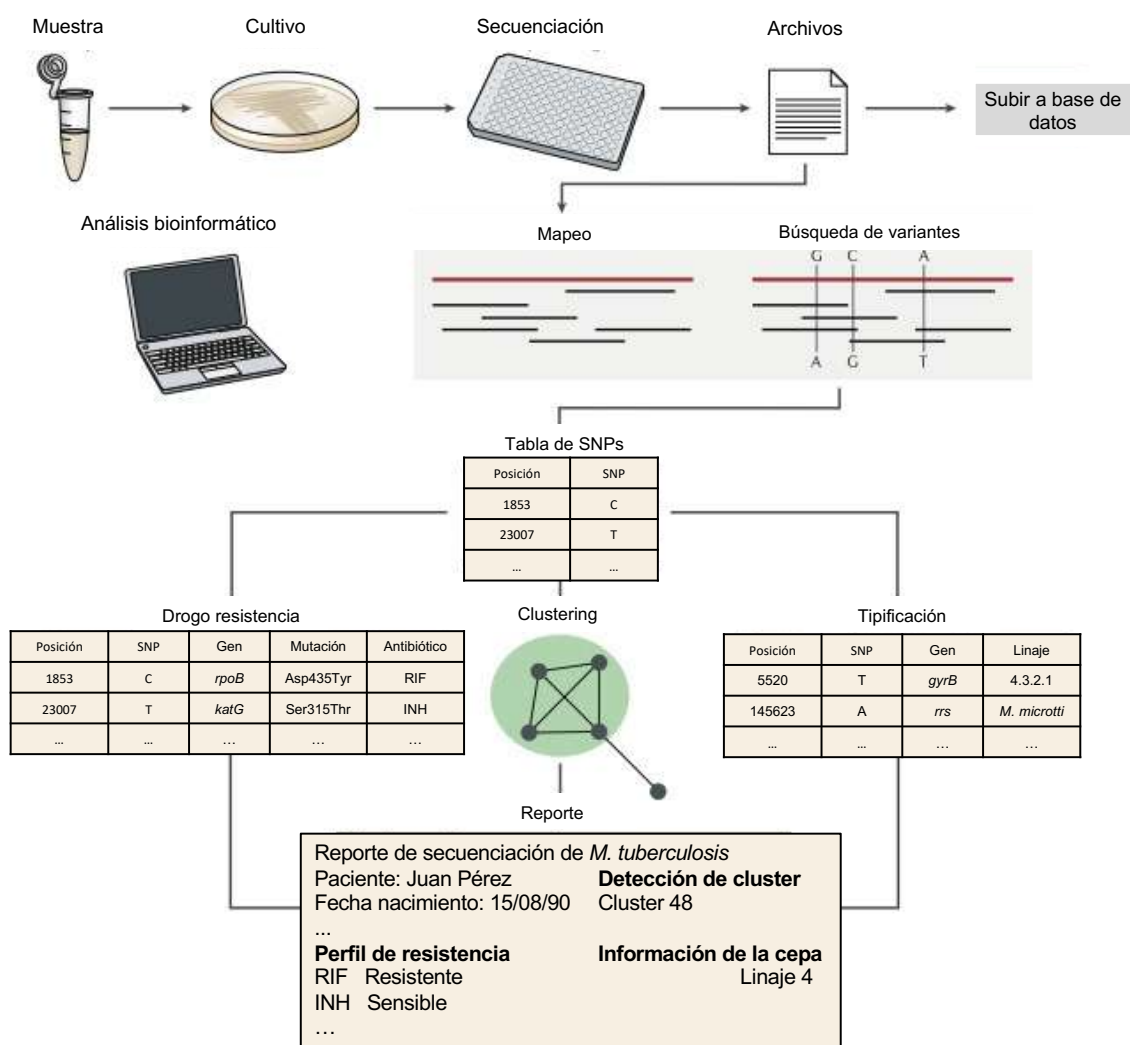


Figura 3. Flujo de trabajo estándar para el análisis genómico de cepas del complejo MTB. INH, isoniazida; RIF, rifampicina. Modificado de Meehan et al., 2019.

El estudio genómico de las cepas de MTB ha permitido llevar a cabo comparaciones entre cepas PTB y EPTB mediante la búsqueda de variaciones en el genoma utilizando diferentes estrategias. Así, en 2013 Das y colaboradores secuenciaron 5 cepas causantes de EPTB y realizaron el análisis para la búsqueda de variantes compartidas por la totalidad de las cepas mediante el mapeo contra el genoma de referencia H37Rv; sin embargo, no lograron encontrar dichas variaciones (Das et al., 2013).

En 2016 en Malasia, se secuenciaron 8 cepas causantes de TB meníngea, con la finalidad de identificar determinantes genéticos que pudieran asociarse con neurotropismo mediante la comparación con cepas causantes de TB pulmonar. Las comparaciones genómicas revelaron reordenamientos (traslocaciones, inversiones, inserciones y eliminaciones) y SNPs no sinónimos en las cepas derivadas de líquido cefalorraquídeo que no se observaron en los genomas de origen respiratorio de MTB. Estos reordenamientos poseen una gran cantidad de genes PE (prolina-glutamato) /PPE (prolina-prolina-glutamato), proteínas transcripcionales y de membrana. De manera similar, la mayoría de los SNPs comunes en las cepas meníngeas se observaron en genes que codifican proteínas PE y PPE. Aunque se reportaron variaciones exclusivas de las cepas EPTB en este trabajo, no se lograron identificar determinantes genéticos o variaciones comunes a todas las cepas aisladas de pacientes con EPTB. Posteriormente, en 2017 se seleccionaron 5 cepas correspondientes al Linaje 3 que fueron aisladas de pacientes con EPTB en la India (Sharma et al., 2017). En estas cepas se realizó un análisis similar a los dos anteriores, utilizando para la comparación cepas PTB depositadas en bases de datos. Los resultados mostraron similitudes con los trabajos previamente citados, reportándose 279 variaciones genómicas exclusivas de las cepas EPTB pero que no se comparten entre ellas.

Finalmente, en 2018 se llevaron a cabo dos estudios en los que se utilizó una mayor cantidad de cepas EPTB que se compararon con cepas PTB aisladas de la misma región geográfica. El primero de ellos utilizó 322 cepas, de las cuales 106 fueron EPTB y 206 PTB (Ruesen et al., 2018), mientras que en el segundo se analizó un total de 293, incluyendo 73 EPTB y 220 PTB (Faksri et al., 2018). Estos trabajos utilizaron distintas estrategias metodológicas para obtener determinantes genéticos asociados a TB meníngea. Sin embargo, no se encontraron variantes exclusivas de las cepas EPTB compartidas entre la mayoría de las cepas que ocasionan este fenotipo clínico. Dichos resultados pueden atribuirse a varios factores como: a) el bajo número de cepas de MTB causantes de EPTB, b) el hecho de que las cepas analizadas pertenecen a una sola región geográfica, por lo que existe un sesgo asociado a la predominancia de un linaje específico, o c) la

estrategia de análisis genómico utilizada. Es por esto que a la fecha no existe un consenso sobre los determinantes genéticos que contribuyen al fenotipo EPTB, por lo que el análisis genómico comparativo derivado de la construcción del pangenoma puede ser una estrategia novedosa para encontrar diferencias que contribuyan a explicar la presentación clínica de la enfermedad.

REFERENCIAS

- Ben, Ayed., H., Koubaa, M., Marrakchi, C., Rekik, K., Hammami, F., & Smaoui, F. 2018. Extrapulmonary tuberculosis: update on the epidemiology, risk factors and prevention strategies. *Int J Trop Dis*. 1(006):1-6
- Caws, M., Thwaites, G., Dunstan, S., Hawn, T.R., Lan, N.T., Thuong, N.T., et al. 2008. The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis*. *PLoS Pathog*. 4(3): e1000034. <https://doi.org/10.1371/journal.ppat.1000034>
- Chiner-Oms, Á., Sánchez-Busó, L., Corander, J., Gagneux, S., Harris, S. R., Young, D., González-Candelas, F., & Comas, I. 2019. Genomic determinants of speciation and spread of the *Mycobacterium tuberculosis* complex. *Science Advances*. 5(6): eaaw3307. <https://doi.org/10.1126/sciadv.aaw3307>
- Conradie, F., Diacon, A. H., Ngubane, N., Howell, P., Everitt, D., Crook, A. M., Mendel, C. M., Egizi, E., Moreira, J., Timm, J., McHugh, T. D., Wills, G. H., Bateson, A., Hunt, R., Van Niekerk, C., Li, M., Olugbosi, M., Spigelman, M., & Nix-TB Trial Team. 2020. Treatment of highly drug-resistant pulmonary tuberculosis. *The New England Journal of Medicine*. 382(10): 893–902. <https://doi.org/10.1056/NEJMoa1901814>
- Cooper, A.M. 2009. T cells in mycobacterial infection and disease. *Current Opinion in Immunology*. 21(4): 378–384. <https://doi.org/10.1016/j.coi.2009.06.004>
- Correa-Macedo, W., Cambri, G., & Schurr, E. 2019. The interplay of human and *Mycobacterium tuberculosis* genomic variability. *Frontiers in Genetics*. 10: 865. <https://doi.org/10.3389/fgene.2019.00865>
- 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>
- Coscolla, M., Gagneux, S., Menardo, F., Loiseau, C., Ruiz-Rodriguez, P., Borrell, S., Otchere, I. D., Asante-Poku, A., Asare, P., Sánchez-Busó, L., Gehre, F., Sanoussi, C. N., Antonio, M., Affolabi, D., Fyfe, J., Beckert, P., Niemann, S., Alabi, A. S., Grobusch, M. P., Kobbe, R., ... Brites, D. 2021. Phylogenomics

- of *Mycobacterium africanum* reveals a new lineage and a complex evolutionary history. *Microbial Genomics*. 7(2): 000477. <https://doi.org/10.1099/mgen.0.000477>
- Das, S., Roychowdhury, T., Kumar, P., Kumar, A., Kalra, P., Singh, J., Singh, S., Prasad, H. K., & Bhattacharya, A. 2013. Genetic heterogeneity revealed by sequence analysis of *Mycobacterium tuberculosis* isolates from extra-pulmonary tuberculosis patients. *BMC Genomics*. 14: 404. <https://doi.org/10.1186/1471-2164-14-404>
- de Martino, M., Lodi, L., Galli, L., & Chiappini, E. 2019. Immune response to *Mycobacterium tuberculosis*: a narrative review. *Frontiers in Pediatrics*. 7: 350. <https://doi.org/10.3389/fped.2019.00350>
- Diarra, B., Kone, M., Togo, A., Sarro, Y., Cisse, A. B., Somboro, A., Degoga, B., Tolofoudie, M., Kone, B., Sanogo, M., Baya, B., Kodio, O., Maiga, M., Belson, M., Orsega, S., Krit, M., Dao, S., Maiga, I. I., Murphy, R. L., Rigouts, L., et al. 2018. *Mycobacterium africanum* (Lineage 6) shows slower sputum smear conversion on tuberculosis treatment than *Mycobacterium tuberculosis* (Lineage 4) in Bamako, Mali. *PloS one*. 13(12): e0208603. <https://doi.org/10.1371/journal.pone.0208603>
- Dormans, J., Burger, M., Aguilar, D., Hernandez-Pando, R., Kremer, K., Roholl, P., Arend, S. M., & van Soolingen, D. 2004. Correlation of virulence, lung pathology, bacterial load and delayed type hypersensitivity responses after infection with different *Mycobacterium tuberculosis* genotypes in a BALB/c mouse model. *Clinical and experimental immunology*, 137(3): 460–468. <https://doi.org/10.1111/j.1365-2249.2004.02551.x>
- Faksri, K., Xia, E., Ong, R. T., 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>

- Firdessa, R., Berg, S., Hailu, E., Schelling, E., Gumi, B., Erenso, G., Gadisa, E., Kiros, T., Habtamu, M., Hussein, J., Zinsstag, J., Robertson, B. D., Ameni, G., Lohan, A. J., Loftus, B., Comas, I., Gagneux, S., Tschopp, R., Yamuah, L., Hewinson, G., ... Aseffa, A. 2013. Mycobacterial lineages causing pulmonary and extrapulmonary tuberculosis, Ethiopia. *Emerging infectious diseases*, 19(3): 460–463. <https://doi.org/10.3201/eid1903.120256>
- Gagneux, S. 2013. Genetic diversity in *Mycobacterium tuberculosis*. *Current topics in microbiology and immunology*, 374: 1–25. https://doi.org/10.1007/82_2013_329
- Khawbung, J. L., Nath, D., & Chakraborty, S. 2021. Drug resistant tuberculosis: a review. *Comparative Immunology, Microbiology and Infectious Diseases*, 74: 101574. <https://doi.org/10.1016/j.cimid.2020.101574>
- Kong, Y., Cave, M. D., Yang, D., Zhang, L., Marrs, C. F., Foxman, B., Bates, J. H., Wilson, F., Mukasa, L. N., & Yang, Z. H. 2005. Distribution of insertion and deletion-associated genetic polymorphisms among four *Mycobacterium tuberculosis* phospholipase C genes and associations with extrathoracic tuberculosis: a population-based study. *Journal of clinical microbiology*, 43(12): 6048–6053. <https://doi.org/10.1128/JCM.43.12.6048-6053.2005>
- Liu, H., Jiang, Y., Dou, X., Wang, H., Zhao, X., Zhang, W., Wan, L., Zhang, Z., Chen, C., & Wan, K. 2013. pstS1 polymorphisms of *Mycobacterium tuberculosis* strains may reflect ongoing immune evasion. *Tuberculosis*, 93(5): 475–481. <https://doi.org/10.1016/j.tube.2013.05.003>
- Madacki, J., Mas Fiol, G., & Brosch, R. 2019. Update on the virulence factors of the obligate pathogen *Mycobacterium tuberculosis* and related tuberculosis-causing mycobacteria. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 72: 67–77. <https://doi.org/10.1016/j.meegid.2018.12.013>
- Magee, M. J., Foote, M., Ray, S. M., Gandhi, N. R., & Kempker, R. R. 2016. Diabetes mellitus and extrapulmonary tuberculosis: site distribution and risk of mortality. *Epidemiology and Infection*, 144(10): 2209–2216. <https://doi.org/10.1017/S0950268816000364>
- Malik, A. N., & Godfrey-

- Faussett, P. 2005. Effects of genetic variability of *Mycobacterium tuberculosis* strains on the presentation of disease. *The Lancet infectious Diseases*, 5(3): 174-183.
- Manca, C., Tsenova, L., Freeman, S., Barczak, A. K., Tovey, M., Murray, P. J., Barry, C., & Kaplan, G. 2005. Hypervirulent *M. tuberculosis* W/Beijing strains upregulate type I IFNs and increase expression of negative regulators of the Jak-Stat pathway. *Journal of interferon & cytokine research*, 25(11): 694–701. <https://doi.org/10.1089/jir.2005.25.694>
- Marimani, M., Ahmad, A., & Duse, A. 2018. The role of epigenetics, bacterial and host factors in progression of *Mycobacterium tuberculosis* infection. *Tuberculosis*, 113: 200–214. <https://doi.org/10.1016/j.tube.2018.10.009>
- Maulén, N. P., & Cifuentes, L. 2018. Polimorfismos genéticos asociados a la inmunidad innata y la susceptibilidad genética a la tuberculosis. *Revista chilena de enfermedades respiratorias*, 34(4) 226-235.
- Müller, A. M., Osório, C. S., Silva, D. R., Sbruzzi, G., de Tarso, P., & Dalcin, R. 2018. Interventions to improve adherence to tuberculosis treatment: systematic review and meta-analysis. *The international journal of tuberculosis and lung disease*, 22(7): 731–740. <https://doi.org/10.5588/ijtld.17.0596>
- Narendran, G., Kavitha, D., Karunaianantham, R., Gil-Santana, L., Almeida-Junior, J. L., Reddy, S. D., Kumar, M. M., Hemalatha, H., Jayanthi, N. N., Ravichandran, N., Krishnaraja, R., Prabhakar, A., Manoharan, T., Nithyananthan, L., Arjunan, G., Natrajan, M., Swaminathan, S., & Andrade, B. B. 2016. Role of LTA4H polymorphism in tuberculosis-associated immune reconstitution inflammatory syndrome occurrence and clinical severity in patients infected with HIV. *PLoS one*, 11(9): e0163298. <https://doi.org/10.1371/journal.pone.0163298>
- Ngabonziza, J., Loiseau, C., Marceau, M., Jouet, A., Menardo, F., Tzfidia, O., Antoine, R., Niyigena, E. B., Mulders, W., Fissette, K., Diels, M., Gaudin, C., Duthoy, S., Ssenkooba, W., André, E., Kaswa, M. K., Habimana, Y. M., Brites, D., Affolabi, D., Mazarati, J. B., ... Supply, P. 2020. A sister lineage of the *Mycobacterium tuberculosis* complex discovered in the African Great Lakes

- region. *Nature communications*, 11(1): 2917. <https://doi.org/10.1038/s41467-020-16626-6>
- Pareek, M., Evans, J., Innes, J., Smith, G., Hingley-Wilson, S., Loughheed, K. E., Sridhar, S., Dediccoat, M., Hawkey, P., & Lalvani, A. 2013. Ethnicity and mycobacterial lineage as determinants of tuberculosis disease phenotype. *Thorax*, 68(3): 221–229. <https://doi.org/10.1136/thoraxjnl-2012-201824>
- Parwati, I., Alisjahbana, B., Apriani, L., Soetikno, R. D., Ottenhoff, T. H., van der Zanden, A. G., van der Meer, J., van Soolingen, D., & van Crevel, R. 2010. *Mycobacterium tuberculosis* Beijing genotype is an independent risk factor for tuberculosis treatment failure in Indonesia. *The Journal of infectious diseases*, 201(4): 553–557. <https://doi.org/10.1086/650311>
- Reed, M. B., Domenech, P., Manca, C., Su, H., Barczak, A. K., Kreiswirth, B. N., Kaplan, G., & Barry, C. E. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature*, 431(7004): 84–87. <https://doi.org/10.1038/nature02837>
- Rengarajan, J., Bloom, B. R., & Rubin, E. J. 2005. Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proceedings of the National Academy of Sciences of the United States of America*, 102(23): 8327–8332. <https://doi.org/10.1073/pnas.0503272102>
- Rosain, J., Kong, X. F., Martinez-Barricarte, R., Oleaga-Quintas, C., Ramirez-Alejo, N., Markle, J., Okada, S., Boisson-Dupuis, S., Casanova, J. L., & Bustamante, J. 2019. Mendelian susceptibility to mycobacterial disease: 2014-2018 update. *Immunology and cell biology*, 97(4): 360–367. <https://doi.org/10.1111/imcb.12210>
- Ruesen, C., Chaidir, L., van Laarhoven, A., Dian, S., Ganiem, A. R., Nebenzahl-Guimaraes, H., Huynen, M. A., Alisjahbana, B., Dutilh, B. E., & van Crevel, R. 2018. Large-scale genomic analysis shows association between homoplastic genetic variation in *Mycobacterium tuberculosis* genes and meningeal or pulmonary tuberculosis. *BMC genomics*, 19(1): 122. <https://doi.org/10.1186/s12864-018-4498-z>

- Sasseti, C. M., Boyd, D. H., & Rubin, E. J. 2003. Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular microbiology*, 48(1): 77–84. <https://doi.org/10.1046/j.1365-2958.2003.03425.x>
- Saw, S. H., Tan, J. L., Chan, X. Y., Chan, K. G., & Ngeow, Y. F. 2016. Chromosomal rearrangements and protein globularity changes in *Mycobacterium tuberculosis* isolates from cerebrospinal fluid. *PeerJ*, 4: e2484. <https://doi.org/10.7717/peerj.2484>
- Sharma, K., Verma, R., Advani, J., Chatterjee, O., Solanki, H. S., Sharma, A., Varma, S., Modi, M., Ray, P., Mukherjee, K. K., Sharma, M., Dhillon, M. S., Suar, M., Chatterjee, A., Pandey, A., Prasad, T., & Gowda, H. 2017. Whole genome sequencing of *Mycobacterium tuberculosis* Isolates from extrapulmonary sites. *Omics*, 21(7): 413–425. <https://doi.org/10.1089/omi.2017.0070>
- Singh, J., & Dinkar, A. 2018. Positive outcome of pulmonary tuberculosis associated with extraordinary extensive extrapulmonary tuberculosis in an immunocompetent adult. *International journal of mycobacteriology*, 7(2): 183–185. https://doi.org/10.4103/ijmy.ijmy_61_18
- Thwaites, G., Caws, M., Chau, T. T., D'Sa, A., Lan, N. T., Huyen, M. N., Gagneux, S., Anh, P. T., Tho, D. Q., Torok, E., Nhu, N. T., Duyen, N. T., Duy, P. M., Richenberg, J., Simmons, C., Hien, T. T., & Farrar, J. 2008. Relationship between *Mycobacterium tuberculosis* genotype and the clinical phenotype of pulmonary and meningeal tuberculosis. *Journal of clinical microbiology*, 46(4): 1363–1368. <https://doi.org/10.1128/JCM.02180-07>
- Tundup, S., Mohareer, K., & Hasnain, S. E. 2014. *Mycobacterium tuberculosis* PE25/PPE41 protein complex induces necrosis in macrophages: Role in virulence and disease reactivation?. *FEBS open bio*, 4: 822–828. <https://doi.org/10.1016/j.fob.2014.09.001>
- Velayati, A. A., Farnia, P., & Masjedi, M. R. 2013. The totally drug resistant tuberculosis (TDR-TB). *International journal of clinical and experimental medicine*, 6(4): 307–309.

- Yáñez-Lema, G., Toapanta-Guayta, B., Ramírez-Jaramillo, E., & Zambrano-Leal, M. 2021. Tuberculosis meníngea en un paciente pediátrico: reporte de un caso clínico. *Revista Ecuatoriana de Neurología*, 30(1): 218-222. <https://doi.org/10.46997/revecuatneurol30100218>
- Yang, Z., Yang, D., Kong, Y., Zhang, L., Marrs, C. F., Foxman, B., Bates, J. H., Wilson, F., & Cave, M. D. (2005). Clinical relevance of *Mycobacterium tuberculosis* plcD gene mutations. *American journal of respiratory and critical care medicine*, 171(12): 1436–1442. <https://doi.org/10.1164/rccm.200408-1147OC>
- Zhang, S., Chen, J., Shi, W., Liu, W., Zhang, W., & Zhang, Y. 2013. Mutations in panD encoding aspartate decarboxylase are associated with pyrazinamide resistance in *Mycobacterium tuberculosis*. *Emerging microbes & infections*, 2(6): e34. <https://doi.org/10.1038/emi.2013.38>

JUSTIFICACIÓN

La tuberculosis causada por MTB es la novena causa mundial de muerte y la primera por enfermedades infecciosas bacterianas. Esta enfermedad afecta a grandes grupos de población, particularmente de áreas marginadas y grupos vulnerables donde predomina la pobreza, la desnutrición y el hacinamiento. Se calcula que una tercera parte de la población mundial tiene TB latente, término aplicado a las personas infectadas por el bacilo pero que aún no han enfermado ni pueden transmitir la infección. Aunque la TB es fundamentalmente una enfermedad pulmonar, otros órganos y tejidos también pueden afectarse, incluyendo a los sistemas nervioso central, linfático, circulatorio y genitourinario, así como huesos y articulaciones. Las formas extrapulmonares de la TB podrían ser la expresión clínica de un defecto inmune del hospedero y pueden conllevar morbilidad y riesgo de mortalidad elevado, sobre todo en las formas meníngeas y miliares; además, presentan dificultades de diagnóstico y la duración mínima eficaz de su tratamiento no está bien determinada. Sin embargo, se sabe que el fondo genético de la bacteria también juega un papel importante en el establecimiento de este fenotipo clínico. A pesar de esto, a la fecha no existe un consenso sobre qué determinantes genéticos contribuyen a las formas EPTB.

En la actualidad se ha acumulado una cantidad considerable de secuencias genómicas completas disponibles de MTB, lo que ofrece nuevas oportunidades para estudios de genómica comparativa. Con base en esta información y mediante la utilización de herramientas bioinformáticas avanzadas, la presente propuesta implementa una estrategia novedosa que permite analizar genomas de MTB de una manera global, accediendo a una cantidad de información previamente no explorada para la comparación de cepas pulmonares y extrapulmonares de esta bacteria. Este trabajo, con aplicación directa a un problema importante en salud pública como es la TB, fortalece también el campo de la bioinformática en el país.

HIPÓTESIS

Existe un fondo genómico exclusivo de las cepas de MTB causantes de EPTB que determina su capacidad para causar este fenotipo clínico.

OBJETIVOS

OBJETIVO GENERAL

Identificar y caracterizar regiones genómicas de MTB que contribuyan a explicar las diferencias fenotípicas entre las cepas causantes de PTB y EPTB.

OBJETIVOS ESPECÍFICOS

- Evaluar la asociación entre el linaje/sublinaje de las cepas de MTB con la presentación clínica de la TB.
- Evaluar la asociación entre las mutaciones presentes en genes de resistencia a antibióticos de las cepas de MTB y la presentación clínica de la TB.
- Identificar variaciones que contribuyan a explicar las diferencias fenotípicas entre cepas PTB y EPTB.
- Identificar genes exclusivos de cepas de MTB que estén relacionados con la presentación clínica de la enfermedad.

RESULTADOS

CAPÍTULO 1: Whole-genome comparative analysis at the lineage/sublineage level discloses relationships between *Mycobacterium tuberculosis* genotype and clinical phenotype.



Whole-genome comparative analysis at the lineage/sublineage level discloses relationships between *Mycobacterium tuberculosis* genotype and clinical phenotype

Submitted 10 May 2021
Accepted 17 August 2021
Published 8 September 2021

Corresponding author

Ma. Soledad Vázquez-Garcidueñas,
soledad.vazquez@umich.mx

Academic editor

Mario Alberto Flores-Valdez

Additional Information and
Declarations can be found on
page 12

DOI 10.7717/peerj.12128

© Copyright
2021 Negrete-Paz et al.

Distributed under
Creative Commons CC-BY 4.0
OPEN ACCESS

How to cite this article Negrete-Paz AM, Vázquez-Marrufo G, Vázquez-Garcidueñas MS. 2021. Whole-genome comparative analysis at the lineage/sublineage level discloses relationships between *Mycobacterium tuberculosis* genotype and clinical phenotype. *PeerJ* 9:e12128 <http://doi.org/10.7717/peerj.12128>

Andrea Monserrat Negrete-Paz^{1,*}, Gerardo Vázquez-Marrufo^{1,*} and Ma. Soledad Vázquez-Garcidueñas²

¹ Centro Multidisciplinario de Estudios en Biotecnología, Facultad de Medicina Veterinaria y Zootecnia, Universidad Michoacana de San Nicolás de Hidalgo, Tarímbaro, Michoacán, Mexico

² División de Estudios de Posgrado, Facultad de Ciencias Médicas y Biológicas “Dr. Ignacio Chávez”, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, Mexico

* These authors contributed equally to this work.

ABSTRACT

Background. Human tuberculosis (TB) caused by members of the *Mycobacterium tuberculosis* complex (MTBC) is the main cause of death among infectious diseases worldwide. Pulmonary TB (PTB) is the most common clinical phenotype of the disease, but some patients develop an extrapulmonary (EPTB) phenotype in which any organ or tissue can be affected. MTBC species include nine phylogenetic lineages, with some appearing globally and others being geographically restricted. EPTB can or not have pulmonary involvement, challenging its diagnosis when lungs are not implicated, thus causing an inadequate treatment. Finding evidence of a specific *M. tuberculosis* genetic background associated with EPTB is epidemiologically relevant due to the virulent and multidrug-resistant strains isolated from such cases. Until now, the studies conducted to establish associations between *M. tuberculosis* lineages and PTB/EPTB phenotypes have shown inconsistent results, which are attributed to the strain predominance from specific *M. tuberculosis* lineages/sublineages in the samples analyzed and the use of low-resolution phylogenetic tools that have impaired sublineage discrimination abilities. The present work elucidates the relationships between the MTBC strain lineages/sublineages and the clinical phenotypes of the disease as well as the antibiotic resistance of the strains.

Methods. To avoid biases, we retrieved the raw genomic reads (RGRs) of all ($n = 245$) the *M. tuberculosis* strains worldwide causing EPTB available in databases and an equally representative sample of the RGRs ($n = 245$) of PTB strains. A multiple alignment was constructed, and a robust maximum likelihood phylogeny based on single-nucleotide polymorphisms was generated, allowing effective strain lineage/sublineage assignment.

Results. A significant Odds Ratio (OR range: 1.8–8.1) association was found between EPTB and the 1.1.1, 1.2.1, 4.1.2.1 and ancestral Beijing sublineages. Additionally, a significant association between PTB with 4.3.1, 4.3.3, and 4.5 and Asian African 2 and Europe/Russia B0/W148 modern Beijing sublineages was found. We also observed a significant association of Lineage 3 strains with multidrug resistance (OR 3.8; 95% CI [1.1–13.6]), as well as between modern Beijing sublineages and antibiotic resistance (OR

4.3; 3.8–8.6). In this work, it was found that intralinesage diversity can drive differences in the immune response that triggers the PTB/EPTB phenotype.

Subjects Bioinformatics, Genomics, Microbiology

Keywords Pulmonary tuberculosis, Extrapulmonary tuberculosis, Drug resistance, Ancient sublineages, Modern sublineages, SNP

INTRODUCTION

Tuberculosis (TB) represents the main cause of death among infectious diseases worldwide, with its drug-resistant manifestations constituting a major global health concern (Floyd *et al.*, 2018). Pulmonary TB (PTB) is the most common clinical disease phenotype, but some patients develop an extrapulmonary TB (EPTB) phenotype in which practically any organ or tissue can be affected, including the aggressive manifestations of lymph node and central nervous system TB (Golden & Vikram, 2005). EPTB represents approximately 15% of reported TB cases globally, whereas as many as 40% of TB cases in several high-income countries are EPTB (WHO, 2020).

Human TB is caused by members of the *Mycobacterium tuberculosis* complex (MTBC), which have >99% nucleotide sequence identity at the genomic level (Gagneux, 2018). The human-adapted species of the MTBC are *M. tuberculosis sensu stricto* and *Mycobacterium africanum*, which are divided into nine phylogenetic lineages: L1, or Indo-Oceanic; L2, or East Asian; L3, or East African-Indian; L4, or Euro-American; L5, or *M. africanum* West-African 1; L6, or *M. africanum* West-African 2; L7, or Ethiopia (Firdessa *et al.*, 2013; Coscolla & Gagneux, 2014); L8, or *M. tuberculosis* from the African Great Lakes (Ngabonziza *et al.*, 2020); and the recently described *M. africanum* L9 (Coscolla *et al.*, 2021). Global phylogeographic reconstruction of *M. tuberculosis* suggests that each lineage has become specifically adapted to defined human populations (Gagneux, 2018), with some occurring globally and others showing strong geographical restriction. L4, L2, and L3 are the most commonly found lineages; nevertheless, L4 is the most widespread lineage worldwide. L3 is mostly found in the Middle East, India, and East Africa, while L2 is found predominantly in East Asia (McHenry *et al.*, 2020). This geographical restriction has been found even at the sublineage level, as in sublineage 4.6/Uganda, found only in Uganda and neighboring countries. This sublineage has been shown to possess highly conserved T cell epitopes and a restricted geographic distribution, suggesting a possible adaptation to a specific human population (Stucki *et al.*, 2016). On the other hand, the high incidence of infections associated with nongeographically restricted strains might imply that these strains are more effective in causing the disease (Malik & Godfrey-Faussett, 2005).

Several studies have been conducted to identify possible relationships between *M. tuberculosis* phylogenetic lineages and the PTB or EPTB phenotype of the disease (Feng *et al.*, 2008; Click *et al.*, 2012), but the results show a lack of consistency. Different factors contribute to explaining the observed discrepancies among the conducted studies in an attempt to establish genotype-phenotype relationships. In the first instance, there were differences in the sample size (Coscolla & Gagneux, 2014), as well as in the nonhomogeneous

distribution of lineages, among the set (*Kato-Maeda & Nahid, 2012*) of analyzed strains. Moreover, biased associations might arise due to a lack of data or failure to control for possible confounders associated with known risk factors for EPTB, such as human immunodeficiency virus (HIV) infection comorbidity in patients from whom strains were isolated (*Coscolla & Gagneux, 2014*). Furthermore, studies use different operational definitions for EPTB (*Kato-Maeda & Nahid, 2012*), and some lack appropriate tools to index genomic diversity and classify strains into lineages in some studies (*Coscolla & Gagneux, 2010*). Whole-genome comparative analysis has allowed the categorization of *M. tuberculosis* lineages into sublineages using single-nucleotide polymorphism (SNP) analysis (*Coll et al., 2014; Stucki et al., 2016*). This subtle level of strain differentiation suggests that some sublineages might drive the observed associations of an *M. tuberculosis* genotype with a specific disease phenotype, such as EPTB (*Feng et al., 2008*), but an analysis to provide evidence in support of this hypothesis has not been conducted. Interestingly, the frequency of the sublineages assigned to isolated strains from the East Asian (L2) lineage differs among populations settled in different geographical areas, which might also explain why some studies associate L2 lineage strains with EPTB (*Feng et al., 2008*), whereas others associate it with PTB (*Dale et al., 2005*), and still others do not find any association at all (*Svensson et al., 2011*). Thus, intralocus diversity requires a detailed exploration to clarify disease phenotype variation and its relationship with MTBC genotypes (*Kato-Maeda & Nahid, 2012*). Additionally, genomic evidence has revealed a close relationship among specific *M. tuberculosis* lineages and sublineages with drug resistance (*Wang et al., 2015*), a relevant public health phenotype that might hinder successful TB treatment. Horizontal transfer of drug resistance genes has not been reported for *M. tuberculosis*, but resistance mostly arises from chromosomal mutations under the selective pressure of antibiotic use (*Nguyen, 2016*). Thus, the characterization of mutations associated with resistance phenotypes in strains with different genotypes can help to reveal lineage-/sublineage-specific microevolutionary processes of epidemiological relevance. Drug-resistance-associated mutations have been hypothesized to modify strain fitness and the ability of a strain to cross the blood-brain barrier, causing EPTB, specifically tuberculous meningitis (*Faksri et al., 2018*).

Unclear relationships of lineages/sublineages with *M. tuberculosis* PTB/EPTB disease and drug resistance phenotypes hinder the generation of adequate epidemiological transmission chains and timely successful treatments. Therefore, this work aims to elucidate relationships between strain genotypes at the lineage/sublineage level and clinical disease phenotypes as well as antibiotic resistance. We retrieved the raw read datasets from the genomes of all the *M. tuberculosis* strains causing EPTB worldwide and deposited in databases. The raw datasets of the same number of genomes of strains causing PTB isolated from the same countries of origin as EPTB strains were selected to avoid sources of possible biases related to previous studies originating from (i) a low number of analyzed strains, (ii) unequal PTB/EPTB strains analyzed, or (iii) differences in regional clinical TB phenotype incidences. The epidemiological and public health relevance of the identified relationships is discussed.

MATERIALS & METHODS

Data retrieval

A total of 490 raw datasets of genome sequence reads were retrieved, which corresponded to 245 *M. tuberculosis* strains causing PTB and 245 strains causing EPTB (Table S1). We retrieved all available genomes of EPTB strains deposited in the NCBI-SRA database. Such genomes correspond to clinical cases for which the major site of infection reported in databases was not pulmonary or miliary and an additional infection site either was not specified or was specified but not as pulmonary or miliary (Click *et al.*, 2012). The first criterion for the genome selection of the PTB strains was the country of the isolation of the EPTB strains for which the genomes were available. The objective of such a criterion was to analyze the genomes of *M. tuberculosis* phenotypes from similar clinical and human population backgrounds. The second criterion to select PTB strains was a negative HIV status reported in the associated metadata. In fact, this second criterion was applied for both PTB and EPTB genomes. Additionally, due to the large number of available genomes of PTB strains from different countries, it was possible to discard other comorbidities. All these criteria were used to avoid possible biases generated by population-genotype associations and HIV, or other comorbidities.

The sequence quality of the FASTQ reads was checked using FASTQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) and subsequently filtered to a Phred score of 30 using TrimGalore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/).

In silico typing

In silico spoligotyping was performed using SpoTyping program version 2.0 (Xia, Teo & Ong, 2016) for next-generation sequencing reads with default parameters. To determine lineage, the TB INSIGHT (<http://tbinsight.cs.rpi.edu>) server was used based on the identified spoligotypes.

Phylogenetic reconstruction

To find SNPs for phylogenetic reconstruction (Homolka *et al.*, 2012; Coll *et al.*, 2014; Merkeret *et al.*, 2015), the sequencing reads of the studied strains were aligned to the reference strain of *M. tuberculosis* H37Rv (accession no. NC_000962.3) using the MTBseq program version 1.0.3 (Kohl *et al.*, 2018) with default values. A frequency of allelic variation greater or equal to 75% and with a phred value >20 was used. Strains where the percentage of reads mapped against the reference genome was less than 80 were excluded due to possible contamination. Also, strains with a median depth coverage >30x were removed. The minimum coverage depth to support a SNP was 8x. For lineage and sublineage assignment, variant positions belonging to repeated regions and resistance genes were excluded. Of the remaining variant positions, those where data quality is below thresholds in >5% of samples were discarded (Jajou *et al.*, 2019). The sublineage phylogeny was rooted using a *Mycobacterium microti* strain (SRR2667442). Phylogenetic inferences were conducted with the maximum likelihood (ML) criterion using the IQ-TREE package (Nguyen *et al.*, 2015) with a general time-reversible (GTR) model of nucleotide substitution and a gamma

model of rate heterogeneity. Phylogenetic trees were constructed based on 1,000 bootstrap replicates, and their visualization was performed using iTOL (Letunic & Bork, 2016).

***In silico* determination of drug resistance**

Raw FASTQ sequencing files were uploaded to TB-Profiler version 3.0.4 (Coll et al., 2015), a tool to determine *in silico* drug resistance. TB-Profiler can determine genotypic drug resistance by aligning raw sequences against the reference genome *M. tuberculosis* H37Rv in order to identify mutations (1541 SNPs and indels) previously associated with phenotypic drug-resistance from a curated database. This package also determines the *M. tuberculosis* lineage based on a 90-SNP barcode. The TB-Profiler-predicted resistance mutations were validated using the results of MTBseq, which reports a list of mutations in genes associated with antimicrobial resistance for every processed strain.

Data analysis

Data entry and statistical analyses were performed in SPSS version 16 (SPSS Inc., Illinois, USA). Univariate analyses were performed using two-tailed Fisher's exact test to estimate the association between each variable (*M. tuberculosis* lineage, sublineage or genotypic resistance) and extrapulmonary tuberculosis relative to pulmonary tuberculosis. Additionally, each variable was compared to every anatomical site of TB disease (central nervous system, bones and joints, lymph nodes, and the genitourinary system). Odds ratios with 95% confidence intervals were considered an effect size of the association. To confirm that the association was not an artifact of demographic differences between the geographic regions, a logistic regression model was performed in which the EPTB group was included as a dependent variable. The variables geographic region (the country of isolation), sublineage and genotypic resistance were included in the model, and adjusted ORs with 95% confidence intervals were calculated.

RESULTS

Anatomical site of infection of strains selected for comparative genomics analysis

A total of 490 raw genome datasets of *M. tuberculosis* strains isolated from seven different countries from individuals with a negative HIV infection status (Table S1) were genotyped. The selected strains included those from countries where specific lineages predominate according to the SITVIT website (http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE/), such as Thailand and Indonesia for the East Asian lineage and India for the East African-Indian lineage. Ninety percent of the selected strains were isolated from Indonesia, Thailand, and Russia. This high percentage is since most of the EPTB strains deposited in the NCBI-SRA database were isolated from these countries. Unfortunately, EPTB strains from African and American countries were not included in the conducted analysis because of the lack of genomes in the NCBI-SRA database until the completion of this study.

The EPTB phenotype strains were clustered into five major groups according to the anatomical site of infection, with a predominance of the central nervous system (72.24%).

Other anatomical sites of the disease were bones and joints (17.14%), lymph nodes (4.49%), and the genitourinary system (2.86%); the remaining 3.27% were from sites specified as having only EPTB strains in databases.

Distribution and association of lineages and genotypes with clinical phenotype

The mean coverage obtained for the analyzed strain set was 109.82, whereas the mean percentage of mapped reads was 98.28, resulting in good genome coverage. The strains were assigned to a major *M. tuberculosis* genetic lineage according to the spoligotype using the TB lineage search option of the TB insight web server. These results correlated with the clustering pattern generated by the SNP-based phylogeny. The phylogenetic analysis of *M. tuberculosis* strains from PTB and EPTB phenotypes is shown in Fig. 1. The predominant lineage of the analyzed strains was East Asian (54.28%), followed by Euro-American (34.70%), Indo-Oceanic (9.6%), and East African-Indian (1.42%). The EPTB case percentage differed among these phylogenetic lineages, with 15.51% for Indo-Oceanic, 48.98% for East Asian, 2.45% for East African-Indian, and 33.06% for Euro-American. The same variation in phylogenetic lineages was observed for strains from PTB cases, with 59.59%, 36.32%, 3.67%, and 0.40% for the East Asian, Euro-American, Indo-Oceanic, and East African-Indian lineages, respectively. Fisher's exact test showed that PTB was significantly associated with East Asian lineage strains (OR 1.5; 95% CI [1.1–2.2]) and that EPTB was significantly associated with the Indo-Oceanic lineage strains (OR 4.8; 95% CI [2.2–10.1]) (Table S2). After logistic regression analysis, the associations between L1 and EPTB (OR 4.4; 95% CI [1.9–10.0] $P = 0.000$) and L2 and PTB (OR 1.4; 95% CI [1.1–1.9]) strains was confirmed, thus eliminating any potential bias due to the isolation regions of *M. tuberculosis* strains (Table S2).

We searched for a relationship between the phylogenetic lineages found in the analyzed strains and anatomical sites of infection, including lungs, lymph nodes, the genitourinary system, the central nervous system, and bones and joints. The results showed that strains belonging to the Indo-Oceanic lineage were significantly associated with central nervous system infection (OR 3.9; 95% CI [2.1–7.4]) (Table S3), whereas those of the East Asian lineage were significantly associated with bone and joint infection (OR 1.5; 95% CI [1.3–1.9]) and with PTB strains, as previously described. Further division of identified lineages was conducted using SNP analysis, distinguishing 27 unique sublineages, with the predominant sublineages being 2.2.1 ($n = 218$ strains), 2.2.1.1 ($n = 25$), 4.1.2.1 ($n = 18$), 4.3.1 ($n = 25$), 4.8 ($n = 51$), and 1.2.1 ($n = 22$). As expected, statistical analysis showed a significant relationship of some of these sublineages with major infection sites. In this way, sublineage 1.1.1 was associated with central nervous system infection (OR 2.8; 95% CI [1.0–7.7]), and the same relationship was observed for sublineages 1.2.1 (OR 6.7; 95% CI [2.0–22.1]) and 4.1.2.1 (OR 6.3; 95% CI [1.9–21.1]). On the other hand, sublineages 4.3.1 (OR 2.9; 95% CI [1.2–7.7]), 4.3.3 (OR 1.8; 95% CI [1.4–2.2]), and 4.5 (OR 3.8; 95% CI [1.0–13.8]) were associated with the pulmonary phenotype.

A deeper whole-genome SNP-based classification to discriminate the proto-Beijing from the Beijing strains allowed us to identify the outbreak sublineages of Asia Ancestral 1,

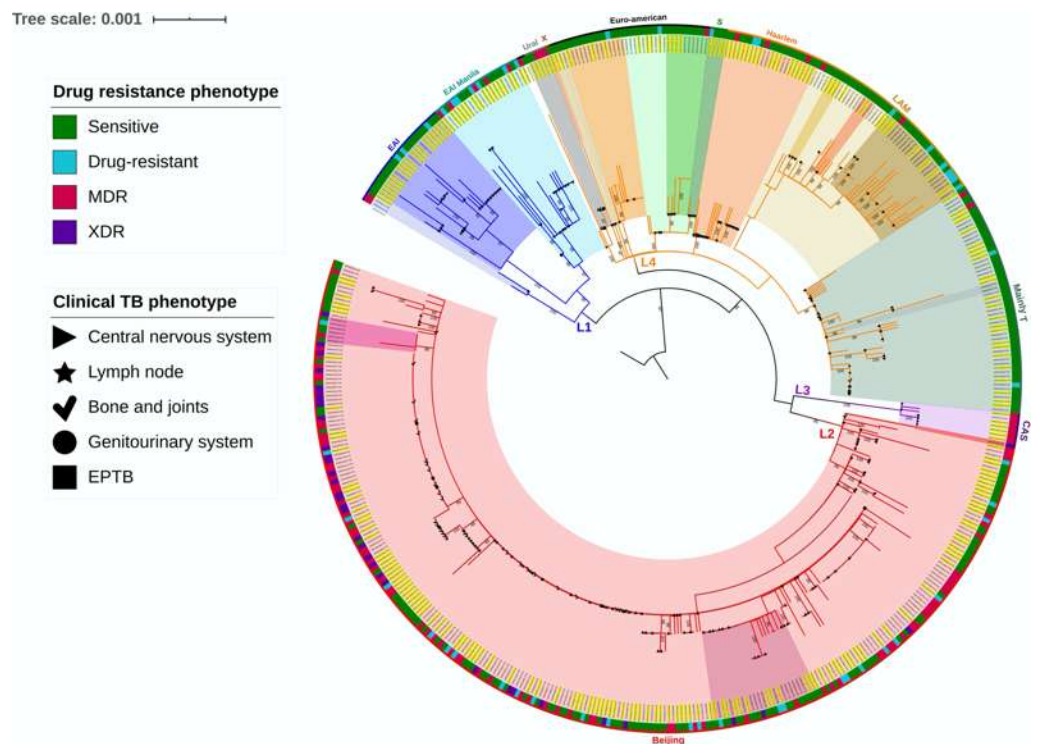


Figure 1 Phylogenetic analysis of *M. tuberculosis* strains from PTB and EPTB phenotypes of the disease. Sensitive, does not present genotypic resistance; Drug-resistant, resistant to at least one antibiotic; MDR, resistant to at least isoniazid and rifampicin. Sensitive, does not present genotypic resistance; Drug-resistant, resistant to at least one antibiotic, MDR, resistant to at least isoniazid and rifampicin, XDR, resistant to isoniazid and rifampicin plus any fluoroquinolone and at least one of three injectable second-line drugs. The phylogenetic tree was inferred using the maximum likelihood (ML) criterion with a general time-reversible model of nucleotide substitution and a gamma model of rate heterogeneity. Yellow highlighted letters indicate EPTB strains. Support values correspond to bootstrap values. The topology was rooted with a *Mycobacterium microti* strain.

Full-size DOI: 10.7717/peerj.12128/fig-1

Asia Ancestral 3, Asian African 2, Asian African 2/RD142, Asian African 3, Pacific RD150, Central Asia, and Europe/Russia B0/W148 and a group of modern unclassified strains using informative genetic markers found in the genomes of the studied strains (Fig. 2). Of all these genotypes, the Asian African 2 sublineage (OR 2.3; 95% CI [1.1–5.2]) and Europe/Russia B0/W148 outbreak sublineage (OR 2.7; 95% CI [1.3–5.4]) were significantly associated with PTB. In the same way, the Ancestral Beijing sublineages (Asia Ancestral 1 and Asia Ancestral 3) (OR 2.4; 95% CI [1.2–4.8]) and Central Asia subgroup (OR 8.2; 95% CI [2.9–22.9]) were significantly associated with EPTB.

The most frequent spoligotype among L2 strains (85.1%) was SIT1 (Table S4). SIT 53 and SIT 19 were the most frequent spoligotypes for the L4 (22.9%) and L1 (27.6%) strains, respectively. For the L3 strains, 42% had an unclassified SIT spoligotype (703775740003771). Spoligotyping allowed the identification of twelve major genotypic families, including Beijing (53.7%), Central Asia (CAS) (1.4%), EAI (9.8%), Haarlem (4.4%), Latin American–Mediterranean (LAM) (7.5%), S (0.2%), T (19.8%), X (0.4%)

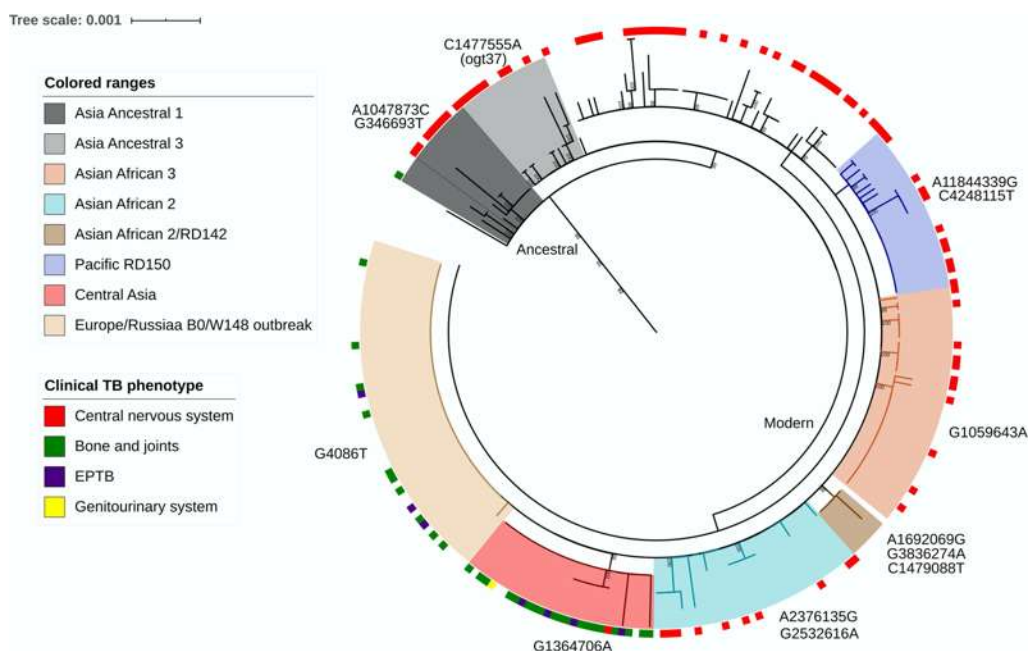


Figure 2 Phylogeny of 263 *Mycobacterium tuberculosis* L2 strains. The phylogeny was constructed by the maximum likelihood (ML) criterion. Classification of strains into sublineages and informative genetic markers are shown. Support values correspond to bootstrap values.

Full-size DOI: [10.7717/peerj.12128/fig-2](https://doi.org/10.7717/peerj.12128/fig-2)

and Family33-36 (2%). Interestingly, 13.26% of the strains corresponded to spoligotypes not previously reported in the SITVIT database (Table S4). A significant association was found between the EAI2 family (OR 6.7; 95% CI [2.0–22.1]) and EPTB, of which 13 strains caused tuberculous meningitis and two of them led to lymph node infection.

Mutations associated with drug-resistant TB

A total of 180 previously reported mutations distributed in 18 genes known to confer resistance to first- and second-line drugs for TB treatment were identified (Table S5). Isoniazid was the antibiotic with the highest predicted resistance in 38.77% of the studied strains, followed by rifampicin in 33.46%, streptomycin in 31.02% and ethambutol in 27.14%.

The most common mutation associated with isoniazid resistance was *katG* Ser315Thr, which was found in 82.63% of the genotypic resistant strains. *rpoB* Ser450Leu was found in 52.43% of the rifampicin-resistant strains, *rpsL* Lys43Arg in 72.36% of streptomycin-resistant strains and *embB* Met306Val in 33.83% of ethambutol-resistant strains.

Extrapulmonary strains showed greater diversity of mutations but appeared less frequently than pulmonary strains. Mutations *gyrA* (Ala90Val) and *thyX* (Glu40Gly) were found only in pulmonary strains. On the other hand, mutations in the *rpoB* (Leu430Pro), *rrs* (907 A>C), *rpsL* (Lys88Thr), *embA* (16C>G), *embB* (Ala356Val, Ser347Ile), *ethA* (Gln269*, Gly385Asp), *katG* (589 insGG), and *pncA* (Ala3Glu, Cys138Arg, Cys72Arg, Gln10Pro, His51Asp, Gly97Asp, Pro54Leu, Thr135Pro) genes were found only in extrapulmonary

strains. There were no statistically significant differences among these variations with the tuberculosis clinical phenotype.

Differences between the distribution of drug-resistant (26.9% PTB, 15.9% EPTB) and drug-sensitive (23.1% PTB, 34.1% EPTB) strains were observed in the EPTB and PTB groups. Four resistance profiles were determined: sensitive, drug resistance (DR), multidrug resistance MDR, and extensive drug resistance (XDR). The frequency for each EPTB/PTB group is shown in [Table S2](#). The bone and joint group presented the highest number of strains with antibiotic resistance: 76.1% were resistant to isoniazid, 73.8% to streptomycin, 66.6% to rifampicin, 54.7% to ethambutol, 42.8% to pyrazinamide and ethionamide, 14.2% to fluoroquinolones and 9.5% to aminoglycosides and paraaminosalicylic acid ([Table S6](#)). A total of 13.5% of central nervous system strains showed resistance to isoniazid, and 54.4% of lymph node strains were resistant to rifampicin. Sensitive strains were statistically associated with EPTB (OR: 1.7; 95% CI [1.3–2.3]), and extensive drug resistance (XDR; OR: 6.0; 95% IC 2.5–14.4) was more strongly associated with PTB strains than with EPTB strains. After logistic regression analysis including the geographic region of isolation, XDR strains were still associated with pulmonary rather than extrapulmonary disease (OR: 6.0; 95% IC [2.3–14.4] $P = 0.000$). We also observed a significant association of Lineage 3 strains with MDR (OR: 3.8; 95% CI [1.1–13.6]) and between the L2 lineage and XDR resistance (OR: 8.1; 5.4–8.6).

DISCUSSION

In this work, we used both raw genome datasets and whole-genome SNP-based phylogeny to genotype the same number of PTB and EPTB *M. tuberculosis* strains. All the strains were included in one of the six lineages defined by Coll ([Coll et al., 2014](#)). When seeking lineage-clinical phenotype relationships, we found that the East Asian ‘modern’ *M. tuberculosis* lineages were associated with PTB, whereas the Indo-Oceanic so-called an ‘ancient lineage’ were associated with EPTB. It is known that strains from ‘modern’ lineages induced a slighter inflammatory response than those from ‘ancient’ lineages, which has been related to a selective advantage of strains from ‘modern’ lineages, resulting in impaired bacterial control by the host, faster disease progression, and enhanced transmission ([Portevin et al., 2011](#)). These clinical differences contribute to explaining the association of the East Asian lineage with the PTB phenotype, supporting the hypothesis that since this is the disease contagious phenotype, patients infected with strains of such a lineage are prone to develop pulmonary disease at a higher frequency than patients infected with strains from other genotypes, which is consistent with the increased transmissibility of strains from this lineage ([Thwaites et al., 2018](#)).

It has been proposed that strains belonging to the Indo-Oceanic lineage are ‘less virulent’ than those from other lineages and cause a specific exacerbated inflammatory response ([Chakraborty et al., 2018](#)), which might be attributed to the presence of unique cell envelope lipids in this lineage, such as phenol phthiocerol dimycocerosate ([Krishnan et al., 2011](#)). Nevertheless, it is still unknown whether these differences in mycobacterial cell envelope lipid composition can explain lineage-related phenotypic differences such as

the EPTB phenotype. Interestingly, the percentage of EPTB cases associated with the East African-Indian lineage was as high as 85%, but this finding was not statistically significant due to the small number of EPTB raw genomes ($n = 7$) from this lineage available in databases. In this regard, the present work reveals database gaps relevant for its public health and epidemiological implications, *i.e.*, the need to include more East African-Indian lineage genomes to clarify its relationships with EPTB.

In the search for a detailed relationship between phylogenetic lineages and the anatomical site of infection, strains belonging to the East Asian lineage were associated with the infection of both bones and joints and PTB. The East Asian lineage comprises two major clades or sublineages, designated proto-Beijing (2.1) and Beijing (2.2) ([Ajawatanawong et al., 2019](#)). Sublineage 2.2, or the Beijing family, as defined by spoligotyping, is composed of several sublineages broadly categorized into the ancestral and modern Beijing strains ([Mokrousov et al., 2005](#)). A subtle SNP-based classification was recently proposed that allows the discrimination of proto-Beijing from Beijing strains ([Shitikov et al., 2017](#)). Such classification divides the Beijing group into the ancestral Beijing clade and the modern Beijing clade, which comprises two groups, one including three strains (Asia Ancestral 1, Asia Ancestral 2, Asia Ancestral 3) and the other including seven strains (Asian African 1, Asian African 2, Asian African 2/RD142, Asian African 3, Pacific RD150, Europe/Russia B0/W148 outbreak and Central Asia). This classification allows us to associate the Asian African 2 sublineage and Europe/Russia B0/W148 outbreak sublineage with PTB, as well as the ancestral Beijing sublineages and Central Asia subgroup with EPTB. Differences in the pathogenicity of Beijing sublineages have been previously reported ([Feng et al., 2008](#)), and sublineage specific patterns of induced cytokine production by macrophages have also been observed ([Sarkar et al., 2012](#)). In this regard, a macrophage infection model revealed that ancestral Beijing strains induce a higher production of the proinflammatory cytokines TNF- α and IL-6 than the modern Beijing sublineage ([Chen et al., 2014](#)). High IFN- γ expression and cytokine production have also been reported in peripheral blood mononuclear cells of ancestral Beijing strains ([Faksri et al., 2014](#)). Such results suggest that the ancestral Beijing strains are as highly immunogenic ([Kato-Maeda et al., 2012](#)) as Lineage 1 strains (EAI) ([Rakotosamimanana et al., 2010](#)). Interestingly, these two sublineages (EAI and ancestral Beijing) were associated with the EPTB phenotype in this study. In contrast, the other sublineage associated with EPTB in this work, 4.1.2.1, has been reported to induce cytokine levels similar to those of H37Rv (Haarlem family strains), which are associated with a low immune response ([Wang et al., 2010](#)). This supports the idea that strains of different sublineages vary by many phenotypes, such as the tendency to develop drug resistance, virulence levels, and immune response, which influence disease severity and clinical presentation.

Interestingly, sublineages 4.1.2.1 and ancestral Beijing, belonging to Lineages 4 and 2, respectively, show a high prevalence worldwide, representing more than 50% of the strains in certain areas and/or subpopulations ([Ajawatanawong et al., 2019](#)). In contrast, Lineage 1 (EAI) strains are commonly reported in countries around the Indian Ocean and are one of the most geographically restricted families. However, EAI strains have been reported in lower percentages in countries such as the Netherlands, Australia, the USA,

Sweden, Saudi Arabia, Tunisia, Taiwan, Panama, and Mexico. The East Asian India 2 spoligotype of Lineage 1 corresponds to the Nonthaburi (EAI2-Nonthaburi) genotype and the Manila (EAI2-Manila) genotype (Couvain, Reynaud & Rastogi, 2019). Recently, Coker (Coker et al., 2016) reported a polymorphism in the genome of Nonthaburi strains from three patients with tuberculous meningitis, consisting of a 500 bp deletion covering ppe50 that was not present in the reference strain *M. tuberculosis* H37Rv. They reported three mutations (T28910C, C1180580T and, C152178T) until now found only in these meningeal Nonthaburi strains. These mutations could represent part of the genetic background that could be shared by strains that cause EPTB and that also belong to the EAI, 4.1.2.1, and ancestral Beijing sublineages. Nonetheless, further investigation is required to determine whether these mutations are shared, which could be the functional consequences and the probable relation of such polymorphisms with TB disease phenotype.

Regarding the genotype relationship with antibiotic resistance, to the best of our knowledge, a significant association of Lineage 3 with drug resistance found here has not been previously reported. This result might be useful to optimize TB treatment in geographical areas where this lineage is frequent. Studies from Asia, Europe and Africa have shown varying associations between drug resistance and MTB lineages (Coscolla & Gagneux, 2010; Singh et al., 2015); however, as was found here, Beijing strains have been associated with MDR and XDR in several cases (Rodríguez-Castillo et al., 2017). However, it must be highlighted that the lack of experimental drug resistance assays is a limitation of this study. The bioinformatic approach to determine the drug resistance of the studied isolates is based on software packages that use different sets of mutations in their analysis. This generates a variation in the drug-resistant genotypes obtained with the different bioinformatic tools with the consequent risk of over- or underestimating the true drug-resistant behavior of the analyzed strains. The statistical inferences that depend on this determination are also subject to such biases as the association of drug resistance with affected organs or mycobacterial lineages. Thus, it is strongly recommended to perform the experimental determination of the drug resistance phenotype of the strains in which the genome will be deposited in public databases.

The advantages of the present study regarding previous works seeking relationships of the *M. tuberculosis* genotype with clinical phenotype include the use of a significantly higher number of strains, an equal number of EPTB and PTB strains, a clear assignment of the strains considered to be associated with EPTB, and the exclusion of the strains from HIV patients, with the latter being a controlled variable, a caution not commonly considered in similar works. However, other possible confounders or known risk factors for EPTB could not be considered in the analysis conducted here, partly due to the lack of additional metadata of clinical information in the database where the genomes of the strains were retrieved. This lack of metadata is not a source of bias in the relationships found here because it has been shown that there is an independent association between lineages and EPTB rather than ethnic factors after a stratified analysis (Click et al., 2012). Additionally, it must be noted that the inclusion of genomes previously published that can be generated for different purposes can generate a bias by sampling protocols to seek specific pathogen genotypes or tuberculosis phenotypes. This was a not controlled variable of this study.

These results support the assertion that the relationship between sublineages and clinical disease phenotypes is not attributable to regional-specific factors; therefore, if this association exists, it does not depend only on unknown clinical factors. However, the contribution of the genetic background of the host to the occurrence of the different clinical phenotypes of tuberculosis should not be ruled out. Several polymorphisms associated with EPTB have been previously reported in specific human populations (Fox *et al.*, 2014). Therefore, associations found here must be further tested with additional epidemiological data to clarify the possible relationship between the intrinsic pathogen sublineage characteristics and host factors as genetic background for the establishment of the infection in a particular organ.

CONCLUSIONS

Overall, the obtained results suggest that intralinear diversity could drive differences in the immune response that trigger the different clinical phenotypes of tuberculosis. The immune response caused by ancient lineage (L1) and ancestral Beijing strains could be similar, eliciting a high inflammatory response. Our results highlight the need both to analyze the genomic background shared by these strains and to perform *in vitro/in vivo* studies that help to elucidate the mechanism by which they could disseminate through the body, causing extrapulmonary disease. We demonstrated that the lack of consistency regarding clinical phenotype-strain genotype associations in the results obtained by previous studies is partially due to the use of inadequate phylogenetic classification tools, which do not allow discrimination between sublineages. Additionally, the present results were not biased by the predominance of a specific lineage/sublineage in a specific human population because we included all the EPTB strains available in databases and an evenly representative sample of PTB strains. Biases originating from unknown sampling procedures of the strains for which genomes are available in databases and due to unknown phenotypic drug resistance are the main limitations of the present study.

ACKNOWLEDGEMENTS

We appreciate the criticism of Dr. Sebastien Gagneux and an anonymous reviewer that contributed to significantly improve the manuscript.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Coordinación de la Investigación Científica UMSNH Research Program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

The Coordinación de la Investigación Científica UMSNH Research Program. Graduate scholarship No. 701889 granted by CONACYT-Mexico to A.M.N.P.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Andrea Monserrat Negrete-Paz performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Gerardo Vázquez-Marrufo and Ma. Soledad Vázquez-Garcidueñas conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw data are available in the [Supplemental Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.12128#supplemental-information>.

REFERENCES

- Ajawanawong P, Yanai H, Smittipat N, Disratthakit A, Yamada N, Miyahara R, Nedsuwan S, Imasanguan W, Kantipong P, Chaiyasirinroje B, Wongyai J, Plitphonganhim S, Tantivitayakul P, Phelan J, Parkhill J, Clark TG, Hibberd ML, Ruangchai W, Palittapongarnpim P, Juthayothin T, Thawornwattana Y, Viratyosin W, Tongsima S, Mahasirimongkol S, Tokunaga K, Palittapongarnpim P. 2019. A novel ancestral Beijing sublineage of *Mycobacterium tuberculosis* suggests the transition site to Modern Beijing sublineages. *Scientific Reports* 9:1–12 DOI [10.1038/s41598-019-50078-3](https://doi.org/10.1038/s41598-019-50078-3).
- Chakraborty P, Kulkarni S, Rajan R, Sainis K. 2018. *Mycobacterium tuberculosis* strains from ancient and modern lineages induce distinct patterns of immune responses. *Journal of Infection in Developing Countries* 11:904–911 DOI [10.3855/jidc.8596](https://doi.org/10.3855/jidc.8596).
- Chen YY, Chang JR, Huang WF, Hsu SC, Kuo SC, Sun JR, Dou HY. 2014. The pattern of cytokine production in vitro induced by ancient and modern Beijing *Mycobacterium tuberculosis* strains. *PLOS ONE* 9(4):e94296 DOI [10.1371/journal.pone.0094296](https://doi.org/10.1371/journal.pone.0094296).
- Click ES, Moonan PK, Winston CA, Cowan LS, Oeltmann JE. 2012. Relationship between *Mycobacterium tuberculosis* phylogenetic lineage and clinical site of tuberculosis. *Clinical Infectious Disease* 54:211–219 DOI [10.1093/cid/cir788](https://doi.org/10.1093/cid/cir788).
- Coker OO, Chaiprasert A, Ngamphiw C, Tongsima S, Regmi SM, Clark TG, Ong RT, Teo YY, Prammananan T, Palittapongarnpim P. 2016. Genetic signatures of *Mycobacterium tuberculosis* Nonthaburi genotype revealed by whole genome analysis of strains from tuberculous meningitis patients in Thailand. *PeerJ* 4:1–20 DOI [10.7717/peerj.1905](https://doi.org/10.7717/peerj.1905).

- Coll F, Mc Nerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, Portugal I, Pain A, Martin N, Clark TG. 2014. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nature Communications* 5:4812 DOI 10.1038/ncomms5812.
- Coll F, Mc Nerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G, Mallard K, Nair M, Miranda A, Alves A, Perdigão J, Viveiros M, Portugal I, Hasan Z, Hasan R, Glynn JR, Martin N, Pain A, Clark TG. 2015. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Medicine* 7:51–61 DOI 10.1186/s13073-015-0164-0.
- Coscolla M, Gagneux S. 2010. Does *M. tuberculosis* genomic diversity explain disease diversity? *Drug Discovery Today Disease Mechanisms* 7(1):e43–e59 DOI 10.1016/j.ddmec.2010.09.004.
- Coscolla M, Gagneux S. 2014. Consequences of genomic diversity in *Mycobacterium tuberculosis*. *Seminars in Immunology* 26:431–444 DOI 10.1016/j.smim.2014.09.012.
- Coscolla M, Gagneux S, Menardo F, Loiseau C, Ruiz-Rodriguez P, Borrell S, Otchere ID, Asante-Poku A, Asare P, Sánchez-Busó L, Gehre F, Sanoussi CN, Antonio M, Affolabi D, Fyfe J, Beckert P, Niemann S, Alabi AS, Grobusch MP, Kobbe R, Parkhill J, Beisel C, Fenner L, Böttger EC, Meehan CJ, Harris SR, De Jong BC, Yeboah-Manu D, Brites D. 2021. Phylogenomics of *Mycobacterium africanum* reveals a new lineage and a complex evolutionary history. *Microbial Genomics* 7:1–14 DOI 10.1099/mgen.0.000477.
- Couvin D, Reynaud Y, Rastogi N. 2019. Two tales: worldwide distribution of Central Asian (CAS) versus ancestral East-African Indian (EAI) lineages of *Mycobacterium tuberculosis* underlines a remarkable cleavage for phylogeographical, epidemiological and demographical characteristics. *PLOS ONE* 14(7):e0219706 DOI 10.1371/journal.pone.0219706.
- Dale JW, Bothamley GH, Drobniewski F, Gillespie SH, McHugh TD, Pitman R. 2005. Origins and properties of *Mycobacterium tuberculosis* strains in London. *Journal of Medical Microbiology* 54:575–582 DOI 10.1099/jmm.0.45959-0.
- Faksri K, Chaiprasert A, Pardieu C, Casali N, Palaga T, Prammananan T, Palitapongarnpim P, Prayoonwivat N, Drobniewski F. 2014. Heterogeneity of phenotypic characteristics of the modern and ancestral Beijing strains of *Mycobacterium tuberculosis*. *Asian Pacific Journal of Allergy and Immunology* 32:124–132 DOI 10.12932/AP0361.32.2.2013.
- Faksri K, Xia E, Ong RT, Tan JH, Nonghanphithak D, Makhao N, Thamnongdee N, Thanormchat A, Phurattanakornkul A, Rattanarangsee S, Ratanajaraya C, Suriyaphol P, Prammananan T, Teo YY, Chaiprasert A. 2018. Comparative whole-genome sequence analysis of *Mycobacterium tuberculosis* isolated from tuberculous meningitis and pulmonary tuberculosis patients. *Scientific Reports* 8:4910 DOI 10.1038/s41598-018-23337-y.
- Feng JY, Su WJ, Tsai CC, Chang SC. 2008. Clinical impact of *Mycobacterium tuberculosis* W-Beijing genotype strain infection on aged patients in Taiwan. *Journal of Clinical Microbiology* 46:3127–3129 DOI 10.1128/JCM.01132-08.

- Firdessa R, Berg S, Hailu E, Schelling E, Gumi B, Erenso G, Gadisa E, Kiros T, Habtamu M, Hussein J, Zinsstag J, Robertson BD, Ameni G, Lohan AJ, Loftus B, Comas I, Gagneux S, Tschopp R, Yamuah L, Hewinson G, Gordon SV, Young DB, Aseffa A. 2013. Mycobacterial lineages causing pulmonary and extrapulmonary tuberculosis, Ethiopia. *Emerging Infectious Disease* 19:460–463 DOI 10.3201/eid1903.120256.
- Floyd K, Glaziou P, Zumla A, Raviglione M. 2018. The global tuberculosis epidemic progress in patient care, prevention and control efforts in year 3 of the End TB Era. *Lancet Respiratory Medicine* 6:299–314 DOI 10.1016/S2213-2600(18)30057-2.
- Fox GJ, Sy DN, Nhung NV, Yu B, Ellis MK, Van Hung N, Cuong NK, Lien LThi, MarksGB, Saunders BM, Britton WJ. 2014. Polymorphisms of SP110 are associated with both pulmonary and extra-pulmonary tuberculosis among the Vietnamese. *PLOS ONE* 9(7):e99496 DOI 10.1371/journal.pone.0099496.
- Gagneux S. 2018. Ecology and evolution of *Mycobacterium tuberculosis*. *Nature Reviews Microbiology* 16:202–213 DOI 10.1038/nrmicro.2018.8.
- Golden MP, Vikram HR. 2005. Extrapulmonary tuberculosis: an overview. *American Family Physician* 72:1761–1768.
- Homolka S, Projahn M, Feuerriegel S, Ubben T, Diel R, Nübel U, Niemann S. 2012. High resolution discrimination of clinical *Mycobacterium tuberculosis* complex strains based on single nucleotide polymorphisms. *PLOS ONE* 7:1–11 DOI 10.1371/journal.pone.0039855.
- Jajou R, Kohl TA, Walker T, Norman A, Cirillo DM, Tagliani E, Niemann S, De Neeling A, Lillebaek T, Anthony RM, Van Soolingen D. 2019. Towards standardisation: comparison of five whole genome sequencing (WGS) analysis pipelines for detection of epidemiologically linked tuberculosis cases. *Eurosurveillance* 24:1–10 DOI 10.2807/1560-7917.ES.2019.24.50.1900130.
- Kato-Maeda M, Nahid P. 2012. *Mycobacterium tuberculosis* lineage-what's in your lungs? *Clinical Infectious Disease* 54:220–224 DOI 10.1093/cid/cir795.
- Kato-Maeda M, Shanley CA, Ackart D, Jarlsberg LG, Shang S, Obregon-Henao A, Harton M, Basaraba RJ, Henao-Tamayo M, Barrozo JC, Rose J, Kawamura LM, Coscolla M, Fofanov VY, Koshinsky H, Gagneux S, Hopewell PC, Ordway DJ, Orme IM. 2012. Beijing sublineages of *Mycobacterium tuberculosis* differ in pathogenicity in the guinea pig. *Clinical Vaccine Immunology* 19:1227–1237 DOI 10.1128/CVI.00250-12.
- Kohl TA, Utpatel C, Schleusener V, De Filippo MR, Beckert P, Cirillo DM, Niemann S. 2018. MTBseq: a comprehensive pipeline for whole genome sequence analysis of *Mycobacterium tuberculosis* complex strains. *PeerJ* 6:e5895 DOI 10.7717/peerj.5895.
- Krishnan N, Malaga W, Constant P, Caws M, Tran TH, Salmons J, Nguyen TN, Nguyen DB, Daffé M, Young DB, Robertson BD, Guilhot C, Thwaites GE. 2011. *Mycobacterium tuberculosis* lineage influences innate immune response and virulence and is associated with distinct cell envelope lipid profiles. *PLOS ONE* 6(9):e23870 DOI 10.1371/journal.pone.0023870.

- Letunic I, Bork P. 2016.** Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* **44**:W242–W245 DOI [10.1093/nar/gkw290](https://doi.org/10.1093/nar/gkw290).
- Malik AN, Godfrey-Faussett P. 2005.** Effects of genetic variability of *Mycobacterium tuberculosis* strains on the presentation of disease. *The Lancet Infectious Disease* **5**:174–183 DOI [10.1016/S1473-3099\(05\)01310-1](https://doi.org/10.1016/S1473-3099(05)01310-1).
- McHenry ML, Bartlett J, Igo RP, Wampande EM, Benchek P, Mayanja-Kizza H, Benchek P, Mayanja-Kizza H, Fluegge K, Hall NB, Gagneux S, Tishkoff SA, Wejse C, Sirugo G, Boom WH, Joloba M, Williams SM, Stein CM. 2020.** Interaction between host genes and *Mycobacterium tuberculosis* lineage can affect tuberculosis severity: evidence for coevolution? *PLOS Genetics* **16**(4):e1008728 DOI [10.1371/journal.pgen.1008728](https://doi.org/10.1371/journal.pgen.1008728).
- Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, Willery E, Blum MG, Rüscher-Gerdes S, Mokrousov I, Aleksic E, Allix-Béguec C, Antierens A, Augustynowicz-Kopeć E, Ballif M, Barletta F, Beck HP, Barry 3rd CE, Bonnet M, Borroni E, Campos-Herrero I, Cirillo D, Cox H, Crowe S, Crudu V, Diel R, Drobniewski F, Fauville-Dufaux M, Gagneux S, Ghebremichael S, Hanekom M, Hoffner S, Jiao WW, Kalon S, Kohl TA, Kontsevaya I, Lillebæk T, Maeda S, Nikolayevskyy V, Rasmussen M, Rastogi N, Samper S, Sanchez-Padilla E, Savic B, Shamputa IC, Shen A, Sng LH, Stakenas P, Toit K, Varaine F, Vukovic D, Wahl C, Warren R, Supply P, Niemann S, Wirth T. 2015.** Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nature Genetics* **47**:242–249 DOI [10.1038/ng.3195](https://doi.org/10.1038/ng.3195).
- Mokrousov I, Ly HM, Otten T, Lan NN, Vyshnevskiy B, Hoffner S, Narvskaya O. 2005.** Origin and primary dispersal of the *Mycobacterium tuberculosis* Beijing genotype: clues from human phylogeography. *Genome Research* **15**:1357–1364 DOI [10.1101/gr.3840605](https://doi.org/10.1101/gr.3840605).
- Ngabonziza JCS, Loiseau C, Marceau M, Jouet A, Menardo F, Tzfadia O, Antoine R, Niyigena EB, Mulders W, Fissette K, Diels M, Gaudin C, Duthoy S, Ssengooba W, André E, Kaswa MK, Habimana YM, Brites D, Affolabi D, Mazarati JB, De Jong BC, Rigouts L, Gagneux S, Meehan CJ, Supply P. 2020.** A sister lineage of the *Mycobacterium tuberculosis* complex discovered in the African Great Lakes region. *Nature Communications* **11**:2917 DOI [10.1038/s41467-020-16626-6](https://doi.org/10.1038/s41467-020-16626-6).
- Nguyen L. 2016.** Antibiotic resistance mechanisms in *M. tuberculosis*: an update. *Archives of Toxicology* **90**:1585–1604 DOI [10.1007/s00204-016-1727-6](https://doi.org/10.1007/s00204-016-1727-6).
- Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015.** IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**:268–274 DOI [10.1093/molbev/msu300](https://doi.org/10.1093/molbev/msu300).
- Portevin D, Gagneux S, Comas I, Young D. 2011.** Human macrophage responses to clinical strains from the *Mycobacterium tuberculosis* complex discriminate between ancient and modern lineages. *PLOS Pathogens* **7**(3):e1001307 DOI [10.1371/journal.ppat.1001307](https://doi.org/10.1371/journal.ppat.1001307).

- Rakotosamimanana N, Raharimanga V, Andriamandimby SF, Soares JL, Doherty TM, Ratsitorahina M, Ramarokoto H, Zumla A, Huggett J, Rook G, Richard V, Gicquel B. 2010. Variation in gamma interferon responses to different infecting strains of *Mycobacterium tuberculosis* in acid-fast bacillus smear-positive patients and household contacts in Antananarivo, Madagascar. *Clinical Vaccine Immunology* 17:1094–1203 DOI 10.1128/CVI.00049-10.
- Rodríguez-Castillo JG, Pino C, Niño LF, Roza JC, Llerena-Polo C, Parra-López CA, Tauch A, Murcia-Aranguren MI. 2017. Comparative genomic analysis of *Mycobacterium tuberculosis* Beijing-like strains revealed specific genetic variations associated with virulence and drug resistance. *Infection Genetics and Evolution* 54:314–323 DOI 10.1016/j.meegid.2017.07.022.
- Sarkar R, Lenders L, Wilkinson KA, Wilkinson RJ, Nicol MP. 2012. Modern lineages of *Mycobacterium tuberculosis* exhibit lineage-specific patterns of growth and cytokine induction in human monocyte-derived macrophages. *PLOS ONE* 7(8):e43170 DOI 10.1371/journal.pone.0043170.
- Shitikov E, Kolchenko S, Mokrousov I, Bespyatykh J, Ischenko D, Ilina E, Govorun V. 2017. Evolutionary pathway analysis and unified classification of East Asian lineage of *Mycobacterium tuberculosis*. *Scientific Reports* 7:9227 DOI 10.1038/s41598-017-10018-5.
- Singh J, Sankar MM, Kumar P, Couvin D, Rastogi N, Singh S. 2015. Genetic diversity and drug susceptibility profile of *Mycobacterium tuberculosis* isolated from different regions of India. *Journal of Infectology* 71:207–219 DOI 10.1016/j.jinf.2015.04.028.
- Stucki D, Brites D, Jeljeli L, Coscolla M, Liu Q, Trauner A, Fenner L, Rutaihua L, Borrell S, Luo T, Gao Q, Kato-Maeda M, Ballif M, Egger M, Macedo R, Mardassi H, Moreno M, Vilanova GTudo, Fyfe J, Globan M, Thomas J, Jamieson F, Guthrie JL, Asante-Poku A, Yeboah-Manu D, Wampande E, Ssengooba W, Joloba M, Henry Boom W, Basu I, Bower J, Saraiva M, Vaconcellos SEG, Suffys P, Koch A, Wilkinson R, Gail-Bekker L, Malla B, Ley SD, Beck HP, De Jong BC, Toit K, Sanchez-Padilla E, Bonnet M, Gil-Brusola A, Frank M, Penlap Beng VN, Eisenach K, Alani I, Wangui Ndung'u P, Revathi G, Gehre F, Akter S, Ntoumi F, Stewart-Isherwood L, Ntinginya NE, Rachow A, Hoelscher M, Cirillo DM, Skenders G, Hoffner S, Bakonyte D, Stakenas P, Diel R, Crudu V, Moldovan O, Al-Hajoj S, Otero L, Barletta F, Jane Carter E, Diero L, Supply P, Comas I, Niemann S, Gagneux S. 2016. *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. *Nature Genetics* 48:1535–1543 DOI 10.1038/ng.3704.
- Svensson E, Millet J, Lindqvist A, Olsson M, Ridell M, Rastogi N, Western Sweden Tuberculosis Epidemiology Study Group. 2011. Impact of immigration on tuberculosis epidemiology in a low-incidence country. *Clinical Microbiology and Infection* 17:881–887 DOI 10.1111/j.1469-0691.2010.03358.
- Thwaites G, Caws M, Chau TT, D'Sa A, Lan NT, Huyen MN, Gagneux S, Anh PT, Tho DQ, Torok E, Nhu NT, Duyen NT, Duy PM, Richenberg J, Simmons C, Hien TT, Farrar J. 2018. Relationship between *Mycobacterium tuberculosis* genotype and the

clinical phenotype of pulmonary and meningeal tuberculosis. *Journal of Clinical Microbiology* **46**:1363–1368 DOI [10.1128/JCM.02180-07](https://doi.org/10.1128/JCM.02180-07).

Wang C, Peyron P, Mestre O, Kaplan G, Van Soolingen D, Gao Q, Gicquel B, Neyrolles O. 2010. Innate immune response to *Mycobacterium tuberculosis* Beijing and other genotypes. *PLOS ONE* **5**(10):e13594 DOI [10.1371/journal.pone.0013594](https://doi.org/10.1371/journal.pone.0013594).

Wang XH, Ma AG, Han XX, Gu XM, Fu LP, Li PG, Li FY, Wang QZ, Liang H, KatarA, Wang LJ. 2015. Correlations between drug resistance of Beijing/W lineage clinical strains of *Mycobacterium tuberculosis* and sublineages: a 2009–2013 prospective study in Xinjiang province, China. *Medical Science Monitor* **21**:1313–1318 DOI [10.12659/MSM.892951](https://doi.org/10.12659/MSM.892951).

Xia E, Teo YY, Ong RT. 2016. SpoTyping: fast and accurate in silico *Mycobacterium* spoligotyping from sequence reads. *Genome Medicine* **8**:19 DOI [10.1186/s13073-016-0270-7](https://doi.org/10.1186/s13073-016-0270-7).

**CAPÍTULO II: Caracterización genómica y fenotípica de cepas de
Mycobacterium tuberculosis aisladas en el estado de Michoacán, México**

INTRODUCCIÓN

La tuberculosis (TB) es uno de los padecimientos más antiguos de la humanidad; sin embargo, sigue siendo una de las enfermedades transmisibles con mayor tasa de mortalidad. Esta enfermedad afecta a grandes grupos de población, particularmente de áreas marginadas y grupos vulnerables donde predomina la pobreza, la desnutrición y el hacinamiento. En 2020, más de 10 millones de personas enfermaron de TB y cerca de 1.8 millones murieron por esta enfermedad, entre ellos, 0.4 millones de personas con VIH (OMS, 2021). Más del 95% de las muertes por TB se producen en países de ingresos bajos y medianos. La rapidez con que se ha propagado la epidemia de VIH en otras regiones del mundo ha provocado un aumento del número de casos de TB. A pesar de que es una enfermedad curable, se necesita de cuatro antibióticos administrados durante seis a nueve meses, lo que redundo en una alta tasa de abandono del tratamiento. Esta situación ha promovido recaídas y el surgimiento de cepas resistentes a varios antibióticos, lo cual complica aún más el tratamiento al incrementar su costo y toxicidad (Khawbung et al., 2021). Gracias a las estrategias globales de control, los casos de TB van decayendo lentamente; sin embargo, dado que la mayoría de las muertes por TB son evitables, la mortalidad de esta enfermedad sigue siendo inaceptablemente alta.

De acuerdo con la información disponible en el componente de TB del Sistema Nacional de Vigilancia Epidemiológica (SINAVE, 2021), en México se registraron 20,075 casos nuevos en 2021, de los cuales 16,008 pertenecen a casos de tuberculosis pulmonar (PTB), 455 casos pertenecen a TB meníngea, una de las formas más severas de la enfermedad (Hernandez-Pando et al., 2013), y 3,612 a otras formas de TB extrapulmonar (EPTB). Esto representa una importante pérdida de años potenciales de vida saludable y, por consiguiente, un obstáculo para el desarrollo socioeconómico de México, ya que el promedio de edad de muerte por este padecimiento es de 54 años. Cabe destacar que se ha observado una baja en el número de casos nuevos reportados en comparación con los años 2019 y 2020 a nivel mundial y nacional; sin embargo, no es debido a que la población enferme menos, sino a causa de las interrupciones en el acceso a los servicios de

diagnóstico y tratamiento desde que inicio la pandemia por COVID-19. Otra razón es que, en el contexto de los confinamientos y otras restricciones a la movilidad, los pacientes y otras personas potencialmente contagiadas han dejado de acudir a los centros de salud a solicitar una consulta médica o a dar seguimiento a su tratamiento antifímico. La consecuencia es que, según la OMS, ahora existen alrededor de 4.1 millones de personas que padecen TB que no han sido diagnosticadas. Se cree que más del 40% de casos de TB no se diagnosticaron en 2020, ocasionando un aumento en la tasa de transmisión que se verá reflejada durante los próximos años.

A pesar de que Michoacán se encuentra por debajo de la media nacional de casos de TB, en los últimos años se ha observado una tendencia de incremento en la tasa de incidencia de PTB y EPTB (Figura 4). Como se puede observar para el año 2021 se reporta un decremento en la incidencia de PTB del 63%, fenómeno atribuido a una disminución en el reporte de nuevos casos probablemente debido a la pandemia por COVID-19. Este aumento en el número de casos de TB en el estado en los últimos años puede deberse a diversos factores como las características inmunogenéticas de la población, sus entornos ecológicos y socioculturales, la coevolución de las cepas de MTB y el fondo genómico propio de las cepas circulantes.

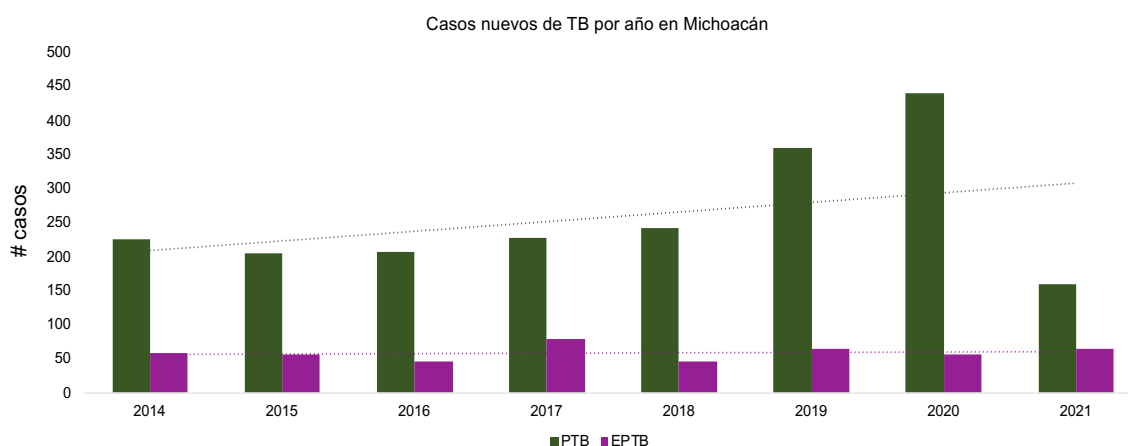


Figura 4. Incidencia por año de casos de PTB y EPTB en Michoacán. PTB, tuberculosis pulmonar; EPTB, tuberculosis extrapulmonar; TB, tuberculosis. Fuente: SINAVE, 2021.

Se habla de coevolución entre MTB y el ser humano debido a la marcada distribución geográfica y amplitud temporal de las cepas del patógeno. En este proceso, la bacteria y su hospedero se han adaptado a eventos biológicos de sus contrapartes. La coevolución ocurre cuando los cambios evolutivos en el patógeno que aumentan la infectividad de este son contrarrestados por cambios evolutivos en el hospedero que aumentan la resistencia a la infección (Brites & Gagneux, 2015). Diversos estudios han demostrado una fuerte asociación entre las cepas de MTB y diversas poblaciones humanas, en donde a pesar de los fenómenos de migración y la consecuente dispersión de cepas, estas asociaciones se mantienen (Liang et al., 2020). En este sentido, existen cepas con una fuerte restricción geográfica y algunas otras distribuidas mundialmente. Para realizar el seguimiento epidemiológico de las cepas circulantes de MTB se utilizan diversas técnicas de tipificación molecular como la espoligotipificación, la tipificación del número variable de copias repetidas en tándem (VNTR por sus siglas en inglés), la tipificación del número de copias del elemento transponible IS6110 por análisis del polimorfismo de fragmentos de restricción (RFLP IS6110) y la determinación de linaje o sublinaje mediante la búsqueda de SNPs en el genoma secuenciado.

La espoligotipificación es una técnica basada en la presencia o ausencia de secuencias espaciadoras en la región de repetidos directos (DR) del cromosoma de MTB (Gori et al., 2005). Los loci DR constan de secuencias repetidas de 36 pares de bases (pb) separadas por secuencias espaciadoras de 35 a 41 pb. Esta técnica sirve para caracterizar familias de micobacterias en una escala intraespecífica utilizando un código octal o código binario para la identificación de espoligotipos, lo que ha permitido la creación de bases de datos mundiales que permiten la comparación inter-laboratorios como SITVIT. En estas bases de datos se encuentran depositados los espoligotipos, cada uno de los cuales tiene asignado un SIT o "*spoligo international type*", un código binario, un código octal, una distribución geográfica y una familia (Brudey et al., 2006). En relación con las unidades interespaciadoras micobacterianas repetidas-número variable de repeticiones en tándem (MIRU-VNTR), esta se basa en amplificaciones por PCR de múltiples loci con una cantidad variable de repeticiones consecutivas usando primers específicos

para las regiones que flanquean cada locus repetido, obteniéndose un amplicón cuyo tamaño refleja el número de copias de MIRU-VNTR; los perfiles de MIRU están representados como códigos numéricos y cada dígito representa el número de copias en un locus (Supply et al., 2000).

Una técnica que fue muy utilizada fue la tipificación del número de copias del elemento transponible IS6110 por análisis del polimorfismo de fragmentos de restricción (RFLP IS6110). La secuencia de inserción IS6110 se ha utilizado ampliamente como marcador genético debido a su aparente gran movilidad y su alto número de copias en la mayoría de las muestras clínicas. Es un elemento transponible exclusivo del CMTB de 1355 pb que se encuentra más o menos distribuido al azar a través del genoma de la bacteria. Los números de copias pueden variar desde 26 copias a la ausencia de estas (Arora et al., 2020). El hecho de presentarse en múltiples copias y en diferentes lugares del genoma permitió que fuera utilizado como el método de tipificación “estándar de oro” durante muchos años (Coros et al, 2008); sin embargo, presenta un poder discriminatorio bajo en las muestras que tienen menos de 6 copias de IS6110 y es casi nula la discriminación con las que tienen una o dos copias de la misma secuencia de inserción.

La técnica de discriminación empleada en la actualidad es la tipificación por SNPs en el genoma. Se trata de una técnica basada en la evaluación de SNPs previamente conocidos y asociados a un linaje específico. Los SNPs son los marcadores filogenéticos más confiables para el estudio evolutivo del CMTB ya que poseen niveles despreciables de homoplasia (Comas et al., 2009). Por lo que es considerada la técnica más discriminativa para medir diversidad genética en MTB. Previamente se han sugerido paneles de SNPs para llevar a cabo la clasificación de cepas en linajes y sublinajes (Coll et al., 2014; Filliol et al., 2016), lo que ha permitido la identificación de 9 linajes principales. Los linajes 2 y 4 conocidos como linajes “modernos” son los responsables de la mayor parte de los casos de TB a nivel mundial y ciertas cepas pertenecientes a estos linajes han estado implicadas en brotes importantes de TB sensible y drogo resistente. Además, se ha reportado que las cepas pertenecientes a estos linajes ocasionan una progresión más rápida a la enfermedad y son más virulentas en comparación con los linajes ancestrales

(Coscolla & Gagneux, 2014). Su origen monofilético y su distribución global sugieren que pudieron haber infectado a la población africana que realizó las primeras migraciones y posteriormente se llevó a cabo un proceso de diversificación en otras poblaciones humanas.

Se ha observado que ciertos linajes e incluso sublinajes están más extendidos que otros, por ejemplo, la marcada restricción geográfica de los linajes 5, 6 y 7 en comparación con los linajes 2 y 4. Además, dentro de los sublinajes pertenecientes al linaje 4 también se pueden encontrar estas diferencias; los sublinajes Harleem y LAM se encuentran distribuidos mundialmente, mientras que los sublinajes Camerún o Uganda difícilmente se encuentran en América Latina (Brites & Gagneux, 2015). Por lo tanto, la distribución diferencial de linajes y sublinajes de MTB también podría reflejar características intrínsecas de las cepas. Por ejemplo, algunos linajes podrían considerarse “generalistas”, es decir, capaces de persistir en diferentes poblaciones humanas, y otros más “especialistas”, es decir, capaces de persistir solo en una o unas pocas poblaciones particulares. Todo esto resalta la importancia de llevar a cabo una caracterización de cepas regionales de MTB con la finalidad de identificar características propias de las mismas que ayuden a explicar su comportamiento fenotípico.

En este trabajo se llevó a cabo la caracterización genómica y fenotípica de 10 cepas causantes de TB en el estado de Michoacán, con la finalidad de identificar a qué linajes y sublinajes pertenecen, así como evaluar su morfología colonial, el fenotipo formador de cuerdas y el patrón de resistencia a antibióticos.

MATERIALES Y MÉTODOS

Cepas bacterianas

Las cepas de estudio fueron donadas por el Laboratorio Estatal de Salud Pública del estado de Michoacán. Corresponden a 10 cepas causantes de TB, siendo 5 PTB y 5 EPTB. Las características de los aislados se muestran en la Tabla 1, a cada uno de ellos se le asignó una clave que comienza con el prefijo MYC, seguido de un número.

Tabla 1. Características epidemiológicas de las cepas de estudio.

CLAVE	SEXO	EDAD	MUNICIPIO	TIPO TB	ENF. ASOCIADAS
MYC036	Femenino	21	Lázaro Cárdenas	Meníngea	Ninguna
MYC052	Masculino	No reportado	No reportado	Renal	Ninguna
MYC061	Masculino	35	Morelia	Pulmonar	Ninguna
MYC072	Masculino	63	Morelia	Pulmonar	Ninguna
MYC075	Masculino	No reportado	No reportado	Pulmonar	Ninguna
MYC078	Masculino	1	Lázaro Cárdenas	Meníngea	Ninguna
MYC087	Masculino	74	Chilchota	Linfática	Ninguna
MYC093	Femenino	No reportado	Charo	Pulmonar	Ninguna
MYC096	Masculino	54	Zamora	Pulmonar	Ninguna
MYC136	Masculino	24	No reportado	Digestiva	Ninguna

Cultivo de *M. tuberculosis*

Las cepas seleccionadas se cultivaron en medio Lowenstein-Jensen y se incubaron a 37°C durante 6 semanas. Los cultivos fueron revisados a las 24 y 48 h para descartar contaminación. Posteriormente se resembró un inocúlo en el mismo medio y se incubó a 37°C hasta observar crecimiento abundante.

Caracterización genómica

Extracción de ADN

Las colonias crecidas en medio Lowenstein-Jensen se suspendieron en 600 μ L de regulador de lisis (Tris-HCl 100 mM, pH 8.0, SDS 0.2% NaCl 100 mM y EDTA 50 mM) y se inactivaron por incubación a 80°C durante 1 h. Posteriormente, la suspensión se enfrió en hielo durante 2 minutos, al término de los cuales se agregaron 600 μ L de mezcla de fenol y cloroformo (1:1 v/v) y se agitó vigorosamente por 10 min. La mezcla se centrifugó a 13.4 x g durante 10 min. El sobrenadante se recuperó en un tubo nuevo y se añadió un volumen igual de mezcla de fenol y cloroformo para posteriormente repetir los pasos de agitación y centrifugado. El sobrenadante se recuperó una vez más en un tubo nuevo y se agregaron 2 volúmenes de isopropanol frío. Hasta este punto, el manejo se llevó a cabo en una campana de bioseguridad nivel 2, con prácticas de bioseguridad nivel 3. La mezcla se incubó a -20°C por 20 min, después de los cuales se centrifugó a 13.4 x g por 10 min. La pastilla de ADN se lavó con etanol al 70%, centrifugada a 13.4 x g por 10 min y secada al aire hasta evaporación completa del etanol. La pastilla de ADN obtenida se resuspendió en 20 μ L de agua bidestilada estéril. Para constatar la presencia e integridad del DNA, cada muestra se analizó mediante electroforesis en gel de agarosa al 0.8 % (p/v) teñido con bromuro de etidio (10 mg/ml). Finalmente, la concentración de ADN se determinó mediante el sistema espectrofotométrico NanoDrop (Thermo Scientific) de acuerdo con las especificaciones del proveedor.

Purificación de ADN

El ADN se purificó mediante la técnica de *glass-milk* mediante un protocolo desarrollado en el laboratorio. De cada muestra se escindió el fragmento del gel de agarosa que contenía el ADN de alto peso molecular visualizado en los geles de agarosa al final de la electroforesis descrita anteriormente. Cada uno de los fragmentos de agarosa obtenidos se colocó en un tubo para microcentrífuga de 1.5 ml, se adicionaron 480 μ l de yoduro de sodio 6M y se incubaron a 50°C por 5 min, mezclando esporádicamente. Se añadieron 50 μ l de *glass-milk* a la solución

resultante y se incubó a 4°C por 10 min, centrifugando posteriormente a 9.7 g por 30 s. La pastilla obtenida se lavó 3 veces con 500 µl de una solución preparada con 50 ml de etanol absoluto y 50 ml de una solución Tris-HCl 20 mM, EDTA 1 mM y NaCl 0.1 M. Después de estos pasos de lavado la pastilla resultante se resuspendió con 20 µl de agua bidestilada y se incubó durante 10 min a 55°C, posteriormente se centrifugó a 9.7 g por 30 s. Se removió el DNA eluido y se colocó en un tubo nuevo. Finalmente, se colocaron 10 µl de agua bidestilada a la pastilla de *glass-milk*, se incubó por 10 min a 55°C y se centrifugó a 9.7 g por 30 s, se recuperó el sobrenadante y se mezcló con el de la primera elución. Se verificó la calidad y cantidad del ADN purificado mediante electroforesis en gel de agarosa y determinación espectrofotométrica, respectivamente, como se describió anteriormente.

Secuenciación del genoma

Las muestras de ADN purificadas se enviaron a la empresa Genewiz (Seattle, USA) para su secuenciación, la cual se realizó con el kit HiSeq200 en un secuenciador HiSeq de la plataforma Illumina (Illumina, CA., USA), produciendo lecturas (*reads*) de secuenciación de una longitud de 250 pb en formato “.fastq”. Se obtuvieron dos archivos con en formato de acuerdo con el sentido de las lecturas: 5’-> 3’ (forward) y 3’-> 5’ (reverse), por cada genoma. Por lo que los análisis subsecuentes se realizaron a partir de dichos archivos.

Análisis bioinformático

Revisión de calidad de las lecturas

Con la finalidad de visualizar la calidad de las lecturas de secuenciación, se empleó el programa Fastqc (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>), el cual evalúa las lecturas y provee una serie de estadísticas que permiten inferir la calidad del proceso de secuenciación. Los criterios de calidad principalmente evaluados fueron: tamaño en nucleótidos, número de lecturas, calidad de las secuencias por base, contenido de adaptadores y proporción del contenido de

guanina-citosina. Las lecturas que presentaron secuencias adaptadoras y/o un valor de calidad phred por debajo de 30 fueron filtradas utilizando TrimGalore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Este índice de calidad se otorga de forma individual a cada nucleótido, reflejando la probabilidad de que un nucleótido determinado sea erróneo (P). Un índice de calidad Phred 30 determina que esa probabilidad de error sea de 1 entre 1000, asegurando una precisión del 99.9 %.

Tipificación *in silico*

Para llevar a cabo la predicción de espoligotipo a partir de las lecturas de secuenciación se utilizó la herramienta Spotyping versión 2.0 (Xia et al., 2016). Spotyping acepta archivos de secuencia de pares de extremo en formato fastq y devuelve como resultado un código binario y octal asociado a un espoligotipo específico. Se ejecutó con el parámetro de número mínimo de aciertos sin errores para respaldar la presencia de un espaciador igual a 5 -MIN (valor predeterminado). La asignación de familia y código SIT (Spoligo International Type) de las cepas secuenciadas se llevó a cabo a partir del código octal resultante mediante la base de datos TB-insight (<https://tbinsight.cs.rpi.edu>) en la opción SPOTCLUST (https://tbinsight.cs.rpi.edu/run_spotclust.html) utilizando el modelo SpolDB3-based y el sistema en línea SITVIT2 (<http://www.pasteur-guadeloupe.fr:8081/SITVIT2/description.jsp>). Los espoligotipos que no fue posible identificar en el portal se consideraron como cepas huérfanas (Rodwell et al., 2012).

Para la asignación de linaje y sublinaje se utilizó el programa MTBSeq versión 1.0.3 (Kohl et al., 2018). Se realizó la detección de SNPs para clasificación filogenética a partir del mapeo de las lecturas de secuenciación filtradas contra el genoma de referencia H37Rv (NC000962.3). Las variantes alélicas se basaron en los siguientes parámetros: cuatro lecturas “forward” y cuatro “reverse” para indicar la presencia de un alelo, valor de phred mayor a 20 y frecuencia alélica mayor a 75%. Para la construcción del alineamiento los genes pertenecientes a la familia PE, PPE y PGRS, además de los genes involucrados en la resistencia a antibióticos, fueron excluidos. A partir del alineamiento, se construyó la inferencia filogenética

con el criterio de máxima verosimilitud (ML) mediante el paquete IQ-TREE (Nguyen et al., 2015) con un modelo general de tiempo reversible (GTR) de sustitución de nucleótidos y una distribución gamma utilizando 1000 bootstrap. La filogenia resultante fue visualizada mediante iTOL (Letunic & Bork, 2016).

Determinación genotípica de resistencia a antibióticos

Para el análisis de fármaco-resistencia *in silico* se utilizó la herramienta TB-profiler v3.0.4 (Coll et al., 2015). Este programa puede determinar la resistencia genotípica a antibióticos mediante el alineamiento contra el genoma de referencia para identificar SNPs previamente asociados a resistencia fenotípica a partir de una base de datos. La versión 3.0.4 de TB-profiler utiliza un total de 1541 SNPs para la identificación de los perfiles de resistencia.

Identificación de macro y microvariantes

Para la identificación de macrovariantes o rearrreglos genómicos en comparación con el genoma de referencia H37Rv, se utilizó la herramienta MAUVE (Darling et al., 2004) manteniendo los valores predeterminados para su análisis. La identificación de microvariantes específicas de las cepas aisladas en Michoacán se llevó a cabo a partir de los archivos de salida de MTBseq, en donde se reportan los SNPs, inserciones y deleciones por cepa (.gatk_positions_variants_cf4.tab), utilizando un script en MySQL al comparar contra 490 cepas de *M. tuberculosis* aisladas en diferentes regiones geográficas (Anexo 1). Se concatenaron los archivos variants.tab y se filtraron las variantes asociadas a la clasificación filogenética y resistencia a antibióticos, además de aquellas con un valor de phred menor a 30. A partir de estos datos se identificaron: la inserción exclusiva más común de cepas MYC, la deleción exclusiva más común de cepas MYC y SNP exclusivo más común de cepas MYC.

Búsqueda de secuencias de inserción

Para la identificación de secuencias de inserción se utilizó el programa ISMAPPER (Hawkey et al., 2015), el cual realiza la determinación de ISs a partir de los archivos

de secuenciación (.fastq) sin procesar, es decir, sin necesidad de ensamblar el genoma. Como resultado se obtiene un archivo de texto que contiene el número de secuencias de inserción identificadas, su orientación, regiones flanqueantes anotadas y si se trata de una inserción nueva o conocida, respecto al genoma de referencia H37Rv. Adicionalmente, se comprobaron los resultados obtenidos mediante BLAST utilizando la base de datos ISFinder (Siguier et al., 2011) para la obtención de la secuencia de referencia de IS6110. Para la comprobación con BLAST primero se creó una base de datos utilizando los 10 genomas de las cepas MYC con makeblastdb, incluyendo la opción -dbtype nucl, posteriormente se utilizó blastn para comparar dicha base de datos contra la secuencia blanco (IS6110 en formato fasta), y finalmente se realizó el filtrado de los resultados en donde se tomaron en cuenta los siguientes parámetros: a) longitud del alineamiento (eliminar fragmentos cortos en relación al tamaño de la secuencia de inserción reportada), b) e-value (igual a 0), y c) porcentaje de identidad (99%).

Ensamblado y anotación genómica

Partiendo de las lecturas filtradas en formato fastq, se llevó a cabo el ensamblado de los genomas estudiados con el programa SPAdes, basado en diagramas de *Brujin* (Bankevich et al., 2012). El programa fue utilizado con los parámetros predeterminados, desactivando la opción de filtrado de secuencias, indicando los archivos de las lecturas *forward* y *reverse* y reduciendo el número de errores e *indels* cortos utilizando la opción -careful. La validación de los ensamblados realizados se llevó a cabo con el programa QUAST v4.6.0 (Gurevich et al., 2013). QUAST evalúa los ensamblados de los genomas, utilizando varias características métricas que incluyen el número de *contigs*, longitud total del genoma, % de GC, N50, L50, N75, L75, longitud de los *contigs*, entre otros, y puede evaluar la calidad del ensamblado incluso sin un genoma de referencia. A pesar de que no es necesario un genoma de referencia para este análisis, el utilizarlo da un soporte adicional a la calidad del ensamblado obtenido, por lo que se utilizó el genoma H37Rv para el análisis.

Con el objetivo de realizar una mejora de los ensamblados se realizó un filtrado de *contigs* con menos de 500 pb y un reordenamiento con ABACAS (Assefa

et al., 2009). Esta herramienta se basa en la utilización de un genoma de referencia para el ordenamiento y orientación de los contigs obtenidos en un ensamblado. Las bases indeterminadas se denominan con Ns en las secuencias obteniendo finalmente *scaffolds* representativos del genoma secuenciado.

La anotación de los genomas ensamblados se realizó mediante el servidor RAST (<https://rast.nmpdr.org>) (Aziz et al., 2008) seguido de una re-anotación con Prokka v1.10 (Seemann, 2014). Con este fin, se descargó el archivo que contiene las secuencias de aminoácidos del genoma de referencia H37Rv y el archivo .gff del mismo (RefSeq NC_000962.3).

Caracterización fenotípica

Morfología colonial

A partir de los cultivos en medio Lowenstein-Jensen se realizaron resiembras de las cepas seleccionadas en medio BBL Seven H11 adicionado con sangre humana, con la finalidad de obtener una cantidad mayor de colonias en un periodo menor de tiempo. La adición de sangre a medios de cultivo para micobacterias ha sido previamente reportada (Drancourt & Raoult, 2007; Tran et al., 2019). Para la preparación del medio de cultivo adicionado con sangre humana se utilizó una preparación de suplemento de enriquecimiento no comercial que contiene 2.5 g de albúmina bovina, 30 μ l de aceite de oliva extra virgen y 0.425 g de cloruro de sodio en 50 ml de agua destilada. Esta solución se esterilizó en autoclave a 121°C durante 10 min y posteriormente se dejó enfriar a 50°C para añadir 0.002 g de catalasa.

Se siguieron las instrucciones del fabricante para la preparación del medio base BBL Seven H11 sustituyendo 50 ml de suplemento OADC por 50 ml del suplemento de enriquecimiento preparado y se añadieron 25 ml de sangre humana. Los medios de cultivo se incubaron por 48 h a 37°C para descartar una contaminación. Finalmente fueron inoculados con las cepas de estudio y la cepa de referencia H37Rv e incubados a 37°C por un periodo de 4 a 8 semanas con aireación semanal en campana. Se realizó una revisión y reporte semanal del crecimiento y la morfología colonial.

Determinación del fenotipo formador de cuerdas

Se empleó la técnica de Ziehl-Neelsen para determinar la formación microscópica de cuerdas en las semanas 1, 2, 3, 4, 5, 6, 7 y 8 de incubación, partiendo de los cultivos en medio Lowenstein-Jensen y medio suplementado con sangre humana. Reportándose la presencia o ausencia de formación de cuerdas y morfología bacilar.

Determinación de resistencia fenotípica a antibióticos

La resistencia a fármacos de primera línea (isoniazida, rifampicina, estreptomina, etambutol y pirazinamida fue realizada en el Laboratorio Estatal de Salud Pública del Estado de México mediante el método automatizado BACTEC MGIT 960 (Becton Dickinson Diagnostic Instrument System, Inc.).

RESULTADOS

Caracterización genómica

Las características de las muestras de ADN obtenidas y secuenciadas se describen en la Tabla 2.

Tabla 2. Características del ADN extraído de las cepas seleccionadas.

Cepa	Volumen (μ l)	Concentración (ng/ μ l)	Concentración total (ng)	Pureza (260/280)
MYC36	19	57.78	1097.82	1.85
MYC52	29	85.71	2585.59	1.87
MYC61	19	86.65	1644.45	2.0
MYC72	27	150.78	4071.60	1.87
MYC75	44	196.50	8659.20	1.92
MYC78	37	69.51	2671.87	1.88
MYC87	24	212.19	5092.87	1.93
MYC93	21	74.22	1558.62	1.88
MYC96	22	139.35	3065.70	1.92
MYC136	29	200.25	5807.50	1.97

Los valores de la relación de absorbancia 260/280 muestran la pureza óptima del ADN obtenido según los estándares aceptados para su utilización en los protocolos de secuenciación genómica, en los cuales dicho valor debe oscilar entre 1.8 y 2.

Los resultados de la secuenciación después del filtrado de calidad de las lecturas se muestran en la Tabla 3.

Se observó que el valor más bajo del número de lecturas corresponde a la cepa MYC75 (2,245,009); sin embargo, se encuentra por encima de la media de lecturas de secuenciación reportada en la base de datos PATRIC (<https://www.patricbrc.org>), lo que garantiza una buena cobertura del genoma.

Tabla 3. Rendimiento de secuenciación por cepa. Se muestra el número total de secuencias posterior al análisis con TrimGalore.

Cepa	% Lecturas eliminadas	% Lecturas a utilizar	Número de lecturas después del filtrado
MYC36	1,67	98,33	2,229,972
MYC52	2,01	97,99	2,517,118
MYC61	1,68	98,32	2,417,474
MYC72	1,54	98,46	2,245,092
MYC75	1,68	98,32	2,376,254
MYC78	1,70	98,30	2,505,854
MYC87	2,29	97,71	2,577,326
MYC93	1,80	98,20	2,413,548
MYC96	2,32	97,68	2,429,252
MYC136	2,45	97,55	2,498,242

Tipificación *in silico*

Predicción de espoligotipo

Mediante el análisis con Spotyping se identificó al 80% de las cepas con un código octal previamente reportado en el sistema SITVIT (Tabla 4). Las cepas MYC072 y MYC078 fueron identificadas como “huérfanas”. De acuerdo a la clasificación de linaje y sublinaje con base en los espoligotipos, las cepas MYC36, MYC72 y MYC75 pertenecen a la familia LAM, las cepas MYC52, MYC78, MYC87, MYC93 y MYC136 pertenecen a la familia X1, la cepa MYC61 a la familia T y la cepa MYC96 a la familia S.

Tabla 4. Asignación de espoligotipo, SIT y familia por Spotyping. EPTB, tuberculosis extrapulmonar; PTB, tuberculosis pulmonar; SIT, spoligo international type; SC, sin coincidencia.

Cepa	Tipo TB	Localización	Código octal	SIT	Familia
MYC036	EPTB	Meníngea	777777607760771	42	LAM
MYC052	EPTB	Ganglionar	777776777760601	137	X2
MYC061	PTB	Pulmonar	777777777760771	53	T
MYC072	PTB	Pulmonar	777777607660471	SC	LAM
MYC075	PTB	Pulmonar	77777607760771	469	LAM1
MYC078	EPTB	Meníngea	757776377760771	SC	X1
MYC087	EPTB	Renal	777776777760771	119	X1
MYC093	PTB	Pulmonar	777776777760771	119	X1
MYC096	PTB	Pulmonar	77637777760751	1356	S
MYC136	EPTB	Digestiva	777716777760771	1826	X1

Determinación de linaje y sublinaje

Mediante el análisis filogenético fue posible ubicar a la totalidad de las cepas en el linaje 4 o Euroamericano. Se identificaron diversos sublinajes (Figura 5) siendo el más abundante el 4.1.1 (40%), seguido por el 4.3.4.2 en un 20% de las cepas. El clado 4.1.1 agrupó exclusivamente cepas extrapulmonares, las cuales representan el 80% de las cepas EPTB secuenciadas.

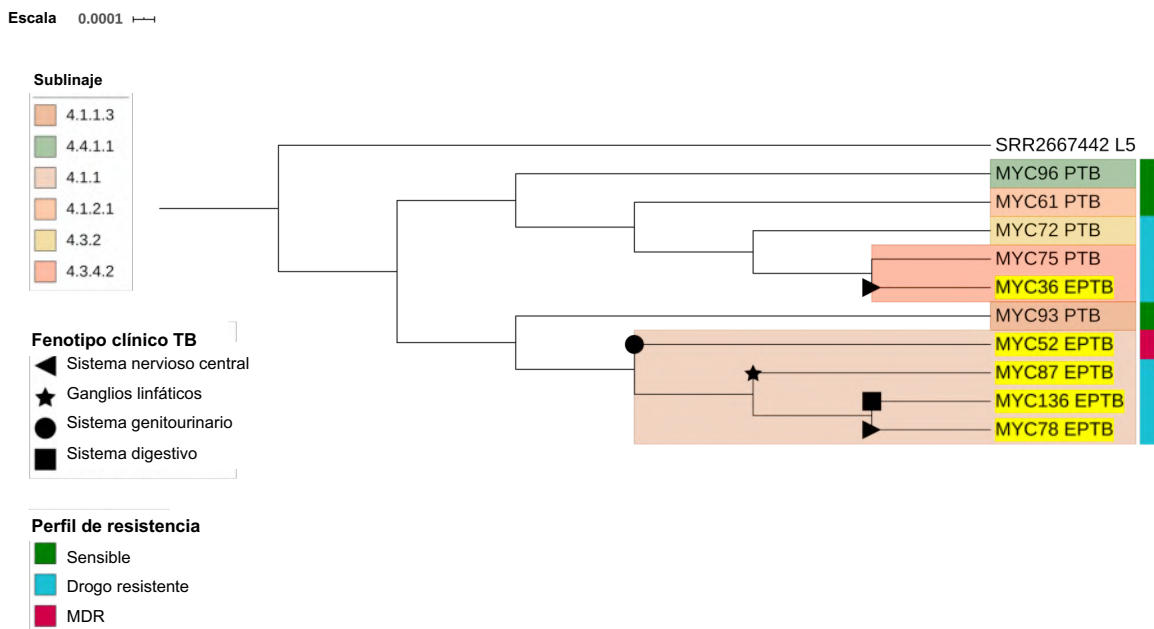


Figura 5. Clasificación filogenética de las cepas MYC. PTB, tuberculosis pulmonar; EPTB, tuberculosis extrapulmonar; MDR, tuberculosis multidrogo resistente. La cepa SRR2667442 perteneciente al linaje 5 fue utilizada como grupo externo.

En la Tabla 5 se muestra la comparación entre la asignación de familias por espoligotipo y sublinaje, se pueden observar discrepancias en los resultados reportados para las cepas MYC052, MYC061 y MYC075.

Determinación genotípica de resistencia a antibióticos

El análisis con TB-profiler permitió identificar a 8 de 10 cepas como sensibles, a la cepa MYC72 como resistente a etambutol y a la cepa MYC como mutlidrogo resistente (MDR) (Tabla 6).

Tabla 5. Comparativo de la asignación de familias por MTBSeq y Spotyping.

Cepa	Linaje	Sublinaje	Familia MTBSeq	Familia Spotyping
MYC036	L4	4.3.4.2	LAM	LAM
MYC052	L4	4.1.1	LAM	X2
MYC061	L4	4.1.2.1	Harleem	T
MYC072	L4	4.3.2	LAM	LAM
MYC075	L4	4.3.4.2	X	LAM1
MYC078	L4	4.1.1	X	X1
MYC087	L4	4.1.1	X	X1
MYC093	L4	4.1.1.3	X	X1
MYC096	L4	4.4.1.1	S	S
MYC136	L4	4.1.1	X	X1

Tabla 6. Resistencia genotípica a fármacos de primera línea. R, resistente; SENS, sensible.

Cepa	Rifampicina	Isoniacida	Pirazinamida	Etambutol	Estreptomicina
MYC036	SENS	SENS	SENS	SENS	SENS
MYC052	R	R	R	SENS	R
MYC061	SENS	SENS	SENS	SENS	SENS
MYC072	SENS	SENS	SENS	R	SENS
MYC075	SENS	SENS	SENS	SENS	SENS
MYC078	SENS	SENS	SENS	SENS	SENS
MYC087	SENS	SENS	SENS	SENS	SENS
MYC093	SENS	SENS	SENS	SENS	SENS
MYC096	SENS	SENS	SENS	SENS	SENS
MYC136	SENS	SENS	SENS	SENS	SENS

Las mutaciones identificadas fueron *embB* Val369Ala para resistencia a etambutol en MYC072, *rpoB* Asp435Tyr para resistencia a rifampicina, *rrs* 514 a>c para resistencia a estreptomicina, *katG* Ser315Thr para resistencia a isoniazida y *pncA* Trp68Leu para resistencia a pirazinamida en MYC052.

Determinación de macro y microvariantes

Los rea-reglos genómicos identificados con MAUVE se muestran en la Figura 6. MAUVE identifica regiones conservadas que aparentan no haber sido alteradas por los reordenamientos e inversiones dentro del genoma. Inicialmente identifica

secuencias sucesivas con similitud exacta compartida por dos o más genomas y, junto a una matriz de distancia basada en la conservación genómica, construye un árbol guía. Posteriormente, selecciona subconjuntos de estas secuencias para hacer el anclaje y realizar un alineamiento usando el algoritmo de alineamientos locales MUSCLE. Estas regiones conservadas se agrupan en bloques locales de linealidad (*locally colinear blocks*, LCB) y representan una secuencia compartida por dos o más genomas incluidos dentro del alineamiento. Es decir, los bloques coloreados representan los segmentos de ADN que coinciden entre el genoma de referencia H37Rv y cada una de las cepas MYC. Se identificaron 351 bloques de linealidad presentes en todos los genomas, 3,983 LCBs están presentes en una proporción de 90 al 10% de las cepas y 2,228 LCBs están presentes en al menos un genoma de las cepas MYC, pero ausentes en la de referencia. De estos se observaron 9 regiones conservadas únicas presentes en la totalidad de las cepas MYC (Tabla 7). Además, se identificaron 9 inversiones que oscilan entre 99 a 70 631 pares de bases (pb) en 4 de las 10 cepas MYC. En la cepa MYC052 se observó una inversión en la posición 1,218,405 con una longitud de 35,996 pb; mientras que en la cepa MYC072 se observaron dos inversiones de 202 y 70,632 pb en las posiciones 2,042,321 y 4,131,715, respectivamente. La cepa MYC087 presentó la mayor cantidad de reordenamientos de este tipo, en las posiciones 3,230,693, 1,146,913, 2,785,313, 3,857,783 y 3,865,042, con longitudes de 36,576, 99, 21,365, 290, y 15,168 pb. El análisis de los genes presentes en estas regiones permitió identificar 71 genes involucrados, pertenecientes principalmente a la familia PE, PPE, PGRS y a proteínas hipotéticas. A partir del análisis de los datos de las regiones invertidas se observó que los genes *pe_pgrs18* en la cepa MYC093 y *ppe34* en las cepas MYC052, MYC087 presentan interrupciones en su secuencia debido a los reordenamientos cromosómicos.

Tabla 7. Características de las regiones exclusivas identificadas en los genomas de las cepas MYC. LCB, bloques locales de linealidad; bp, pares de bases; RNC, región no codificante.

	LCB posición en el genoma								
	1	2	3	4	5	6	7	8	9
MYC036	201511	794961	943664	1846358	3042606	3043023	3043942	3545788	3771392
MYC052	206306	798508	947751	1871232	3070220	3070408	3070808	3573607	3798892
MC061	116088	708277	856981	1758722	2957386	2957517	2957912	3461919	3686606
MYC072	194278	785940	935058	1834134	3033691	3033822	3034169	3537806	3763276
MYC075	201177	794195	944112	1842732	3037960	3038148	3038548	3541399	3767097
MYC078	201175	794735	944522	1860739	3064833	3065021	3065421	3568848	3794234
MY087	201553	793556	942300	1847924	3048742	3048930	3049330	3557890	3783276
MYC093	201425	811086	960644	1869964	3098701	3098818	3099218	3602491	3827978
MYC096	208374	802768	951920	1859426	3063256	3063387	3063787	3569801	3792807
MYC136	201243	794457	943382	1859516	3067545	3067733	3068133	3571753	3797297
Longitud (pb)	183	536	123	618	115	344	863	97	256
Proteína	RNC	Hipotética	Hipotética	RNC	RNC	Hipotética	Hipotética	fadD2	Hipotética

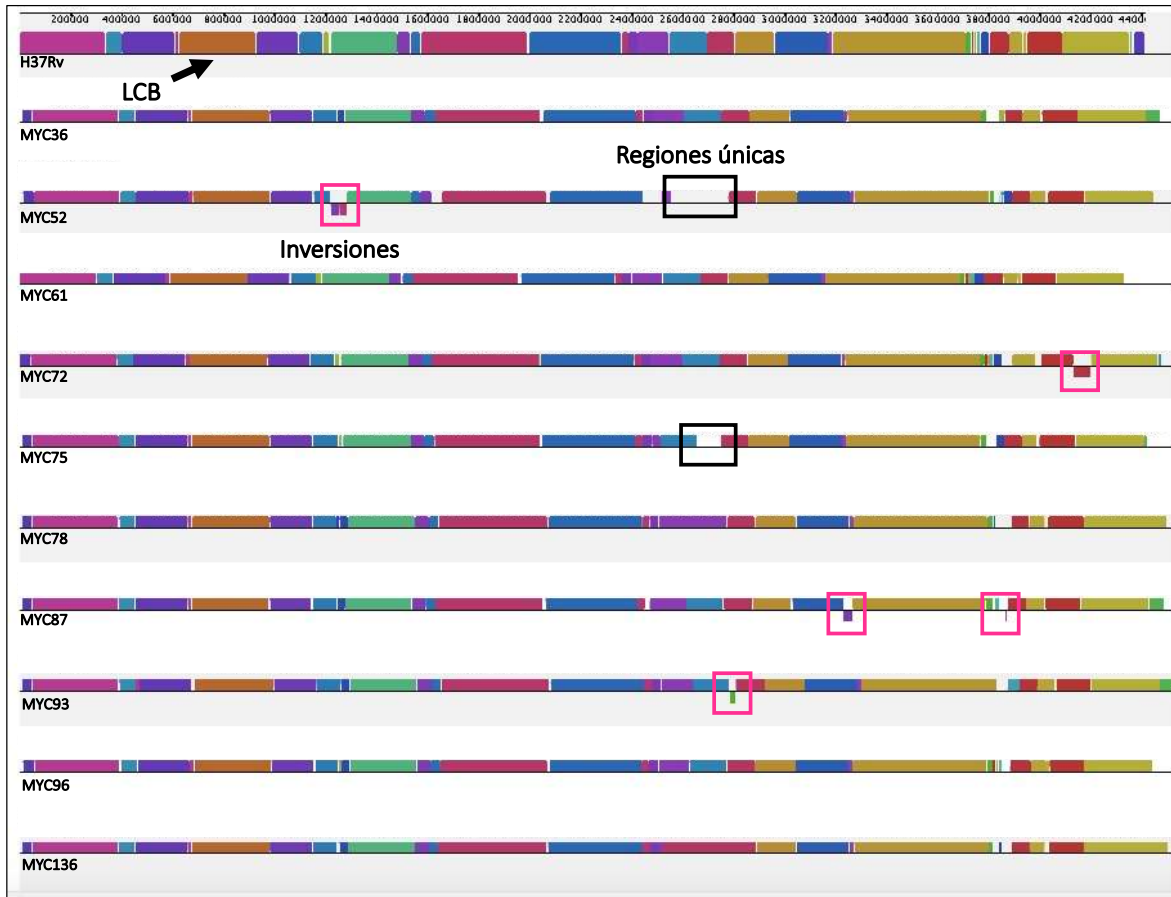


Figura 6. Rearreglos genómicos en las cepas MYC. Cada bloque coloreado representa secuencias homólogas entre los genomas de las cepas comparadas. LCB, bloques locales de linealidad.

El análisis para la búsqueda de microvariantes demostró la presencia de 1,228 a 1,646 mutaciones en las cepas MYC (Tabla 8), siendo la cepa MYC078 la que presentó una mayor cantidad de estas. Al realizar el análisis comparativo contra las cepas obtenidas de bases de datos públicas (n=490), se observó un total de 2,136 mutaciones presentes exclusivamente en las cepas aisladas de Michoacán, de estas 1,062 son deleciones, 37 son inserciones y 1,107 son SNPs no sinónimos. Aunque se observaron una gran cantidad de mutaciones exclusivas de las cepas MYC, no fue posible identificar variaciones exclusivas presentes en el 100% de las cepas. Sin embargo, se observó una mutación en el gen *priA* (sustitución de una adenina por una timina en la posición 217) en el 50% de los genomas. Además, se demostró que el 9.7% (140) de los SNPs exclusivos de cepas MYC están presentes en el 40% de las cepas y el 39.44% se encuentran en regiones intergénicas.

Tabla 8. Microvariantes por cepa identificadas mediante el mapeo contra el genoma de referencia.

Cepa	Número de mutaciones identificadas
MYC036	1230
MYC052	1412
MC061	1398
MYC072	1301
MYC075	1228
MYC078	1646
MY087	1540
MYC093	1644
MYC096	1388
MYC136	1611

Búsqueda de secuencias de inserción

Mediante el análisis con ISMapper se observó un inusual bajo número de copias de la secuencia de inserción IS6110 en relación con el linaje al que pertenecen (Tabla 9). Las cepas fueron clasificadas de acuerdo con el número de IS6110 como: cero (sin copias), bajo (≤ 6) y alto (≥ 7). Con la finalidad de demostrar si el fenómeno observado es propio de las cepas aisladas en el estado, se realizó un análisis utilizando 79 cepas de MTB secuenciadas en Veracruz (n=78) y Jalisco (n=1) (Anexo 2). Éstas fueron obtenidas de la base de datos SRA del NCBI. Después de la revisión de calidad de las lecturas con FASTQC, fueron sometidas a un análisis con ISMAPPER, observándose un 4.4% de cepas cero (sin copias), 45.6% de cepas del grupo bajo (≤ 6) y 50% del grupo alto (≥ 7). Posteriormente se realizó una clasificación filogenética utilizando los mismos parámetros usados para las cepas MYC (Figura 7).

En este conjunto de datos se encontraron 16 diferentes sublinajes. El 28.8% de las cepas pertenecen al sublinaje 4.1.1.3, seguido del sublinaje 4.1.2.1 en un 26.6%. También fue posible observar la presencia de 2 cepas del linaje 1, 2 cepas del linaje 3 y 4 del linaje 2 o Beijing, genotipos no predominantes en esta región geográfica. El 71% de las cepas agrupadas con un bajo número de IS6110 pertenecen al sublinaje 4.1.1.3. Por otro lado, las cepas que no contienen la

secuencia IS6110 se encuentran en los sublinajes 4.7, 4.3.4.1, 4.3.4.2 y 4.1.1.3 y representan 3 cepas MYC y 1 aislada del estado de Jalisco.

Al analizar los datos respecto a la interrupción génica ocasionada por la inserción de la secuencia IS6110 se determinó la presencia de una doble disrupción de los genes *Rv0402c-Rv0403c* y *Rv2813c-Rv2816c* en las cepas pertenecientes al sublinaje 4.1.1.3, a excepción de la cepa MYC093 y tres cepas aisladas del estado de Veracruz (Figura 8). Debido a los datos encontrados se realizó un análisis estadístico utilizando chi-cuadrada con significancia estadística $p < 0.05$, encontrándose una asociación estadísticamente significativa ($p < 0.002$) entre la disrupción de los genes *Rv0402-Rv0403c* y *Rv2813-Rv2816c* con la familia X3 o 4.1.1.3, que presenta un bajo número de copias de IS6110.

Tabla 9. Secuencias de inserción IS6110 encontradas en los genomas de las cepas MYC.

Cepa	Sublinaje	Núm. De IS6110	Clasificación	Posición en el genoma	Genes flanqueantes
MYC36	4.3.4.2	0	Cero	NA	NA
MYC52	4.3.4.2	1	Bajo	3480188-3480568	Rv0835-Rv0836c
MYC61	4.1.2.1	3	Bajo	2039011-2039000	Rv0402c- Rv0403c
MYC72	4.3.3	2	Bajo	947544-947525	Rv0405
MYC75	4.1.1.1	1	Bajo	3122063—1	Rv3924c- Rv2816c
MYC78	4.1.1	0	Cero	NA	NA
MYC87	4.1.1	1	Bajo	2039011-2039000	Rv0402c- Rv0403c
MYC93	4.1.1.3	1	Bajo	483360—1	Rv3924c- Rv0403c
MYC96	4.4.1.1	1	Bajo	3122063—1	Rv3924c- Rv2816c
MYC136	4.1.1	1	Bajo	2039011-2039000	Rv0402c- Rv0403c

NA, No aplica.

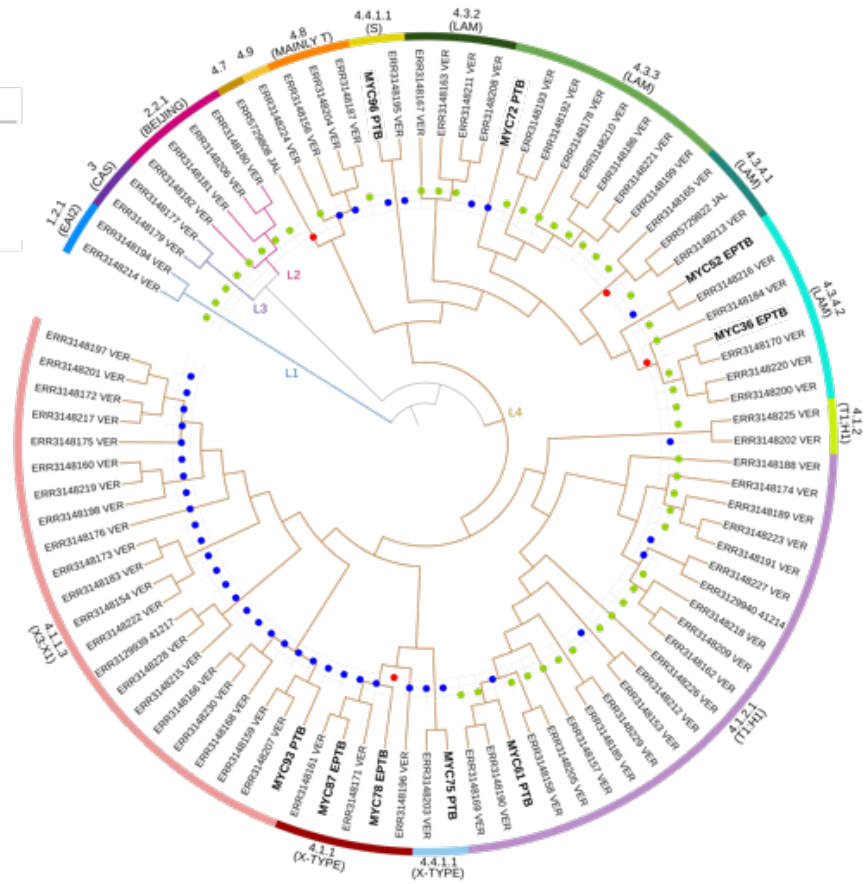
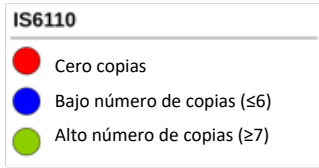


Figura 7. Filogenia de las cepas de *M. tuberculosis* aisladas en México. Las cepas MYC se encuentran resaltadas en negritas. La filogenia fue construida utilizando el criterio de máxima verosimilitud con un total de 1000 bootstrap.

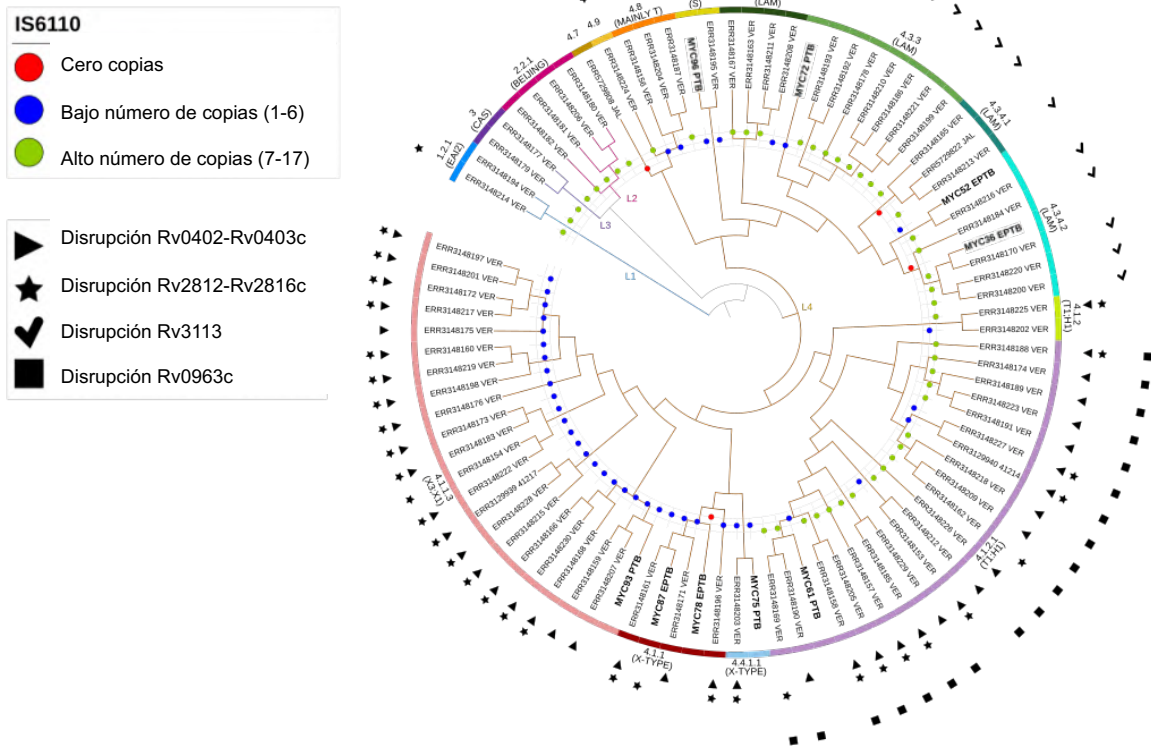


Figura 8. Disrupción génica ocasionada por la inserción de IS6110 en los genomas de cepas de MTB en México. Las cepas MYC se resaltan en negritas.

Adicionalmente, se encontró una asociación estadísticamente significativa ($p < 0.01$) entre la disrupción de los genes *Rv0402-Rv0403c* y *Rv0963c* con la familia T1;H1, que presenta alto número de copias de IS6110. Es de resaltar que, aunque las cepas MYC se agrupan en clados, con marcada tendencia a ciertas características genóticas, no comparten estas variaciones presentes en las cepas de la familia a la que pertenecen.

Ensamblado y anotación

Las características de los ensamblados obtenidos con SPAdes se muestran en la Tabla 10. Se obtuvo una cobertura mayor a 50x en el 100% de las cepas, recomendada para esta especie bacteriana, y un número aceptable de contigs. Para bacterias con un genoma de aproximadamente 4Mb, ensamblados con un valor ≤ 200 contigs se consideran aceptables. Los valores de longitud del genoma

(4341393–4529861 pb) y N50 aseguraron además una calidad adecuada de los ensamblados construidos.

Tabla 10. Características de los ensamblados obtenidos para cada una de las cepas MYC.

Cepa	Longitud del genoma	Contigs	Contenido de GC	N50	Cobertura
MYC036	4473036	81	65.5	99982	80x
MYC052	4504014	83	65.6	111037	86x
MYC061	4341393	89	65.5	101777	91x
MYC072	4519365	88	65.5	104288	86x
MYC075	4468615	86	65.5	99014	87x
MYC078	4499176	75	65.5	123695	95x
MYC087	4486405	83	65.5	113871	86x
MYC093	4529861	74	65.5	119086	84x
MYC096	4499957	76	65.5	106172	86x
MYC136	4501126	79	65.5	125044	93x

Las anotaciones realizadas con Prokka (Tabla 11) mostraron un rango que osciló entre los 4,243 (cepas MYC072 y MY096) a 4,087 genes (cepa MYC061). La cepa MYC052 aislada de sistema genitourinario es la que presentó mayor cantidad de regiones repetidas (n=5), mientras que las cepas MYC072 (pulmonar) y MYC087 (linfática) presentaron menos regiones repetidas en su genoma.

Los resultados de la anotación con RAST se muestran en la Figura 9. El contenido de genes se agrupa en varios subsistemas asociados a una función específica, lo que da una idea de la diferencia en el contenido de genes entre las cepas MYC. Alrededor del 25 al 28% de los genes de las cepas MYC no se agrupan en ningún subsistema y están anotadas principalmente como proteínas hipotéticas.

Tabla 11. Resultados de la anotación en Prokka de los genomas de las cepas MYC.

Cepa	CDS	Genes	misc_rna	rRNA	Regiones repetidas	tRNA	tmRNA
MYC036	4126	4219	37	3	3	52	1
MYC052	4027	4110	33	3	5	49	1
MYC061	3997	4087	36	3	4	50	1
MYC072	4151	4243	36	3	2	52	1
MYC075	4027	4113	34	3	4	48	1
MYC078	4009	4097	34	3	5	50	1
MYC087	4090	4183	36	3	2	53	1
MYC093	4120	4213	37	3	3	52	1
MYC096	4140	4243	35	3	3	53	1
MYC136	4019	4107	34	3	4	50	1

Las diferencias más notorias en la distribución de los genes se encontraron en las categorías de metabolismo de carbohidratos, aminoácidos y sus derivados, ácidos grasos, lípidos e isoprenoides, y en el metabolismo de proteínas.

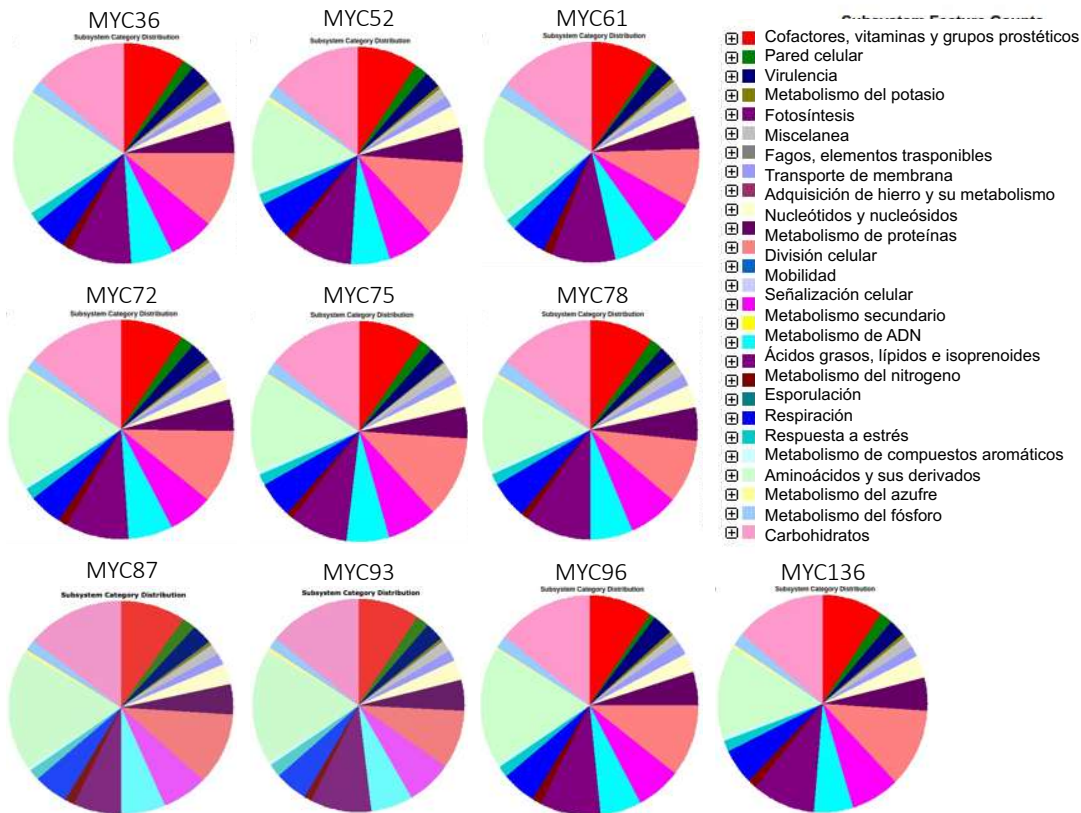


Figura 9. Comparación del contenido de genes basado en la clasificación funcional asignada por RAST para las cepas MYC.

Caracterización fenotípica

Morfología colonial

A partir de los cultivos en medio Lowenstein-Jensen y medio adicionado con sangre humana se determinó la morfología colonial de las cepas MYC, encontrándose 2 morfotipos, uno en el que se presentan colonias grandes, elevadas, rugosas y quebradizas (Figura 10a) y otro en donde se presentan colonias pequeñas, planas e irregulares (Figura 10b). Las cepas que presentaron el morfotipo 1 fueron MYC072, MYC075, MYC078, MYC087 y MYC136, las cepas en las que se observó el morfotipo 2 fueron MYC036, MYC052, MYC061, MYC093 y MYC096. Adicionalmente, se observaron diferencias en la velocidad de crecimiento, en las cepas que presentaron el morfotipo 1 el tiempo de crecimiento fue mayor a 4 semanas, mientras que en el morfotipo 2 el crecimiento fue visible a partir de los 5 días de incubación.

No se observaron diferencias en el morfotipo o velocidad de crecimiento entre los grupos pulmonar y extrapulmonar.

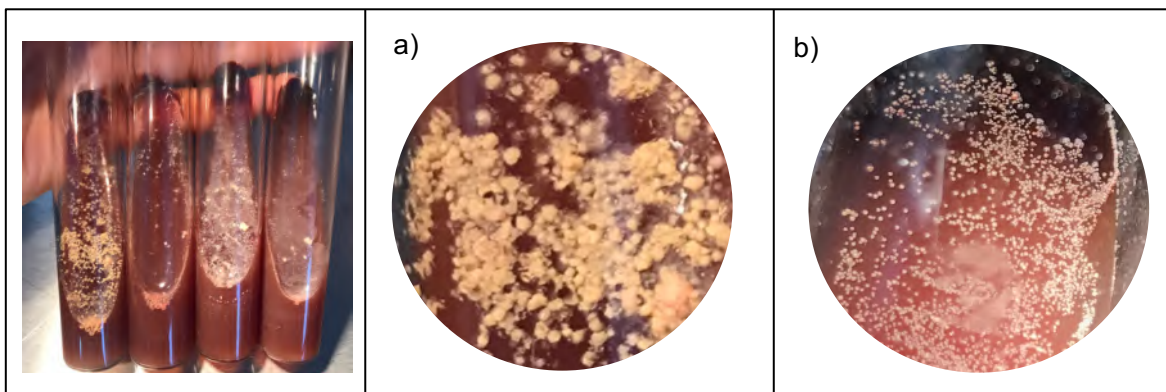


Figura 10. Morfología colonial de las cepas MYC en agar BBL Seven H11 adicionado con sangre humana. a) morfotipo 1, colonias grandes, rugosas y quebradizas; b) morfotipo 2, colonias pequeñas, planas e irregulares.

Determinación del genotipo formador de cuerdas

Mediante la tinción diferencial de Zielh-Neelsen se observó la formación de cuerdas características en las cepas MYC036, MYC052, MYC061, MYC093 y MYC096 (Figura 11). Por otro lado, se determinó la presencia de agrupamientos circulares en las cepas MYC072, MYC078 y MYC087 y la presencia de bacilos sin agrupamientos en MYC075. Se observó que las cepas formadoras de cuerdas corresponden al morfotipo 2 de colonias pequeñas y mayor velocidad de crecimiento, mientras que las cepas que formaron agrupaciones circulares presentan colonias grandes de crecimiento lento. No fue posible encontrar una asociación entre el fenotipo clínico (PTB/EPTB) y la formación de cuerdas microscópicas.

Determinación de resistencia genotípica

Los resultados de la determinación del perfil de resistencia fenotípica a antibióticos de las cepas MYC se presentan en la Tabla 12. Al comparar la resistencia genotípica con la resistencia fenotípica reportada se encontraron múltiples inconsistencias en las cepas. Así, la cepa MYC036 que está predicha genotípicamente como sensible, presenta resistencia fenotípica a pirazinamida y estreptomina, MYC072 que

genotípicamente presenta resistencia a etambutol, es fenotípicamente resistente a rifampicina. Las cepas MYC075, MYC078, MYC087 y MYC136 que genotípicamente se caracterizaron como sensibles, fenotípicamente son resistentes a isoniazida, estreptomicina, isoniazida y pirazinamida y estreptomicina, respectivamente.

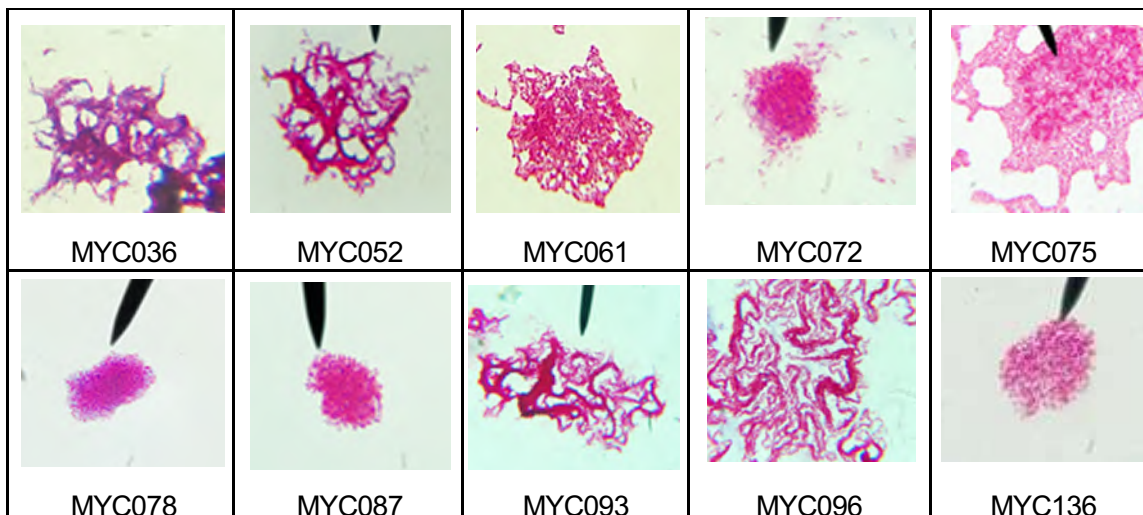


Figura 11. Formación de cuerdas microscópicas en las cepas MYC. Observables mediante tinción diferencial con Ziehl-Neelsen. Objetivo 100x.

Tabla 12. Resistencia fenotípica a antibióticos de primera línea en las cepas MYC. SENS, sensible; R, resistente.

Cepa	Rifampicina	Isoniacida	Pirazinamida	Etambutol	Estreptomicina
MYC036	SENS	SENS	R	SENS	R
MYC052	R	R	R	SENS	R
MYC061	SENS	SENS	SENS	SENS	SENS
MYC072	R	SENS	SENS	SENS	SENS
MUY075	SENS	R	R	SENS	SENS
MYC078	SENS	SENS	SENS	SENS	R
MYC087	SENS	R	R	SENS	SENS
MYC093	SENS	SENS	SENS	SENS	SENS
MYC096	SENS	SENS	SENS	SENS	SENS
MYC136	SENS	SENS	SENS	SENS	R

DISCUSIÓN

La caracterización genómica de las cepas MYC permitió identificar su pertenencia al linaje 4 o Euroamericano. Estudios de epidemiología molecular previos han demostrado una gran diversidad de linajes circulantes en México, siendo el linaje 4 el que se reporta con mayor frecuencia (Juaréz-Eusebio et al., 2017). Dicho linaje es genéticamente diverso y filogeográficamente estructurado, en el cual se han identificado sublinajes especialistas y generalistas. Los sublinajes generalistas como LAM y Harleem se han aislado en todo el mundo, mientras que los sublinajes especialistas mantienen una marcada restricción geográfica. Las cepas MYC036, MYC052 y MYC072 se identificaron como pertenecientes a la familia LAM, mientras que la cepa MY061 pertenece a la familia Harleem, es decir el 40% de las cepas pertenecieron a los sublinajes generalistas, de estas el 75% son cepas extrapulmonares. Las cepas generalistas del linaje 4 se caracterizan por tener una gran diversidad de epítopes para células T altamente conservados, por lo que pueden interactuar con una gama más amplia de alelos del antígeno leucocitario humano (Perdigão et al., 2017). Dentro de la familia LAM existen subfamilias, siendo LAM9 y LAM3 las más distribuidas en México. En la clasificación filogenética realizada en este trabajo solo fue posible identificar una cepa perteneciente a la subfamilia LAM1 (MYC075), que corresponde a un aislado pulmonar. En un estudio en el que se analizaron aislados pertenecientes a esta familia, se observó una diferencia significativa en la velocidad de crecimiento en medio de cultivo, característica relacionada con la capacidad de reproducción bacteriana. Esto genera una mayor carga bacilar, lo que podría promover una mayor dispersión de la familia LAM (Von Groll et al., 2010), por lo que se esperaría asociarla con mayor frecuencia en cepas pulmonares, en comparación con cepas extrapulmonares.

Por otro lado, la cepa MYC093 pertenece al sublinaje 4.1.1.3 o familia X3, el cual ha sido previamente reportado como exclusivo de México y se ha asociado a multidrogo-resistencia (Jiménez-Ruano et al., 2021). Sin embargo, esta cepa no presentó resistencia fenotípica ni genotípica a antibióticos, además de carecer de la doble disrupción génica de *Rv0402c-Rv0403c* y *Rv2813c-Rv2816c* asociada a la

familia a la que pertenece y a un número bajo de copias de secuencia de inserción IS6110. Anteriormente se ha demostrado la existencia de patrones de números de copia específicos del linaje (Roychowdhury et al., 2015), en el que los sublinajes Harleem, S y LAM se han asociado a un alto número de copias, mientras que el sublinaje X presenta un bajo número de copias. En este trabajo se identificó una diferencia entre el número de secuencias de inserción IS6110 esperadas según el sublinaje y lo observado para las cepas MY036 (LAM), MYC052(LAM), MYC061(Harleem), MYC072(LAM) y MYC096 (S). Aunque se cree que un alto número de copias de IS6110 proporciona una ventaja selectiva a las cepas que los poseen, la resistencia antibióticos y la aparición de brotes también se han asociado a cepas con un bajo número de copias (McEvoy et al., 2007). Interesantemente, las cepas MYC con menos de 6 copias de IS6110 presentan resistencia fenotípica a al menos un antibiótico de primera línea.

La importancia de la distribución preferencial de las secuencias IS6110 en ciertas posiciones del genoma de MTB se resalta por las posibles consecuencias funcionales de su inserción. Estas pueden ocasionar la inactivación de un gen si se insertan en una región codificante, la delección de las primeras bases y aparición de mutaciones puntuales si se integran al inicio de la secuencia del gen, la inversión o delección del gen completo si ocurre recombinación entre dos secuencias IS6110, o diferencias en la expresión génica si se inserta en la región promotora del gen (McEvoy et al., 2007). En el caso de las cepas MYC, la inserción de IS6110 en el genoma ocasionó la interrupción de *Rv2813-Rv2816c* en MYC096 perteneciente al sublinaje 4.4.1.1 y MYC087 del sublinaje 4.1.1. Estos genes pertenecen a la región de repetidos directos, en donde se insertan comúnmente fagos y secuencias de inserción y codifican para proteínas hipotéticas cuyas mutaciones se han asociado con alteraciones en la producción de citocinas y la respuesta de polimorfonucleares, confiriéndole a las cepas que los poseen ventajas en la transmisión de la enfermedad (Nebenzahl-Guimaraes et al., 2017). Por otro lado, la inserción de una copia de IS6110 en la secuencia del gen *plcD* se ha relacionado con la forma extrapulmonar de la TB (Yang et al., 2005), pero en este trabajo no se identificaron inserciones en este locus específico.

Otros genes en los que si se observó interrupción de la secuencia son *Rv0402c* y *Rv0403c* que codifican para dos proteínas de membrana implicadas en el metabolismo de ácidos grasos (*mmpL1* y *mmpS1*). Una interrupción en la secuencia de dichos genes podría ocasionar una alteración en la vía de síntesis de dimicolato de trehalosa, glicolípido asociado previamente con la formación de cuerdas microscópicas (Harland et al., 2008). El factor cuerda o dimicolato de trehalosa encargado de la formación de estas estructuras se relaciona con la inducción de secreción de las citocinas IL-1b, IL-6 y TNF-a, es clave en la aparición del tipo cavitario de TB y su metabolismo se ve implicado en el desarrollo de mutaciones asociadas a resistencia a antibióticos (Lee et al., 2019). Por lo que la interrupción de los genes *Rv0402* y *Rv0403c* podría estar jugando un papel importante en la transmisión local del genotipo 4.1.1.3 o X3.

Cabe resaltar que la cepa MYC093, es la única del clado X3 que no posee ninguna interrupción de los genes antes mencionados, y esto coincide con la formación de cuerdas macróscopias durante el cultivo. Este comportamiento “atípico” de las cepas aisladas en Michoacán ya ha sido observado previamente por nuestro grupo de trabajo, en diversos análisis de secuencias de genes involucrados en la vía de síntesis de ácidos micólicos, de resistencia a antibióticos y de la familia *pe*, *ppe* y *pgrs*, encontrándose un inusual alto número de mutaciones en comparación con cepas aisladas de otras regiones geográficas (Frutos-Hernandez, 2015; Negrete-Paz, 2018). Estas inconsistencias también se observan en la comparación genotípica-fenotípica de resistencia a antibióticos de las cepas MYC036, MYC072, MYC075, MYC078, MYC087 y MYC136. En ellas se puede ver un fenotipo resistente, pero no existe ninguna mutación en genes asociados a resistencia a antibióticos. Debido a que en MTB este es el principal mecanismo de resistencia adquirida, dicho fenómeno es poco común. Generalmente, las discrepancias entre la resistencia predicha y observada están más relacionadas con una baja especificidad y sensibilidad de las herramientas utilizadas para la determinación genotípica de resistencia que con una característica fisiológica de la bacteria. Sin embargo, los pocos estudios en donde se reporta un fenómeno común al observado en este trabajo (Zenteno-Cuevas et al., 2020), sugieren la presencia

de mecanismos de resistencia intrínseca no descritos para los fármacos de primera línea. Por ejemplo, la presencia de mutaciones en genes relacionados con la permeabilidad celular, como *fadD2* que confiere resistencia a pirazinamida, o *lipX* que se ha asociado a resistencia a rifampicina (Singh et al., 2016), o con bombas de eflujo que modifican el transporte de antibióticos hacia el espacio intracelular.

Los fenotipos característicos de estas cepas MYC pueden estar determinados por la presencia de ciertas variaciones en su genoma. El genoma de referencia H37Rv posee un total de 4 452 489 pb. En comparación con las cepas estudiadas, solo la cepa MYC61 es de una longitud menor (4 341 393 pb), ya que el resto de las cepas poseen de 16 126 a 77 372 pb más que el genoma de referencia. El menor tamaño del genoma de la cepa MYC61 se debe a la presencia de una mayor cantidad de deleciones, por el contrario, las cepas MYC36, MYC52, MYC72, MYC75, MYC78, MYC87, MYC93, MYC96 y MYC136 presentan inserciones en su genoma. Existe una presión evolutiva significativa dirigida hacia genomas bacterianos más pequeños, ya que los cromosomas más pequeños pueden replicarse más rápido. Si una población bacteriana se traslada a un entorno donde, por ejemplo, abunda un aminoácido esencial, algunos miembros de la población bacteriana, después de muchas generaciones, perderán la capacidad de sintetizar ese aminoácido. Eventualmente, cada miembro de la población también perderá por deleción la secuencia de ADN que codifica para la ruta necesaria en la producción de ese aminoácido. Sin embargo, este proceso ocurre extremadamente lento, ya que la presión evolutiva para disminuir el tamaño del genoma también necesita proteger al organismo de la extinción por la degeneración de su genoma (Arnold, 2007).

Tras el análisis para la búsqueda de macrovariantes, se identificaron 9 regiones homólogas presentes en la totalidad de las cepas MYC, pero ausentes en la cepa de referencia H37Rv. De estas, 5 incluyen genes que codifican para proteínas hipotéticas, 3 corresponden a regiones no codificantes y 1 a una variación dentro del gen *fadD2*. Este gen es un mediador de la resistencia intrínseca a pirazinamida que codifica una ligasa de acetil coenzima A, cuya actividad es necesaria para el mantenimiento de una reserva intracelular de acil-CoA de cadena

larga suficiente para procesos metabólicos como la β -oxidación y la síntesis de triacilglicerol o la síntesis de lípidos de membrana (Rosen et al., 2017).

En el caso de las microvariantes (SNPs, inserciones y deleciones), se identificaron 2,136 mutaciones exclusivas de las cepas MYC, aunque no se encontró ninguna presente en todas si se determinó una variación exclusiva presente en el 50% de las cepas (MYC052, MYC078, MYC087, MYC093 Y MYC136). Esta mutación ocurre en el gen *priA* en la posición 217, ocasionando una sustitución de una adenina por una timina. El gen *priA* es considerado como esencial y está involucrado en la iniciación de la replicación del ADN (Reiche et al., 2017). La presencia de elementos en el genoma que puedan ocasionar modificaciones en la replicación o reparación del ADN como esta mutación en el gen *priA*, pueden ser la causa del aumento en el número elevado de polimorfismos en los genes *pe*, *ppe*, *pgrs* y de la vía de síntesis de ácidos micólicos característico de las cepas MYC.

CONCLUSIÓN

Las cepas MYC poseen variaciones genéticas no presentes en otros genotipos MTB circulantes, lo que les confiere características propias que podrían explicar los fenotipos observados, como la composición lipídica de la pared celular y la presencia de una gran cantidad de polimorfismos en su genoma; sin embargo, es necesaria la realización de experimentos *in vivo/in vitro* para confirmar la asociación de dichas variaciones con los fenotipos.

REFERENCIAS

- Arora, J., Suresh, N., Porwal, C., Pandey, P., Pande, J. N., & Singh, U. B. 2020. Genotyping *Mycobacterium tuberculosis* isolates with few copies of IS6110: Value of additional genetic markers. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*, (81): 104230. <https://doi.org/10.1016/j.meegid.2020.104230>
- Arnold, C. 2007. Molecular evolution of *Mycobacterium tuberculosis*. *Clinical microbiology and infection*, 13(2): 120-128. <https://doi.org/10.1111/j.1469-0691.2006.01637.x>
- Assefa, S., Keane, T. M., Otto, T. D., Newbold, C., & Berriman, M. 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics*, 25(15): 1968–1969. <https://doi.org/10.1093/bioinformatics/btp347>
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Zagnitko, O. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC genomics*, 9(75): 1-15. <https://doi.org/10.1186/1471-2164-9-75>
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology: a journal of computational molecular cell biology*, 19(5): 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Brites, D., & Gagneux, S. 2015. Co-evolution of *Mycobacterium tuberculosis* and *Homo sapiens*. *Immunological reviews*, 264(1): 6–24. <https://doi.org/10.1111/imr.12264>
- Brudey, K., Driscoll, J. R., Rigouts, L., Prodinger, W. M., Gori, A., Al-Hajoj, S. A., Allix, C., Aristimuño, L., Arora, J., Baumanis, V., Binder, L., Cafrune, P., Cataldi, A., Cheong, S., Diel, R., Ellermeier, C., Evans, J. T., Fauville-Dufaux, M., Ferdinand, S., Garcia

- de Viedma, D., Sola, C. 2006. Mycobacterium tuberculosis complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC microbiology*, 6:23. <https://doi.org/10.1186/1471-2180-6-23>
- Coll, F., McNerney, R., Guerra-Assunção, J. A., Glynn, J. R., Perdigão, J., Viveiros, M., Portugal, I., Pain, A., Martin, N., & Clark, T. G. 2014. A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. *Nature communications*, 5: 4812. <https://doi.org/10.1038/ncomms5812>
- Coll, F., McNerney, R., Preston, M. D., Guerra-Assunção, J. A., Warry, A., Hill-Cawthorne, G., Mallard, K., Nair, M., Miranda, A., Alves, A., Perdigão, J., Viveiros, M., Portugal, I., Hasan, Z., Hasan, R., Glynn, J. R., Martin, N., Pain, A., & Clark, T. G. 2015. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome medicine*, 7(1): 51. <https://doi.org/10.1186/s13073-015-0164-0>
- 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>
- Coros, A., DeConno, E., & Derbyshire, K. M. 2008. IS6110, a *Mycobacterium tuberculosis* complex-specific insertion sequence, is also present in the genome of *Mycobacterium smegmatis*, suggestive of lateral gene transfer among mycobacterial species. *Journal of bacteriology*, 190(9): 3408–3410. <https://doi.org/10.1128/JB.00009-08>
- 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>
- Drancourt, M., & Raoult, D. 2007. Cost-effectiveness of blood agar for isolation of mycobacteria. *PloS neglected tropical diseases*, 1(2): e83. <https://doi.org/10.1371/journal.pntd.0000083>

- Filliol, I., Motiwala, A. S., Cavatore, M., Qi, W., Hazbón, M. H., Bobadilla del Valle, M., Fyfe, J., García-García, L., Rastogi, N., Sola, C., Zozio, T., Guerrero, M. I., León, C. I., Crabtree, J., Angiuoli, S., Eisenach, K. D., Durmaz, R., Joloba, M. L., Rendón, A., Sifuentes-Osornio, J., Alland, D. 2006. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *Journal of bacteriology*, 188(2): 759–772. <https://doi.org/10.1128/JB.188.2.759-772.2006>
- Gori, A., Bandera, A., Marchetti, G., Degli Esposti, A., Catozzi, L., Nardi, G. P., Gazzola, L., Ferrario, G., van Embden, J. D., van Soolingen, D., Moroni, M., & Franzetti, F. 2005. Spoligotyping and *Mycobacterium tuberculosis*. *Emerging infectious diseases*, 11(8): 1242–1248. <https://doi.org/10.3201/eid1108.040982>
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. 2013. QAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8): 1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
- Harland, C. W., Rabuka, D., Bertozzi, C. R., & Parthasarathy, R. 2008. The *Mycobacterium tuberculosis* virulence factor trehalose dimycolate imparts desiccation resistance to model mycobacterial membranes. *Biophysical journal*, 94(12): 4718–4724. <https://doi.org/10.1529/biophysj.107.125542>
- Hawkey, J., Hamidian, M., Wick, R. R., Edwards, D. J., Billman-Jacobe, H., Hall, R. M., & Holt, K. E. 2015. ISMapper: identifying transposase insertion sites in bacterial genomes from short read sequence data. *BMC genomics*, 16(1): 667. <https://doi.org/10.1186/s12864-015-1860-2>
- Hernández-Pando, R., Soto-Hernández, J. L., & Salinas-Lara, C. 2013. Aspectos inmunopatológicos de la tuberculosis meningea. *Revista de investigacion clinica; organo del Hospital de Enfermedades de la Nutricion*, 65(4): 349–356.
- Jiménez-Ruano, A.C., Madrazo-Moya, C.F., Cancino-Muñoz, I., Mejía-Ponce, P.M., Licona-Cassani, C., et al. 2021 Whole genomic sequencing-based genotyping reveals a specific X3 sublineage restricted to Mexico and related with multidrug resistance. *Sci Rep* (11): 1870-1878 doi: 10.1038/s41598-020-80919-5.

- Juarez-Eusebio, D. M., Munro-Rojas, D., Muñiz-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: journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 55: 384–391. <https://doi.org/10.1016/j.meegid.2016.09.012>
- Kohl, T.A., Utpatel, C., Schleusener, V., De Filippo, M.R., Beckert, P., Cirillo, D.M., Niemann, S. 2018. MTBseq: a comprehensive pipeline for whole genome sequence analysis of *Mycobacterium tuberculosis* complex strains. *PeerJ* 6:e5895. [10.7717/peerj.5895](https://doi.org/10.7717/peerj.5895).
- Lee, J. J., Lee, S. K., Song, N., Nathan, T. O., Swarts, B. M., Eum, S. Y., Ehrst, S., Cho, S. N., & Eoh, H. 2019. Transient drug-tolerance and permanent drug-resistance rely on the trehalose-catalytic shift in *Mycobacterium tuberculosis*. *Nature communications*, 10(1): 2928. <https://doi.org/10.1038/s41467-019-10975-7>
- Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* 44: W242–W245. [10.1093/nar/gkw290](https://doi.org/10.1093/nar/gkw290).
- Liang, P. K., Zheng, C., Xu, X. F., Zhao, Z. Z., Zhao, C. S., Li, C. H., Couvin, D., Reynaud, Y., Zozio, T., Rastogi, N., & Sun, Q. 2020. Local adaptive evolution of two distinct clades of Beijing and T families of *Mycobacterium tuberculosis* in Chongqing: a Bayesian population structure and phylogenetic study. *Infectious diseases of poverty*, 9(1):59. <https://doi.org/10.1186/s40249-020-00674-7>
- McEvoy, C. R., Falmer, A. A., Gey van Pittius, N. C., Victor, T. C., van Helden, P. D., & Warren, R. M. 2007. The role of IS6110 in the evolution of *Mycobacterium tuberculosis*. *Tuberculosis*, 87(5): 393–404. <https://doi.org/10.1016/j.tube.2007.05.010>
- Nebenzahl-Guimaraes, H., van Laarhoven, A., Farhat, M. R., Koeken, V., Mandemakers, J. J., Zomer, A., van Hijum, S., Netea, M. G., Murray, M., van Crevel, R., & van Soolingen, D. 2017. Transmissible *Mycobacterium tuberculosis* Strains Share Genetic Markers and Immune Phenotypes. *American journal of respiratory and*

critical care medicine, 195(11): 1519–1527. <https://doi.org/10.1164/rccm.201605-1042OC>

- Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32:268–274. [10.1093/molbev/msu300](https://doi.org/10.1093/molbev/msu300)
- Perdigão, J., Clemente, S., Ramos, J., Masakidi, P., Machado, D., Silva, C., Couto, I., Viveiros, M., Taveira, N., & Portugal, I. 2017. Genetic diversity, transmission dynamics and drug resistance of *Mycobacterium tuberculosis* in Angola. *Scientific reports*, 7: 42814. <https://doi.org/10.1038/srep42814>
- Reiche, M. A., Warner, D. F., & Mizrahi, V. 2017. Targeting DNA Replication and Repair for the Development of Novel Therapeutics against Tuberculosis. *Frontiers in molecular biosciences*, (4 75). <https://doi.org/10.3389/fmolb.2017.00075>
- Rodwell, T. C., Kapasi, A. J., Barnes, R. F., & Moser, K. S. 2012. Factors associated with genotype clustering of *Mycobacterium tuberculosis* isolates in an ethnically diverse region of southern California, United States. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 12(8): 1917–1925. <https://doi.org/10.1016/j.meegid.2012.08.022>
- Rosen, B. C., Dillon, N. A., Peterson, N. D., Minato, Y., & Baughn, A. D. 2017. Long-Chain Fatty Acyl Coenzyme A Ligase FadD2 Mediates Intrinsic Pyrazinamide Resistance in *Mycobacterium tuberculosis*. *Antimicrobial agents and chemotherapy*, 61(2): e02130-16. <https://doi.org/10.1128/AAC.02130-16>
- Roychowdhury, T., Mandal, S., & Bhattacharya, A. 2015. Analysis of IS6110 insertion sites provide a glimpse into genome evolution of *Mycobacterium tuberculosis*. *Scientific reports*, 5: 12567. <https://doi.org/10.1038/srep12567>
- Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14): 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
- Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., & Chandler, M. 2006. Isfinder: the reference centre for bacterial insertion sequences. *Nucleic acids research*, 34: D32–D36. <https://doi.org/10.1093/nar/gkj014>
- Singh, P., Rao, R. N., Reddy, J. R., Prasad, R. B., Kotturu, S. K., Ghosh, S., & Mukhopadhyay, S. 2016. PE11, a PE/PPE family protein of *Mycobacterium*

tuberculosis is involved in cell wall remodeling and virulence. *Scientific reports*, 6: 21624. <https://doi.org/10.1038/srep21624>

Supply, P., Mazars, E., Lesjean, S., Vincent, V., Gicquel, B., & Locht, C. 2000. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. *Molecular microbiology*, 36(3): 762–771. <https://doi.org/10.1046/j.1365-2958.2000.01905.x>

Tran, S. T., Le, H. T., Dang, H., Dao, T. M., Van Nguyen, L., Van Nguyen, H., Nguyen, N. T., Nguyen, N. V., Thwaites, G., Caws, M., van Doorn, H. R., & Wertheim, H. F. 2019. Mycobacterial Blood Culture for Diagnosis of Tuberculosis in Vietnamese Children. *The Pediatric infectious disease journal*, 38(11): e309–e312. <https://doi.org/10.1097/INF.0000000000002457>

von Groll, A., Martin, A., Stehr, M., Singh, M., Portaels, F., da Silva, P. E. A., & Palomino, J. C. 2010. Fitness of *Mycobacterium tuberculosis* strains of the W-Beijing and Non-W-Beijing genotype. *PloS one*, 5(4): e10191.

Xia, E., Teo, Y.Y., Ong, R.T. 2016. SpoTyping: fast and accurate in silico *Mycobacterium* spoligotyping from sequence reads. *Genome Medicine* 8:19. [10.1186/s13073-016-0270-7](https://doi.org/10.1186/s13073-016-0270-7).

Yang, Z., Yang, D., Kong, Y., Zhang, L., Marrs, C. F., Foxman, B., Bates, J. H., Wilson, F., & Cave, M. D. 2005. Clinical relevance of *Mycobacterium tuberculosis* plcD gene mutations. *American journal of respiratory and critical care medicine*, 171(12): 1436–1442. <https://doi.org/10.1164/rccm.200408-1147OC>

Zenteno-Cuevas, R., Fernandez, E., Viveros, D., Madrazo-Moya, C. F., Cancino-Muñoz, I., Comas, I., Gonzalez-Covarrubias, V., Barbosa-Amezcu, M., & Cuevas-Cordoba, B. 2020. Characterization of Polymorphisms Associated with Multidrug-Resistant Tuberculosis by Whole Genomic Sequencing: A Preliminary Report from Mexico. *Microbial drug resistance* 26(7): 732–740. <https://doi.org/10.1089/mdr.2019.0054>

CAPÍTULO III: Identificación de polimorfismos asociados a los fenotipos pulmonar y extrapulmonar de cepas de *Mycobacterium tuberculosis*

INTRODUCCIÓN

La tuberculosis (TB) es una de las enfermedades transmisibles con mayor mortalidad a nivel mundial, además de ser una de las enfermedades infecciosas más antiguas y conocidas de la humanidad. Se trata de una afección predominantemente pulmonar, que en humanos se produce principalmente por la especie *Mycobacterium tuberculosis* (MTB) y en menor proporción por las otras especies del complejo *M. tuberculosis*: *M. bovis*, *M. africanum*, *M. microti* y *M. canetti* (Coscolla et al., 2014).

La TB suele clasificarse en pulmonar (PTB) y extrapulmonar (EPTB). Aunque el número de casos de TB ha disminuido significativamente, las proporciones de casos de EPTB han permanecido constantes. Esta enfermedad es el resultado de la diseminación hematogena y linfática de MTB. Las localizaciones extrapulmonares más frecuentes son la pleura, los ganglios linfáticos, el sistema genitourinario, los huesos y articulaciones, las meninges y el peritoneo, no obstante, se puede afectar cualquier órgano o sistema. En la TB meníngea luego de la primoinfección o reactivación, se da una proliferación de bacilos, que se pueden diseminar al sistema nervioso central (SNC). Al producirse la ruptura de un tubérculo en el espacio subaracnoideo, se genera una infección por parte del bacilo en esta zona, ocasionando una reacción inflamatoria del tejido (Rodríguez et al., 2021). En esta localización, la aracnoiditis localizada en la base del cráneo constituye masas fibrosas que envuelven los pares craneales y los vasos, produciendo vasculitis con la consiguiente trombosis e infartos. La tuberculosis meníngea representa del 1 al 5% de todos los casos de TB y es considerada como la forma más severa de la enfermedad, resultando en el fallecimiento del paciente o la presencia de daño neurológico en más de la mitad de los afectados (van Laarhoven et al., 2017). Actualmente se desconoce en gran medida porque ciertas personas desarrollan PTB y otras EPTB. Los factores relacionados con el hospedero juegan un papel importante en el establecimiento de la enfermedad, como lo demuestra una mayor propensión a desarrollar EPTB en pacientes inmunodeprimidos e infantes. Además, los factores genéticos propios del hospedero también pueden ser claves para la aparición de los diferentes fenotipos clínicos de la TB; por ejemplo, variaciones en

los genes que codifican para receptores de vitamina D, P2X7, interleucina 10 e interferón gamma (Sharma et al., 2017). Diversos estudios epidemiológicos y experimentos con modelos animales han tratado de determinar si los linajes o genotipos del patógeno están asociados con la presentación de la enfermedad en sitios extrapulmonares. Por el momento, los hallazgos encontrados han sido inconsistentes y no concluyentes para atribuir la contribución de variaciones genéticas específicas a la presentación de la enfermedad (Coscolla et al., 2014).

La secuenciación del genoma completo provee un nivel de resolución más fino en comparación con las técnicas antes mencionadas para la búsqueda de determinantes genéticos asociados a EPTB. Los estudios que han utilizado esta estrategia para encontrar dichas asociaciones todavía son escasos, y han evaluado principalmente cepas aisladas de pacientes con TB meníngea. Un trabajo realizado en India en 2013 mediante la secuenciación analizó 5 cepas causantes de EPTB detectó heterogeneidad genómica entre las cepas, es posible que la secuenciación genómica se haya realizado en muestras con ADN contaminado por material genómico de más de una cepa de MTB (Das et al., 2013). Otras aproximaciones similares realizadas en Tailandia e Indonesia utilizaron una mayor cantidad de cepas EPTB causantes de TB meníngea y compararon sus genomas contra grupos de cepas PTB aisladas de la misma región geográfica (Faksri et al., 2018; Ruesen et al., 2018). Sus resultados muestran mutaciones exclusivas de cepas causantes de EPTB pero que no se encuentran en la totalidad de estas. Aunque se trata de dos trabajos independientes, ambos emplearon una metodología de comparación similar, pero que no permite la comparación de sus resultados. Por lo anterior, este trabajo se pretende determinar la presencia de microvariantes genómicas asociadas al fenotipo EPTB de la tuberculosis. Se utilizó un número de cepas EPTB significativamente mayor al de estudios previos, empleando genomas de cepas provenientes de las distintas regiones geográficas disponibles en las bases de datos públicas. Esta estrategia de análisis global de cepas buscó evitar un sesgo por la predominancia de linajes particulares en determinadas regiones geográficas.

MATERIALES Y MÉTODOS

Obtención de los genomas

Se recuperaron un total de 490 datos crudos de lecturas de secuenciación de la base de datos NCBI SRA, que corresponden a 245 aislados de MTB causantes de PTB y 245 aislados causantes de EPTB (Anexo 2). Se consideraron como cepas de EPTB aquellas aisladas de casos en los que el sitio de infección reportado en la base de datos no fuera pulmonar o miliar. La calidad de la secuencia de las lecturas FASTQ se verificó mediante FASTQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) y posteriormente se filtró a una puntuación Phred de 30 mediante TRIMGALORE (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). A estas secuencias filtradas y de buena calidad se añadieron las 10 cepas MYC descritas en el capítulo anterior para obtener una n=500, de las cuales 250 son EPTB y 250 son PTB.

Determinación de microvariantes

A partir de los archivos fastq de las cepas seleccionadas se utilizó la herramienta MTBseq versión 1.0.3 (Kohl et al., 2018) con los valores por defecto para realizar el mapeo contra el genoma de referencia H37Rv y obtener los archivos individuales de la presencia de variaciones en cada uno de los genomas. La presencia de variantes alélicas se determinó basándose en los siguientes parámetros: cuatro lecturas “forward” y cuatro “reverse” para indicar la presencia de un alelo, valor de phred mayor a 20 y frecuencia alélica mayor a 75%. La identificación de microvariantes específicas de las cepas PTB y EPTB se llevó a cabo a partir de los archivos de salida de MTBseq en donde se reportan los SNPs, inserciones y deleciones por cepa (.gatk_positions_variants_cf4.tab) utilizando un script en MySQL para comparar ambos grupos e identificar las variaciones totales, únicas, compartidas y exclusivas de los grupos PTB y EPTB. Con este fin, se concatenaron los archivos _variants.tab y se filtraron las variantes asociadas a la clasificación

filogenética y resistencia a antibióticos, además de aquellas con un valor de phred menor a 30.

A partir de estos datos se identificó la inserción exclusiva más común de cepas EPTB y PTB, la delección exclusiva más común de cepas EPTB y PTB, y el SNP exclusivo más común de cepas EPTB y PTB.

Análisis estadístico

El análisis estadístico se realizó en SPSS versión 16 (SPSS Inc., Illinois, EE. UU.). La asociación entre los fenotipos clínicos de MTB y la presencia de SNPs, inserciones o delecciones se estimó mediante la prueba de chi-cuadrada, con intervalos de confianza del 95% para evaluar la significancia estadística.

RESULTADOS

Se obtuvieron un total de 490 datos crudos de secuenciación de cepas de MTB aisladas de siete países diferentes, todas ellas provenientes de pacientes con un estado negativo de infección por VIH (Anexo 1). Las cepas seleccionadas incluyeron aquellos países donde predominan linajes específicos según la base de datos SITVIT (http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE/), como Tailandia e Indonesia para el linaje de Asia Oriental, e India para el linaje indio del este de África. Los genomas pertenecientes al fenotipo EPTB se agruparon en cinco grandes grupos según el sitio anatómico de aislamiento, con predominio del sistema nervioso central (72.24%). Otros sitios anatómicos incluidos fueron huesos y articulaciones (17.14%), ganglios linfáticos (4.49%) y aparato genitourinario (2.86%); el 3.27% restante procede de sitios no especificados pero identificados EPTB en las bases de datos.

A partir del análisis de microvariantes se identificaron 53,440 mutaciones repartidas entre las 500 cepas de estudio, de estas 13,560 son mutaciones compartidas, 21,400 son exclusivas del fenotipo EPTB y 18,480 son exclusivas del fenotipo PTB (Figura 12).

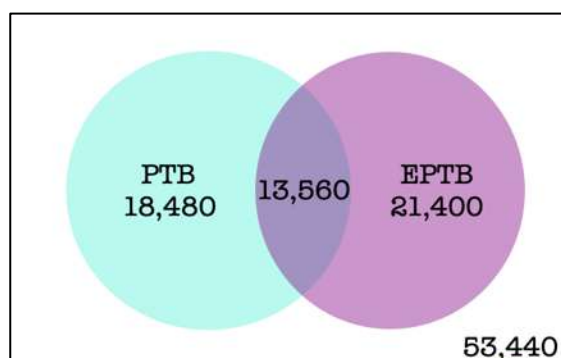


Figura 12. Diagrama de Venn de las microvariantes identificadas en las cepas de *M. tuberculosis*. EPTB, tuberculosis extrapulmonar; PTB, tuberculosis pulmonar.

Del total de mutaciones encontradas 1,406 corresponden a inserciones, 11,545 a deleciones y 40,489 a polimorfismos de un solo nucleótido, de estos últimos 33,277 están presentes en regiones codificantes y 20,298 son mutaciones no sinónimas. Utilizando estos datos se buscaron en los genes exclusivos los SNPs

más frecuentes que solo estuvieran presentes en el fenotipo EPTB, encontrándose una variación en el gen *ppe57* como la más común. Esta mutación ocasiona un cambio en la cadena de aminoácidos en la posición 110 de una alanina por una treonina (Ala110Thr), y aunque no se encuentra en todas las cepas EPTB, si se encuentra asociado a este fenotipo ($p=0.0133$, OR:35.24; IC 95% 2.1-59.0).

Por su parte, respecto al fenotipo PTB se identificaron 2 SNPs en el gen *pe_pgrs10* exclusivos se dicho fenotipo (ocasionando los cambios Ala523Thr, Ala518Gly), y de manera similar, aunque no se encuentran de manera consistente en todas las cepas PTB, si se asocian con dicho fenotipo ($p=0.0036$, OR:64.3, IC 95% 3.91-106.0; $p=0.004$, OR:61.64, IC95% 3.73-106.5).

Las mutaciones compartidas por ambos fenotipos, pero asociadas con el fenotipo EPTB, se muestran en la Tabla 13.

Tabla 13. Mutaciones asociadas al fenotipo extrapulmonar. SNP, polimorfismo de un solo nucleótido; OR, odds ratio; IC, intervalo de confianza; AA, aminoácido.

SNP	Valor de p	OR	IC 95%	CAMBIO AA	Gen	Proteína	Tipo SNP
T2165503A	< 0.0001	2.45	1.63-3.68	Ala603Ala (gca/gcT)	<i>Rv1917c</i>	PPE34	Sinonimo
A3883626G	0.0002	2.08	1.41-3.09	--	--	--	No CDS
G103836T	< 0.0001	4.95	2.56-9.56	Asn276Lys (aac/aaA)	<i>Rv0094c</i>	Proteína hipotética	No sinónimo
G3730582A	< 0.0001	4.21	2.30-7.70	Ala2107Val (gcc/gTc)	<i>Rv3343c</i>	PPE54	No sinónimo
G2439519A	< 0.0001	2.90	1.75-7.80	Arg143Arg (cgc/cgT)	<i>Rv2177c</i>	Transposasa	Sinonimo
C3884791A	< 0.0001	3.9	2.13-7.18	Asn276Lys (aac/aaA)	<i>Rv3467</i>	Proteína hipotética	No sinónimo
G3732344A	0.0007	1.9	1.31-2.81	Ala1531Val (gcc/gTc)	<i>Rv3343c</i>	PPE54	No sinónimo
C104824A	< 0.0001	4.00	2.1-7.6	Ser131Ile (agc/aTc)	<i>Rv0095c</i>	Proteína hipotética	No sinónimo
G3883915A	< 0.0001	4.88	3.04-7.82	Val131Ile (gtc/Atc)	<i>Rv3466</i>	Proteína hipotética	No sinónimo
C80616G	0.0040	1.72	1.18-2.49	--	--	--	No CDS
A1993808T	0.0031	1.81	1.22-2.69	Glu219Val (gag/gTg)	<i>Rv1760</i>	Diacylglycerol acyltransferasa	No sinónimo
G1096323T	0.0002	2.88	1.64-5.07	His43Gln (cac/caA)	<i>Rv0980c</i>	PE_PGRS18	No sinónimo
G1788570T	0.0001	3.34	1.8-6.19	Asn198Lys (aac/aaA)	<i>Rv1587c</i>	Proteína hipotética	No sinónimo
C841494G	0.0001	4.01	2.00-8.04	Leu89Leu (ctc/ctG)	<i>Rv0749</i>	vapC31	Sinonimo

T2030521C	< 0.0001	4.98	2.44-10.16	--	--	--	No CDS
C560664T	< 0.0001	4.57	2.23-9.36	Asn259Asn (aac/aaT)	Rv0469	umaA	Sinonimo
C3119513T	< 0.0001	5.59	2.55-12.23	--	--	--	No CDS
G3895269C	< 0.0001	4.43	2.16-9.10	Glu282Gln (gaa/Caa)	Rv3478	PPE60	No sinónimo

El 27.7% de las mutaciones compartidas de cepas EPTB se encuentran en genes que codifican proteínas hipotéticas, el 22.2% en regiones no codificantes y el 22.2% corresponden a mutaciones sinónimas.

Las mutaciones compartidas por ambos fenotipos, pero asociadas con el fenotipo PTB, se muestran en la Tabla 14.

Tabla 14. Mutaciones asociadas al fenotipo pulmonar. SNP, polimorfismo de un solo nucleótido; OR, odds ratio; IC, intervalo de confianza; AA, aminoácido.

SNP	p value	OR	IC 95%	CAMBIO AA	GEN	PROTEINA	Tipo SNP
G3379730C SNP	<0.0001	3.37	2.27- 5.01	--	--	--	No CDS
T336560C SNP	<0.0001	2.88	1.96- 4.24	_838_ (taa/taG)	Rv0279c	PE_PGRS4	Codón de paro
T3379742C SNP	<0.0001	3.05	2.03- 4.59	--	--	--	No CDS
A3379735C SNP	<0.0001	2.74	1.86- 4.04	--	--	--	No CDS
A336380T SNP	<0.0001	2.34	1.63- 3.36	--	--	--	No CDS
A336405G SNP	<0.0001	2.30	1.61- 3.13	--	--	--	No CDS
C336400G SNP	<0.0001	2.34	1.63- 3.36	--	--	--	No CDS
C336403G SNP	<0.0001	2.30	1.61- 3.13	--	--	--	No CDS
T336537G SNP	<0.0001	2.44	1.67- 3.54	--	--	--	No CDS
C3949532G SNP	<0.0001	3.03	1.96- 4.66	Pro1247Ala (ccc/Gcc)	Rv3514	PE_PGRS57	No sinónimo
G336540T SNP	<0.0001	2.39	1.64- 3.47	--	--	--	No CDS
T336546G SNP	<0.0001	2.34	1.61- 3.40	--	--	--	No CDS
T336592G SNP	<0.0001	2.34	1.61- 3.40	Ile828Leu (atc/Ctc)	Rv0279c	PE_PGRS4	No sinónimo
G336590C SNP	<0.0001	2.30	1.59- 3.34	Ile828Met (atc/atG)	Rv0279c	PE_PGRS4	No sinónimo

G3741572A SNP	<0.0001	2.50	1.68- 3.70	Gly401Gly (ggc/ggT)	Rv3345c	PE_PGRS50	Sinónimo
T336535G SNP	<0.0001	2.28	1.57- 3.30	--	--	--	No CDS
C673564T SNP	<0.0001	2.90	1.86- 4.50	Ala785Thr (gcc/acc)	Rv0578c	PE_PGRS7	No sinónimo
G336617C SNP	<0.0001	2.29	1.57- 3.35	Pro819Pro (ccc/ccG)	Rv0279c	PE_PGRS4	Sinónimo
G336611C SNP	<0.0001	2.20	1.51- 3.20	Ala821Ala (gcc/gcG)	Rv0279c	PE_PGRS4	Sinónimo
G3927595A SNP	<0.0001	3.08	1.94- 4.90	Ala343Thr (gcc/acc)	Rv3507	PE_PGRS53	No sinónimo
G4032625T SNP	<0.0001	2.45	1.64- 3.66	Gly178Gly (ggc/ggA)	Rv3590c	PE_PGRS58	Sinónimo
C3949531G SNP	<0.0001	2.68	1.74- 4.13	Gly1246Gly (ggc/ggG)	Rv3514	PE_PGRS57	Sinónimo

El 50% de las mutaciones compartidas de cepas PTB se encuentran en genes que codifican proteínas de la familia *pe_pgrs*, el 50% en regiones no codificantes y el 18% corresponden a mutaciones sinónimas.

DISCUSIÓN

Con la finalidad de determinar la variación genética asociada al fenotipo clínico de la tuberculosis, se realizó la comparación de 500 cepas de MTB aisladas de distintas regiones geográficas para eliminar sesgos en la elección de los linajes incluidos debido a la distribución geográfica de los genotipos, incluyendo también un número igualitario de cepas PTB y EPTB. Mediante este análisis se identificaron 19 polimorfismos de un solo nucleótido asociados con el fenotipo extrapulmonar de la TB, uno de ellos exclusivo de este grupo, que se localiza en el gen *ppe57*. Aunque se desconoce la función del gen *ppe57*, se sabe que interviene en el reconocimiento de la bacteria por los receptores tipo Toll2, ocasionando la activación de la vía de las MAPK (MAP cinasas) y la activación de factores de transcripción que llevarán a una producción aumentada de IL12 y TNF- α , originando una fuerte respuesta inflamatoria y disminuyendo la tasa de persistencia de la bacteria en el macrófago (Mogensen, 2009). Sin embargo, si existe una mutación que altere la función de esta proteína, las vías de señalización se pueden ver modificadas, ocasionando una menor activación de macrófagos, y por lo tanto una mayor persistencia y probable diseminación de la bacteria observada en las cepas EPTB.

Adicionalmente, se identificaron SNPS que, a pesar de estar presentes en ambos fenotipos, debido a la frecuencia observada se pudieran asociar con PTB o EPTB. Gracias a este análisis se lograron determinar 19 SNPs asociados con EPTB y 22 con PTB. Para EPTB los 19 SNPs se encontraron en los genes *ppe54*, *ppe60*, *pgrs19* y *Rv1760*, además de otros genes que codifican proteínas hipotéticas. Se ha reportado que la proteína codificada por *ppe54* se expresa en macrófagos activados y que interviene en mecanismos de resistencia a estrés oxidativo (Mycobrowser, 2022). Respecto a *Rv1760*, se sabe que es sobreexpresado en condiciones de estrés, por lo que también podría intervenir en estos mecanismos (Mestre et al., 2013). La función de *pgrs18* está asociada a la inhibición de la apoptosis, y *ppe60*, al igual que *ppe54* y *Rv1760*, está involucrado en mecanismos de resistencia a estrés. Se sabe que desencadena una respuesta proinflamatoria con un aumento de la producción de TNF- α y el paso a piroptosis, una forma de

muerte celular programada que se correlaciona con la restricción del crecimiento intracelular y la respuesta inmune mejorada del hospedero (Quian et al., 2020). Las mutaciones encontradas podrían alterar la función de estas proteínas ocasionando modificaciones tanto en los mecanismos de resistencia a estrés, como en la respuesta inmune que podrían ser claves para explicar el fenotipo EPTB.

En el caso de PTB se encontraron 2 polimorfismos en el gen *pgrs10* exclusivos de este fenotipo y, de manera similar, aunque no se encuentran de manera consistente en todas las PTB si se asocian al fenotipo. Se desconoce la función de *pepgrs10*; sin embargo, pertenece a un grupo de proteínas PGRS para las cuales se ha descrito actividad de unión a fibronectina, estas proteínas se encuentran en la superficie celular y son necesarias para la unión a la célula hospedadora y la internalización (Meena, 2019). Una vez que ésta se lleva a cabo, ocurre la formación del fagolisosoma que va a desencadenar en la excreción de péptidos utilizados para la presentación antigénica por las células dendríticas. Una alteración en estas proteínas podría llevar a una disminución en la tasa de activación de los macrófagos.

Por otra parte, respecto a los 22 SNPs compartidos asociados con PTB, estos se encontraron en genes como *pe_pgrs4*, *pe_pgrs7*, *pe_pgrs53* y *pe_pgrs57*. Se sabe que *pe_pgrs4* y *pe_pgrs7* están involucrados en mecanismos de unión a la superficie de células fagocíticas ya que tienen actividad de fibronectina y adhesina (Meena, 2019). Se ha reportado previamente la delección del gen *pe_pgrs4* en cepas Beijing o pertenecientes al linaje 2, e interesantemente la variación asociada a PTB en este gen es la presencia de un codón de paro en la posición 838, ésta se observa en 115 de las 260 cepas Beijing de nuestro conjunto de datos. Finalmente, se sabe los genes *pe_pgrs53* y *pe_pgrs57* que están involucrados en el metabolismo del colesterol; sin embargo, se desconoce su función específica.

Como se comentó anteriormente, se han reportado previamente SNPs asociados tanto a EPTB como a PTB en otros trabajos de análisis genómico comparativo. En el estudio llevado a cabo por Faksri en 2018 se reportó una asociación de un SNP en el gen *pe_pgrs6* con las cepas causantes de TB meníngea (Faksri et al., 2018). En el mismo año, Ruesen y colaboradores reportan un SNP

en el gen *Rv0218* asociado con PTB (Ruesen et al., 2018). Sin embargo, en nuestro análisis el SNP presente en *pe_pgrs6* se encuentra en su mayoría en cepas pertenecientes al fenotipo PTB, mientras que el SNP en *Rv0218*, aunque si se observa con mayor frecuencia en cepas PTB, no se encuentra asociado a dicho fenotipo. Estos estudios difieren del presente trabajo principalmente en la naturaleza y cantidad de las cepas utilizadas en donde no se observa una subrepresentación de los casos EPTB.

CONCLUSIÓN

Las variaciones genéticas diferenciales (SNPs) entre los fenotipos EPTB y PTB se encuentran mayoritariamente en genes pertenecientes a la familia PE/PPE/PGRS. Dichos SNPs pueden ocasionar una alteración en la función de las proteínas codificadas al modificar la secuencia de aminoácidos, provocando variaciones tanto en la respuesta inmune como en los mecanismos de respuesta a estrés que se encuentran asociados a las distintas presentaciones clínicas de la tuberculosis

REFERENCIAS

- 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. K., & Bhattacharya, A. 2013. Genetic heterogeneity revealed by sequence analysis of *Mycobacterium tuberculosis* isolates from extra-pulmonary tuberculosis patients. *BMC genomics*, 14(404). <https://doi.org/10.1186/1471-2164-14-404>
- Faksri, K., Xia, E., Ong, R. T., 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>
- Kohl, T.A., Utpatel, C., Schleusener, V., De Filippo, M.R., Beckert, P., Cirillo, D.M., Nieman, S. 2018. MTBseq: a comprehensive pipeline for whole genome sequence analysis of *Mycobacterium tuberculosis* complex strains. *PeerJ* 6: e5895 DOI 10.7717/peerj.5895.
- Meena, L. S. 2019. Interrelation of Ca²⁺ and PE_PGRS proteins during *Mycobacterium tuberculosis* pathogenesis. *Journal of biosciences*, 44(1):24.
- Mestre, O., Hurtado-Ortiz, R., Dos Vultos, T., Namouchi, A., Cimino, M., Pimentel, M., Neyrolles, O., & Gicquel, B. 2013. High throughput phenotypic selection of *Mycobacterium tuberculosis* mutants with impaired resistance to reactive oxygen species identifies genes important for intracellular growth. *PloS one*, 8(1): e53486. <https://doi.org/10.1371/journal.pone.0053486>

- Mogensen, T. H. 2009. Pathogen recognition and inflammatory signaling in innate immune defenses. *Clinical microbiology reviews*, 22(2): 240–273. <https://doi.org/10.1128/CMR.00046-08>
- Qian, J., Chen, R., Wang, H., & Zhang, X. 2020. Role of the PE/PPE Family in Host-Pathogen Interactions and Prospects for Anti-Tuberculosis Vaccine and Diagnostic Tool Design. *Frontiers in cellular and infection microbiology*, 10: 594288. <https://doi.org/10.3389/fcimb.2020.594288>
- Rodríguez, D., Zapata, A., Molina, F., Pemuolo, G., Burbano, J., & Correa, L. 2021. Meningeal tuberculosis: cases report years 2005-2017. *Revista Chilena de Infectologia: Organo Oficial de la Sociedad Chilena de Infectologia*, 38(3): 410-416.
- Ruesen, C., Chaidir, L., van Laarhoven, A., Dian, S., Ganiem, A. R., Nebenzahl-Guimaraes, H., Huynen, M. A., Alisjahbana, B., Dutilh, B. E., & van Crevel, R. 2018. Large-scale genomic analysis shows association between homoplastic genetic variation in *Mycobacterium tuberculosis* genes and meningeal or pulmonary tuberculosis. *BMC genomics*, 19(1): 122. <https://doi.org/10.1186/s12864-018-4498-z>
- Sharma, K., Verma, R., Advani, J., Chatterjee, O., Solanki, H. S., Sharma, A., Varma, S., Modi, M., Ray, P., Mukherjee, K. K., Sharma, M., Dhillion, M. S., Suar, M., Chatterjee, A., Pandey, A., Prasad, T., & Gowda, H. 2017. Whole Genome Sequencing of *Mycobacterium tuberculosis* Isolates From Extrapulmonary Sites. *Omic: a journal of integrative biology*, 21(7): 413–425. <https://doi.org/10.1089/omi.2017.0070>
- van Laarhoven, A., Dian, S., Ruesen, C., Hayati, E., Damen, M., Annisa, J., Chaidir, L., Ruslami, R., Achmad, T. H., Netea, M. G., Alisjahbana, B., Rizal Ganiem, A., & van Crevel, R. 2017. Clinical Parameters, Routine Inflammatory Markers, and LTA4H Genotype as Predictors of Mortality Among 608 Patients With Tuberculous Meningitis in Indonesia. *The Journal of infectious diseases*, 215(7): 1029–1039. <https://doi.org/10.1093/infdis/jix051>

CAPÍTULO IV: Identificación de genes asociados a los fenotipos pulmonar y extrapulmonar de cepas de *Mycobacterium tuberculosis*

INTRODUCCIÓN

Se estima que actualmente 2 mil millones de personas están infectadas por *Mycobacterium tuberculosis* (MTB), el agente causal de la tuberculosis (TB), enfermedad que afecta anualmente a más de 14 millones de personas y que causa la muerte a casi 2 millones de ellas. Las principales causas que han dado lugar a la reemergencia de la TB son la coinfección con el VIH, la aparición de cepas de MTB resistentes y multirresistentes a los fármacos del tratamiento, y la baja eficacia de la vacuna actual (OMS, 2021). Aunque se estima que una tercera parte de la población mundial está infectada por MTB, solo en un 5% ó 10% de los casos se desarrolla la enfermedad, ya sea en el primer o segundo año tras la infección (TB primaria) o por reactivación de la infección a lo largo de la vida (TB secundaria). En el resto de los casos la infección permanece latente durante toda la vida (OMS, 2021).

Cuando se desarrolla TB activa, la localización de la enfermedad, la severidad, y el desenlace son muy variables. En un 85% de los casos la TB se limita a los pulmones. Mientras que en aproximadamente un 15% de los casos se desarrolla tuberculosis extrapulmonar (EPTB), la cual requiere de procedimientos invasivos como la biopsia para su diagnóstico, por lo que suele ser de difícil diagnóstico. Las formas miliar y meníngea de la TB son consideradas las formas más graves de la enfermedad y generalmente son de pronóstico reservado (Rodríguez et al., 2021). La amplia diversidad en las manifestaciones clínicas y la gravedad de la enfermedad son uno de los aspectos más intrigantes de la TB y se han atribuido principalmente a factores del hospedero como la infección por VIH, la presencia de diabetes, el tabaquismo, la edad y la desnutrición, entre otros. Sin embargo, los factores del hospedero no explican completamente la variación observada. Una gran cantidad de estudios epidemiológicos y experimentales han demostrado que la diversidad genética de las cepas de MTB también es importante (Coscolla, 2017). Dentro de estos estudios, el análisis genómico comparativo ha cobrado especial importancia, ya que permite llevar a cabo comparaciones entre

cepas pulmonares (PTB) y EPTB mediante la búsqueda de determinantes genéticos asociados a tropismo utilizando diferentes estrategias.

En un estudio en donde se analizaron 8 cepas causantes de TB meníngea en pacientes de Malasia (Saw et al., 2016), se lograron identificar determinantes genéticos asociados con el neurotropismo mediante la comparación con cepas causantes de PTB. Las comparaciones genómicas revelaron reordenamientos como traslocaciones, inversiones, inserciones y eliminaciones, así como SNP no sinónimos en las cepas derivadas de líquido cefalorraquídeo que no se observaron en los genomas de origen respiratorio de MTB. Estos reordenamientos poseen una gran cantidad de genes PE (prolina-glutamato) /PPE (prolina-prolina-glutamato), proteínas transcripcionales y de membrana. De manera similar, la mayoría de los SNPs comunes en las cepas meníngeas se observaron en genes que codifican proteínas PE y PPE. Aunque se reportan variaciones exclusivas de las cepas EPTB en este trabajo asociadas a neurotropismo, no se lograron identificar determinantes genéticos o variaciones comunes a todas las cepas aisladas de pacientes con EPTB. Posteriormente, en 2017 se seleccionaron 5 cepas correspondientes al linaje 3 que fueron aisladas de pacientes con EPTB en India (Sharma et al., 2017), las cuales se utilizaron para realizar un análisis similar a los dos anteriores, utilizando cepas PTB depositadas en bases de datos para comparar. Los resultados mostraron similitudes también con los trabajos previos, reportándose 279 variaciones genómicas exclusivas de las cepas EPTB pero que no se comparten entre ellas. Finalmente, en 2018 se llevaron a cabo dos estudios en los que se utilizó una mayor cantidad de cepas EPTB comparando con cepas PTB aisladas de la misma región geográfica. El primero de ellos utilizó 322 cepas, 106 EPTB y 206 PTB (Ruesen, 2018), el segundo un total de 293, 73 EPTB y 220 PTB (Faksri et al., 2018). Los autores utilizaron distintas estrategias metodológicas para obtener determinantes genéticos asociados a TB meníngea; sin embargo, no se reportan variantes exclusivas de las cepas EPTB compartidas entre la mayoría de las cepas que ocasionan este fenotipo clínico. La razón de estos resultados puede deberse a varios factores: a) el bajo número de cepas de MTB causantes de EPTB; b) las cepas analizadas pertenecen a una sola región geográfica, por lo tanto, existe un

sesgo asociado a la predominancia de un linaje específico; y c) la estrategia de análisis genómico utilizada. Es por esto por lo que a la fecha no existe un consenso sobre los determinantes genéticos que contribuyen a las formas extrapulmonares de la TB. El análisis genómico comparativo derivado de la construcción del pangenoma puede ser una estrategia novedosa para encontrar diferencias que contribuyan a explicar la presentación clínica de la enfermedad, algo no evaluado hasta la fecha.

El pangenoma describe la colección de todos los genes en una especie determinada o grupo de microorganismos e incluye un genoma central o genoma núcleo (*core*) y un genoma accesorio. El genoma núcleo consiste en todos los genes presentes en la totalidad de las cepas, mientras que el genoma accesorio está conformado por los genes dispensables que se pueden encontrar en dos o más cepas y por los genes que son exclusivos de una sola cepa. Se cree que el genoma accesorio juega un papel importante en la variación fenotípica y la evolución del genoma (Zahkam et al., 2021). Los genes dispensables y los genes específicos de cepas se consideran genes secundarios, que determinan las características parcialmente compartidas y específicas de las cepas de una especie y no son esenciales para la supervivencia. Estos genes parcialmente compartidos o específicos también son importantes para la adaptación del bacilo a ciertas circunstancias (Yang et al., 2018). Por lo que este estudio pretende evaluar la composición del genoma accesorio del pangenoma construido a partir de cepas PTB y EPTB con la finalidad de identificar aquellos genes asociados con la variación fenotípica en relación con la presentación clínica de la TB.

MATERIALES Y MÉTODOS

Obtención de los genomas

Se recuperaron un total de 490 datos crudos de lecturas de secuenciación de la base de datos NCBI SRA, que corresponden a 245 aislados de MTB causantes de PTB y 245 aislados causantes de EPTB (Anexo 1). Se consideraron como cepas de EPTB aquellas aisladas de casos en los que el sitio de infección reportado en la base de datos no fuera pulmonar o miliar. La calidad de la secuencia de las lecturas FASTQ se verificó mediante FASTQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) y posteriormente se filtró a una puntuación Phred de 30 mediante TRIMGALORE (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). A estas secuencias filtradas y de buena calidad se añadieron las 10 cepas MYC descritas en el capítulo anterior para obtener una n=500, donde 250 son EPTB y 250 son PTB.

Ensamblado y anotación genómica

Partiendo de las lecturas filtradas en formato fastq, se llevó a cabo el ensamblado de los genomas estudiados con el programa SPAdes, basado en diagramas de Brujin (Bankevich et al., 2012). El programa se utilizó con los parámetros predeterminados, desactivando la opción de filtrado de secuencias, indicando los archivos de las lecturas forward y reverse y reduciendo el número de errores e indels cortos utilizando la opción -careful. La validación de los ensamblados se llevó a cabo con el programa QUAST v4.6.0, (Gurevich et al., 2013). QUAST evalúa los ensamblados de los genomas, utilizando varias características métricas (número de contigs, longitud total del genoma, % de GC, N50, L50, N75, L75, longitud de los contigs, entre otros) y puede evaluar la calidad del ensamblado incluso sin un genoma de referencia. A pesar de que no es necesario un genoma de referencia, el utilizar alguno da una mejor visión de que tan exitoso fue el ensamblado, por lo que se utilizó el genoma de la cepa de referencia H37Rv para el análisis.

Con el objetivo de realizar una mejora de los ensamblados se realizó un filtrado de contigs con menos de 500 pb y un reordenamiento con ABACAS (Assefa et al., 2009). Esta herramienta se basa en la utilización de un genoma de referencia para el ordenamiento y orientación de los contigs obtenidos en un ensamblado. Las bases indeterminadas se denominan con Ns en las secuencias obteniendo finalmente scaffolds representativos del genoma secuenciado.

La anotación de los ensamblados se realizó mediante Prokka v1.10 (Seemann, 2014). Con este fin, se descargó el archivo que contiene las secuencias de aminoácidos del genoma de referencia H37Rv y el archivo .gff del mismo (RefSeq NC_000962.3).

Construcción del pangenoma

Se realizó mediante la herramienta Roary (Page et al., 2015) a partir de las anotaciones en archivo gff, utilizando los parámetros predeterminados en donde se incluye un porcentaje de identidad mínima de 90, la creación de un archivo multifasta a partir del alineamiento del genoma núcleo, alineamiento rápido con MAFFT y la creación de plots con R. Roary utiliza el algoritmo CD-HIT (Fu et al., 2012) para realizar un pre-agrupamiento rápido de las secuencias para posteriormente comparar una sola secuencia representativa de cada pre-grupo con las demás mediante BLASTp, y finalmente reagrupar estas secuencias mediante MLC.

Identificación de genes exclusivos

La determinación de los genes accesorios exclusivos a los fenotipos extrapulmonar y pulmonar se realizó con la herramienta query_pan_genome (opción -difference) que permite identificar diferencias en el contenido de genes entre dos grupos específicos. Adicionalmente se utilizó SCOARY (Brynildsrud et al., 2016) que toma los archivos de salida de ROARY en formato .csv que incluyen la información de ausencia y presencia de genes para realizar un análisis de asociación genómico, en

el que se destacan los genes que podrían estar relacionados con un fenotipo o característica clínica específica.

RESULTADOS

Se utilizó un total de 500 genomas de cepas clínicas de MTB para la construcción del pangenoma. Estos incluyen 490 cepas aisladas de siete países diferentes y 10 aisladas del estado de Michoacán que representan la región de Latinoamérica, todas ellas aisladas de pacientes con un estado negativo de infección por VIH (Anexo 1). Las cepas seleccionadas incluyeron aquellos países en donde predominan linajes específicos según la base de datos SITVIT (http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE/), como Tailandia e Indonesia para el linaje de Asia Oriental e India para el linaje indio del este de África.

Los genomas pertenecientes al fenotipo EPTB se agruparon en cinco grandes grupos según el sitio anatómico de aislamiento, con predominio del sistema nervioso central (72.24%). Otros sitios anatómicos incluidos fueron huesos y articulaciones (17.14%), ganglios linfáticos (4.49%) y aparato genitourinario (2.86%); el 3.27% restante procede de sitios no especificados pero identificados EPTB en las bases de datos. A partir de la construcción del pangenoma de estas 500 cepas se identificaron 5,961 agrupamientos de genes que constituyen la totalidad del pangenoma. A partir de la presencia/ausencia de genes en el pangenoma se construyó la filogenia mostrada en la Figura 13, en la cual es posible observar clados que agrupan a cepas altamente similares. El análisis con Roary permitió clasificar el contenido del pangenoma en 4 grandes grupos: a) genes *core*, contenidos en el 100% de los genomas, b) genes *soft-core*, encontrados en el 95 a 99% de los genomas, c) genes *shell*, observados en el 15 al 95% de los genomas incluidos y d) genes *cloud*, contenidos en el 15 al 1% de los genomas. En este estudio se identificaron 2,231 genes *core*, 1,182 genes *soft-core*, 466 genes *shell* y 2081 genes *cloud* (Figura 14). De los 2,231 genes que conforman el genoma núcleo (*core*), 284 corresponden a proteínas hipotéticas o de función desconocida, mientras que 1,947 son genes esenciales asociados con la replicación y reparación de ADN, traducción y regulación proteica, síntesis de ácidos micólicos, regulación transcripcional, virulencia y catabolismo, entre otras funciones.

El pangenoma construido corresponde a un pangenoma abierto, como se puede observar en la Figura 15. Mientras aumenta el número de genomas

incorporados al pangenoma, el número de genes conservados comienza a mantenerse, mientras que el número total de genes sigue aumentando. Se dice que es un pangenoma próximo a cerrarse cuando la línea que representa el número total de genes ya no aumenta significativamente con la adición de nuevos genomas, y esta empieza a aplanarse, fenómeno observado en este pangenoma.

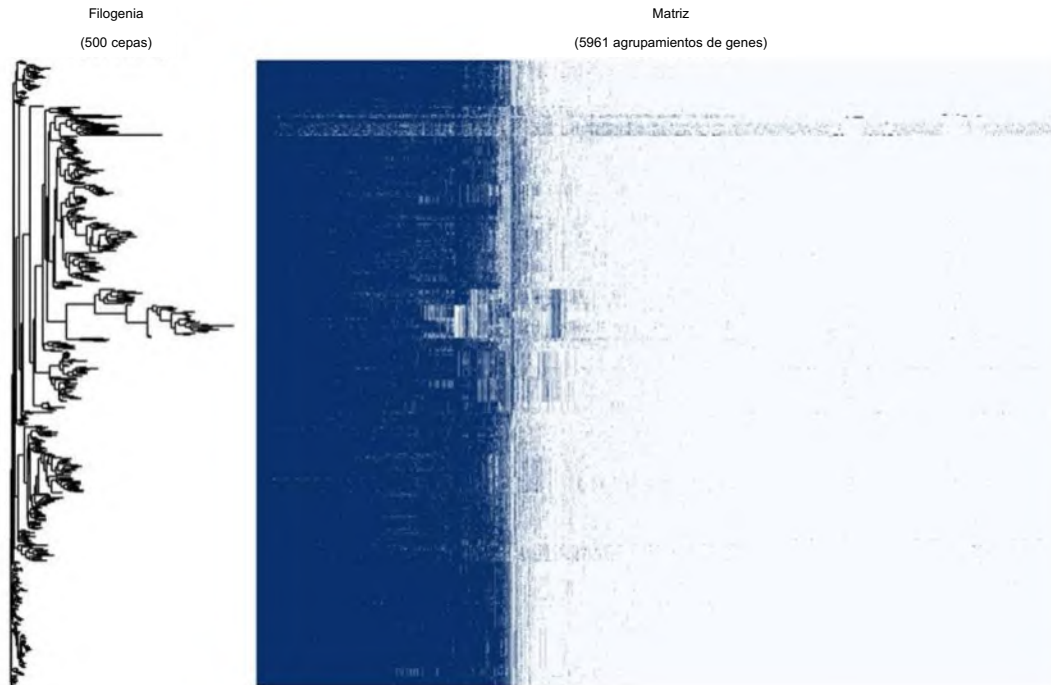


Figura 13. Construcción filogenética realizada por Roary para las cepas de estudio a partir del pangenoma.

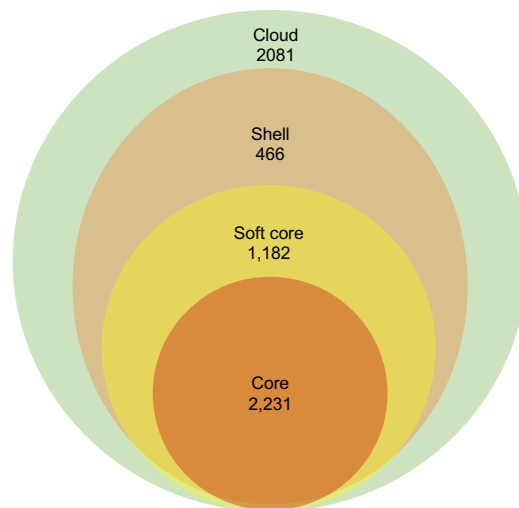


Figura 14. Distribución de los genes que conforman el pangenoma. Los genes *core* están contenidos en el 100% de los genomas, los genes *soft-core* en el 95 a 99%, los genes *shell* en el 15 al 95% y los genes *cloud* en el 1 al 15% de los genomas.

Mediante la búsqueda de genes exclusivos de las cepas pertenecientes al fenotipo EPTB con `query_pan_genome`, se identificaron 2 genes que codifican proteínas exclusivas de este fenotipo: *pstAq1* (proteína transportadora de fosfatos) y el gen de una proteína hipotética; sin embargo, no fue posible establecer una asociación estadística debido a que están presentes en una baja proporción de las cepas EPTB (<15 cepas). Un fenómeno similar se observó para las cepas PTB, en donde se identificaron genes codificantes de proteínas exclusivas *devS1* (regulador de histidina quinasa) y *mtr1* (micotiol reductasa), de igual manera, no fue posible establecer una asociación estadística debido a que están presentes en una baja proporción de las cepas PTB.

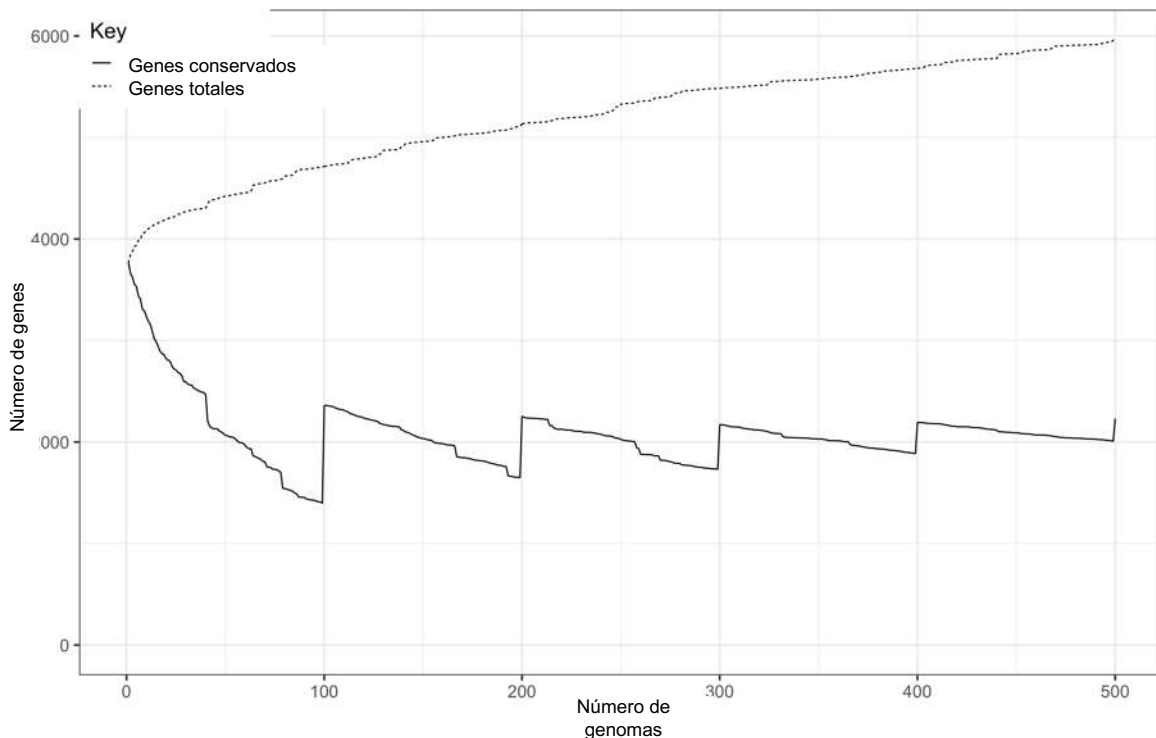


Figura 15. Curvas representativas del contenido de genes en el pangenoma.

En el caso del análisis de genes compartidos con Scoary, se encontró que la presencia del gen *afc* mantiene una relación estadísticamente significativa con el fenotipo EPTB ($p < 0.0001$, OR: 10.5), mientras que la presencia de un gen

codificante para una proteína hipotética se encuentra asociado con el grupo PTB ($p < 0.0001$, OR:10.2).

Mediante un análisis de la distribución de genes de virulencia en el pangenoma accesorio de los fenotipos EPTB y PTB, y su posterior análisis estadístico, se determinó la asociación del fenotipo PTB con la delección de los genes de virulencia *hspR* y *plcD*. Adicionalmente, mediante este análisis se observó que la delección de los genes *pe_pgrs5*, *pe_pgrs25*, *Rv1759c*, *pe_pgrs57*, *Rv817c*, *Rv2818c*, *Rv2816c*, *Rv3740c* y *Rv2550c* en los genomas de MTB se encuentran asociados al fenotipo PTB (Tabla 15). Por su parte, la delección de los genes de virulencia *aceA*, *plcA*, *esxR* y *Rv2275* se asocian con el fenotipo EPTB, además del gen *ppe50*.

Tabla 15. Genes asociados a los fenotipos PTB y EPTB mediante el análisis del genoma accesorio. OR, odds ratio; IC, intervalo de confianza; PTB tuberculosis pulmonar; EPTB, tuberculosis extrapulmonar.

	Gen	Producto	Clasificación	Valor p	OR	IC 95%	Asociación
Factores de virulencia	<i>Rv0353/hspR</i>	Proteína reguladora HspR	Virulencia	0.0483	1.5	1.0034 to 2.4059	PTB
	<i>Rv1755c/plcD</i>	Fosfolipasa C	Virulencia	0.0006	2	1.3487 to 2.9849	PTB
	<i>Rv1790/ppe27</i>	PPE27	Virulencia	0.0540	1.48	0.9933 to 2.2193	
	<i>Rv1915/aceA</i>	Isocitrato liasa	Virulencia	0.0005	2.03	1.3606 to 3.0357	EPTB
	<i>Rv2351c/plcA</i>	Fosfolipasa C	Virulencia	0.007	2.78	1.3136 to 5.9075	EPTB
	<i>Rv3019c/esxR</i>	EsxR	Virulencia	0.0132	1.6	1.1010 to 2.2783	EPTB
	<i>Rv2275</i>	Cyclo(L-tyrosyl-L-tyrosyl) sintasa	Virulencia	0.001	2.9	1.5405 to 5.5661	EPTB
Otros genes	<i>Rv3135/ppe50</i>	PPE50	Pe/ppe	< 0.0001	6.4	3.7599 to 10.7484	EPTB
	<i>Rv0297/pe_pgrs5</i>	PE PGRS5	Pe/ppe	0.0303	1.5	1.0419 to 2.2758	PTB
	<i>Rv1396c/pe_pgrs25</i>	PE PGRS25	Pe/ppe	0.0255	1.6	1.0553 to 2.2802	PTB
	<i>Rv1759c/wag22</i>	Transposasa	Pe/ppe	< 0.0001	2.4	1.6427 to 3.4220	PTB
	<i>Rv3514/pe_pgrs57</i>	PE PGRS57	Pe/ppe	0.0201	1.5	1.0679 to 2.1640	PTB
	<i>Rv1979c</i>	Permeasa transportadora tipo ABC	Procesamiento de pared celular	0.0735	1.3	0.9699 to 1.962	
	<i>Rv2819c</i>	Proteína RAMP Csm5	Proteínas conservadas	0.0396	1.4	1.0178 to 2.0610	PTB
	<i>Rv2817c</i>	Endonucleasa asociada a CRISPR Cas1	Proteínas conservadas	0.0396	1.4	1.0178 to 2.0610	PTB
	<i>Rv2818c</i>	Proteína asociada a CRISPR Csm6	Proteínas conservadas	0.0317	1.5	1.0344 to 2.0955	PTB
	<i>Rv2816c</i>	Endoribonucleasa asociada a CRISPR Cas2	Proteínas conservadas	0.0394	1.4	1.0182 to 2.0635	PTB
	<i>Rv3515c/fadD19</i>	Acil coenzima sintasa	Metabolismo de lípidos	0.0663			
	<i>Rv3740c</i>	Diacilglicerol aciltransferasa	Metabolismo de lípidos	0.0266	1.5	1.0516 to 2.2586	PTB
	<i>Rv2550c</i>	Antitoxina VapB20	Virulencia	0.0001	2.8	1.6896 to 4.5763	PTB

DISCUSIÓN

En hecho de que el pangenoma de MTB sea abierto le permite enriquecer su reserva genética con genes nuevos con el objetivo de conferirle características evolutivas en términos de patogenicidad. Sin embargo, se considera que este es un pangenoma próximo a cerrarse, lo que es congruente con la idea general de que el genoma de MTB es altamente conservado. Esta observación se corrobora con los resultados generados, ya que un genoma de MTB posee alrededor de 4,000 genes, y en este trabajo se identificaron 2,231 genes core que corresponden al 55.77% del genoma total. Los valores de distribución de los genes pertenecientes al genoma core (2,231) y genoma accesorio (3,729) son comparables con algunos de los pangenomas previamente publicados, pero en otros casos no es posible comparar, debido a que se utilizan diferentes parámetros para su construcción. Por ejemplo, en 2018 se construyó un pangenoma que incluye 1,595 cepas de MTB, identificándose 11,039 agrupamientos de los cuales 3,419 corresponden al genoma core y 7,620 al genoma accesorio (Kavvas et al., 2018); un número muy por encima de los valores encontrados en este trabajo. Las diferencias observadas pueden deberse tanto al número de cepas incluidas, como a la utilización de un porcentaje menor (80%) de similitud en la secuencia para la construcción del pangenoma. Por otro lado, existen pangenomas en donde la distribución de los genes es similar, como en el caso del pangenoma construido a partir de 183 cepas pertenecientes al complejo *Mycobacterium tuberculosis* (Zakham et al., 2021), en donde se reporta un core genoma de 1,166 genes y un genoma accesorio de 5,870 genes. Por lo que la distribución de genes en el pangenoma está en función de la cantidad y naturaleza de las cepas incluidas.

La ausencia de genes exclusivos de un fenotipo puede explicarse debido al alto grado de conservación de los genomas de MTB mencionado previamente. Esta variabilidad genética está en función entre otras cosas, de la nula adquisición de material genético mediante transferencia horizontal. Es de resaltar que el grado de diversidad genética dentro del complejo *Mycobacterium tuberculosis* ha sido un tema de debate. Los primeros estudios basados en genes específicos revelaron

baja diversidad en la secuencia de ADN y condujeron a la noción de que la variabilidad de cepas en MTB era insignificante y sin importancia clínica. Sin embargo, estudios posteriores que utilizaron un mayor número de marcadores genéticos demostraron la presencia de diferencias genéticas entre cepas de MTB aisladas de distintas regiones geográficas. Actualmente el análisis genómico comparativo basado en la secuenciación de genomas completos ha revelado diferencias inclusive entre dos cepas que difieren en hasta 2,000 polimorfismos de un solo nucleótido (SNP), una distancia genética que es comparable a la distancia entre especies de *M. tuberculosis* y *M. bovis* (Galagan, 2014). Aunque esta diversidad es sustancialmente menor que la observada en otras especies bacterianas, existe una cantidad de evidencia cada vez mayor de que puede tener consecuencias funcionales y clínicas.

Por otra parte, los procesos de recombinación genética pueden llevar a la presentación de diferencias en el contenido de genes compartidos por los fenotipos EPTB/PTB, como se observó en este trabajo respecto al gen *aftC*. Este gen, asociado a EPTB, participa en el metabolismo del arabinano, necesario para la síntesis de arabinogalactanos, una de los principales polisacáridos de la pared celular micobacteriana (Zhang et al., 2013). La delección del gen *aftC* en las cepas ocasiona una disminución en la cantidad de arabinogalactanos, y por tanto de ácidos micólicos, presentando una permeabilidad incrementada con menor resistencia a estrés. Otros genes en los que se encontraron diferencias en la proporción y que se asociaron a los fenotipos de estudio fueron los genes de virulencia *hspR*, *plcD*, *ppe27*, *aceA*, *plcA*, *esxR* y *Rv2275*. En un pangenoma construido para 1,595 aislados mediante *machine learning*, se identificó que el 93% de los genes de virulencia de MTB se encuentran en el genoma core, entre el 7% restante que se encuentra en el genoma accesorio están los genes *plcA*, *plcD* y *aceA*, identificados en este trabajo también como parte del genoma accesorio, lo que resalta su posible papel en la adquisición de ventajas en términos de patogenicidad. Los genes *plcA* y *plcD* codifican para una fosfolipasa C, enzima involucrada en la utilización de lípidos presentes en la membrana de la célula hospedadora para la producción de ácido araquidónico. El ácido araquidónico es un

inductor de necrosis que a su vez induce la respuesta proinflamatoria (Le Chevalier et al., 2015). Por lo que si se altera la función de esta proteína, podría ocasionar una activación de la apoptosis y con ello un aumento en la posibilidad de diseminación bacilar por las células dendríticas. Es de resaltar que la manipulación de la muerte celular se ha descrito previamente como una estrategia para modular la respuesta inmune.

Otras posibles implicaciones de los genes asociados a los fenotipos de estudio incluyen a la proteína HspR, un modulador transcripcional que reprime la transcripción de genes codificantes de proteínas de choque térmico. Las alteraciones en estas proteínas influyen en los mecanismos de virulencia, disminuyendo la habilidad de la bacteria para establecer una infección crónica. El gen *Rv1759c* codifica para un antígeno de la familia pe_pgrs, con actividad de fibronectina (*wag22*). Esta proteína está involucrada en el mantenimiento de la infección latente en un modelo de infección crónica (Campuzano et al., 2007) e induce la producción de niveles altos de interferón gamma. Todas las funciones de estos genes en los que se observan deleciones podrían estar alteradas, ocasionando modificaciones en la respuesta inmune generada.

CONCLUSIÓN

Si bien no se lograron identificar genes exclusivos de las cepas pertenecientes a los fenotipos PTB y EPTB, si se encontró una relación estadísticamente significativa ente la distribución de genes intrafenotipos derivados del análisis del pangenoma accesorio. Estas características genotípicas pueden ser las responsables en parte, de la aparición de las distintas presentaciones clínicas de la tuberculosis.

REFERENCIAS

- Assefa, S., Keane, T. M., Otto, T. D., Newbold, C., & Berriman, M. 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics*, 25(15): 1968–1969. <https://doi.org/10.1093/bioinformatics/btp347>
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Pribelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology : a journal of computational molecular cell biology*, 19(5): 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Brynildsrud, O., Bohlin, J., Scheffer, L., & Eldholm, V. 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome biology*, 17(1): 238. <https://doi.org/10.1186/s13059-016-1108-8>
- Campuzano, J., Aguilar, D., Arriaga, K., León, J. C., Salas-Rangel, L. P., González-Merchand, J., Hernández-Pando, R., & Espitia, C. 2007. The PGRS domain of *Mycobacterium tuberculosis* PE_PGRS Rv1759c antigen is an efficient subunit vaccine to prevent reactivation in a murine model of chronic tuberculosis. *Vaccine*, 25(18): 3722–3729. <https://doi.org/10.1016/j.vaccine.2006.12.042>
- Coscolla, M. 2017. Biological and Epidemiological Consequences of MTBC Diversity. *Advances in experimental medicine and biology*, 1019: 95–116. https://doi.org/10.1007/978-3-319-64371-7_5
- Faksri, K., Xia, E., Ong, R. T., 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>
- Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23): 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>
- Galagan, J. E. 2014. Genomic insights into tuberculosis. *Nature reviews. Genetics*, 15(5): 307–320. <https://doi.org/10.1038/nrg3664>
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8): 1072–1075. <https://doi.org/10.1093/bioinformatics/btt08>
- 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): 4306. <https://doi.org/10.1038/s41467-018-06634-y>
- Kohl, T.A., Utpatel, C., Schleusener, V., De Filippo, M.R., Beckert, P., Cirillo, D.M., Nieman, S. 2018. MTBseq: a comprehensive pipeline for whole genome sequence analysis of *Mycobacterium tuberculosis* complex strains. *PeerJ* 6: e5895 <https://doi.org/10.7717/peerj.5895>.
- Le Chevalier, F., Cascioferro, A., Frigui, W., Pawlik, A., Boritsch, E. C., Bottai, D., Majlessi, L., Herrmann, J. L., & Brosch, R. 2015. Revisiting the role of phospholipases C in virulence and the lifecycle of *Mycobacterium tuberculosis*. *Scientific reports*, 5: 16918. <https://doi.org/10.1038/srep16918>
- Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., 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>
- Rodríguez, D., Zapata, A., Molina, F., Pemuolo, G., Burbano, J., & Correa, L. 2021. Meningeal tuberculosis: cases report years 2005-2017. *Revista Chilena de Infectología: Organo Oficial de la Sociedad Chilena de Infectología*, 38(3): 410-416.

- Ruesen, C., Chaidir, L., van Laarhoven, A., Dian, S., Ganiem, A. R., Nebenzahl-Guimaraes, H., Huynen, M. A., Alisjahbana, B., Dutilh, B. E., & van Crevel, R. 2018. Large-scale genomic analysis shows association between homoplastic genetic variation in *Mycobacterium tuberculosis* genes and meningeal or pulmonary tuberculosis. *BMC genomics*, 19(1): 122. <https://doi.org/10.1186/s12864-018-4498-z>
- Saw, S. H., Tan, J. L., Chan, X. Y., Chan, K. G., & Ngeow, Y. F. 2016. Chromosomal rearrangements and protein globularity changes in *Mycobacterium tuberculosis* isolates from cerebrospinal fluid. *PeerJ*. 4: e2484. <https://doi.org/10.7717/peerj.2484>
- Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14): 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
- Yang, T., Zhong, J., Zhang, J., Li, C., Yu, X., Xiao, J., Jia, X., Ding, N., Ma, G., Wang, G., Yue, L., Liang, Q., Sheng, Y., Sun, Y., Huang, H., & Chen, F. 2018. Pan-Genomic Study of *Mycobacterium tuberculosis* Reflecting the Primary/Secondary Genes, Generality/Individuality, and the Interconversion Through Copy Number Variations. *Frontiers in microbiology*, 9: 1886. <https://doi.org/10.3389/fmicb.2018.01886>
- Zakham, F., Sironen, T., Vapalahti, O., & Kant, R. 2021. Pan and Core Genome Analysis of 183 *Mycobacterium tuberculosis* Strains Revealed a High Inter-Species Diversity among the Human Adapted Strains. *Antibiotics*, 10(5): 500. <https://doi.org/10.3390/antibiotics10050500>
- Zhang, J., Angala, S. K., Pramanik, P. K., Li, K., Crick, D. C., Liav, A., Jozwiak, A., Swiezewska, E., Jackson, M., & Chatterjee, D. 2011. Reconstitution of functional mycobacterial arabinosyltransferase AftC proteoliposome and assessment of decaprenylphosphorylarabinose analogues as arabinofuranosyl donors. *ACS chemical biology*, 6(8): 819–828. <https://doi.org/10.1021/cb200091m>

DISCUSIÓN GENERAL

La TB es una enfermedad antigua, pero no una enfermedad del pasado. Después de desaparecer de la agenda de salud pública mundial en las décadas de 1960 y 1970, la TB regresó a principios de la década de 1990 por varias razones, incluida la aparición de la pandemia del VIH/SIDA y el aumento de la resistencia a los medicamentos. En el contexto actual de la pandemia por COVID-19, debido a los confinamientos y otras restricciones a la movilidad, los pacientes y otras personas potencialmente contagiadas dejaron de acudir a los centros de salud a solicitar una consulta médica o a dar seguimiento a su tratamiento antifímico. La consecuencia es que, según la OMS, ahora existen alrededor de 4.1 millones de personas que padecen TB que no han sido diagnosticadas. Se cree que más del 40% de casos de TB no se diagnosticaron en 2020, ocasionando un aumento en la tasa de transmisión que se verá reflejada durante los próximos años (OMS,2021).

Para lograr un avance en el control de la TB, es necesario un correcto seguimiento epidemiológico de las cepas circulantes, tanto global como localmente. En México, el presente trabajo forma parte de los escasos estudios que llevan a cabo la tipificación de cepas de MTB con métodos innovadores como el uso de *whole genome sequencing* (WGS) y el análisis *in silico*. En esta investigación se abordaron cuatro aspectos principales para el estudio de la TB haciendo uso del análisis de WGS: i) la genotipificación mediante la identificación de linajes, sublinajes y espoligotipos, ii) la identificación de SNPs asociados con resistencia a fármacos, iii) la identificación de macro y micro variantes genómicas y iv) la construcción del pangenoma. Los resultados mostrados revelaron que las cepas secuenciadas pertenecen al linaje 4. Estudios previos realizados mediante espoligotipificación y MIRU-VNTR en nuestro país han demostrado una gran diversidad de linajes circulantes; sin embargo, el linaje 4 es el reportado con mayor frecuencia (Juaréz-Eusebio et al., 2017). La presencia mayoritaria de cepas del L4 se debe posiblemente a los movimientos migratorios que ocurrieron de Europa hacia América durante el proceso de colonización. Además, una de las principales características de este linaje es que posee una amplia distribución geográfica, lo

cual refleja el gran éxito evolutivo en adaptación y dispersión a diferentes poblaciones humanas, a diferencia de otros linajes (Stucky et al., 2016).

El 40% de las cepas MYC fueron agrupadas dentro de la familia LAM. En un estudio en el que se analizaron aislados pertenecientes a esta familia, se observó una diferencia significativa en la velocidad de crecimiento en medio de cultivo, característica relacionada con la capacidad de reproducción bacteriana, lo que genera una mayor carga bacilar pudiendo promover una mayor dispersión de la familia LAM (Von Groll et al., 2010). Así, se esperaría aislar con mayor frecuencia a integrantes de dicha familia en cepas pulmonares en comparación con las cepas extrapulmonares, esta tendencia no fue observada en las cepas de la colección MYC (50% cepas PTB, 50% cepas EPTB).

Por otro lado, este trabajo permitió identificar diferencias entre el número de secuencias de inserción IS6110 esperadas según el sublinaje y lo observado para las cepas MY036 (LAM), MYC052(LAM), MYC061(Harleem), MYC072(LAM) y MYC096 (S). Aunque se cree que un alto número de copias de IS6110 proporciona una ventaja selectiva a las cepas que los poseen, la resistencia antibióticos y la aparición de brotes también se han asociado a cepas con un bajo número de copias (McEvoy et al., 2007). Interesantemente, las cepas MYC con menos de 6 copias de IS6110 presentan resistencia fenotípica a al menos un antibiótico de primera línea. La interrupción de las secuencias de inserción IS6110 puede ocasionar distintas alteraciones funcionales. En el caso de las cepas MYC la inserción de IS6110 en el genoma ocasionó la interrupción de los genes *Rv2813-Rv2816c*, *Rv0402c* y *Rv0403c* que podría estar jugando un papel importante en la transmisión local del genotipo 4.1.1.3 o X3.

Cabe resaltar que la cepa MYC093, es la única del clado X3 que no posee ninguna interrupción de los genes antes mencionados, y esto coincide con la formación de cuerdas macroscópicas después del cultivo. Este comportamiento “atípico” de las cepas aisladas en Michoacán ya ha sido observado previamente por nuestro grupo de trabajo, en diversos análisis de secuencias de genes involucrados en la vía de síntesis de ácidos micólicos, de resistencia a antibióticos y de la familia *pe*, *ppe* y *pgrs*, encontrándose un inusual alto número de mutaciones en

comparación con cepas aisladas de otras regiones geográficas (Frutos-Hernandez, 2015; Negrete-Paz, 2018). Estas inconsistencias también son observables en la comparación genotípica-fenotípica de resistencia a antibióticos de las cepas MYC036, MYC072, MYC075, MYC078, MYC087 y MYC136. En ellas se puede ver un fenotipo resistente; sin embargo, no existe ninguna mutación en genes asociados a resistencia a antibióticos.

Los fenotipos característicos de estas cepas MYC pueden estar determinados por la presencia de ciertas variaciones en su genoma. Tras el análisis para la búsqueda de macrovariantes, se identificaron 9 regiones homólogas presentes en la totalidad de las cepas MYC, pero ausentes en la cepa de referencia H37Rv. De éstas, 5 incluyen genes que codifican proteínas hipotéticas, 3 corresponden a regiones no codificantes y 1 a una variación dentro del gen *fadD2*. En comparación con las cepas obtenidas de bases de datos públicas, la longitud de las cepas MYC se encuentra cercana a la media de la longitud de los genomas de las cepas de BD (4491647 pb), lo cual confirma el alto grado de conservación del genoma de MTB.

La búsqueda de asociaciones entre el fenotipo (PTB, EPTB) con el genotipo (linaje o sublinaje), permitió relacionar a los sublinajes 1.1.1, 1.2.1, 4.1.2.1 y ancestrales Beijing con la forma extrapulmonar de la TB y a los sublinajes 4.3.1, 4.3.3, 4.5, Asian African 2 y Europe/Russia B0/W148 con la forma pulmonar en un conjunto de datos obtenidos de bases de datos públicas. Aunque la cepa MYC61 (única perteneciente a estos sublinajes) que fue clasificada filogenéticamente como 4.1.2.1 es de origen pulmonar, posee las variaciones genómicas observadas con mayor frecuencia en las cepas extrapulmonares, resaltando el hecho de que la pertenencia a un grupo o familia no explica completamente el fenotipo pulmonar o extrapulmonar. Un aspecto importante que podría explicar este fenómeno es la variación genómica entre las cepas PTB y EPTB. Las variaciones genéticas diferenciales (SNPs) entre los fenotipos EPTB y PTB se encuentran mayoritariamente en genes pertenecientes a la familia PE/PPE/PGRS. El genoma de MTB codifica las proteínas PE/PPE/PE_PGRS, proteínas intrínsecamente desordenadas, redundantes y de naturaleza antigénica. Éstas realizan múltiples

funciones que intensifican la virulencia de MTB principalmente mediante la modulación de las respuestas inmunitarias, lo que afecta la eliminación del patógeno mediada por el sistema inmunitario. La naturaleza altamente repetitiva, redundante y antigénica de las proteínas PE/PPE/PE_PGRS proporciona una ventaja crítica sobre otras proteínas de MTB en términos de impartir un mayor nivel de virulencia y también como moléculas señuelo que enmascaran el efecto de las moléculas efectoras, modulando así la respuesta inmune (Sharma et al., 2021).

De esta manera es posible hipotetizar que dichos SNPs pueden ocasionar una alteración en la función de las proteínas codificadas al modificar la secuencia de aminoácidos, provocando variaciones tanto en la respuesta inmune como en los mecanismos de respuesta a estrés que se encuentran asociados a las distintas presentaciones clínicas de la tuberculosis. Interesantemente, al realizar el análisis de genes diferenciales de los fenotipos PTB y EPTB mediante la construcción del pangenoma y análisis del pangenoma accesorio, se encontraron de manera similar genes pertenecientes a la familia PE/PPE/PE_PGRS como *ppe27* y *Rv1759c* que codifica para un antígeno de la familia *pe_pgrs*, con actividad de fibronectina (*wag22*). Esta proteína está involucrada en el mantenimiento de la infección latente en un modelo de infección crónica (Campuzano et al., 2007) e induce la producción de niveles altos de interferón gamma. Resaltando de nueva cuenta la posible modificación de la respuesta inmune generada.

Por todo lo anterior, en el presente trabajo se sentaron las bases para un estudio experimental más detallado de los genes que poseen variaciones genéticas asociadas a los fenotipos PTB y EPTB, así como de las consecuencias fenotípicas y funcionales de las diferencias proporcionales de genes de virulencia asociados a las distintas presentaciones clínicas de la TB.

CONCLUSIONES GENERALES

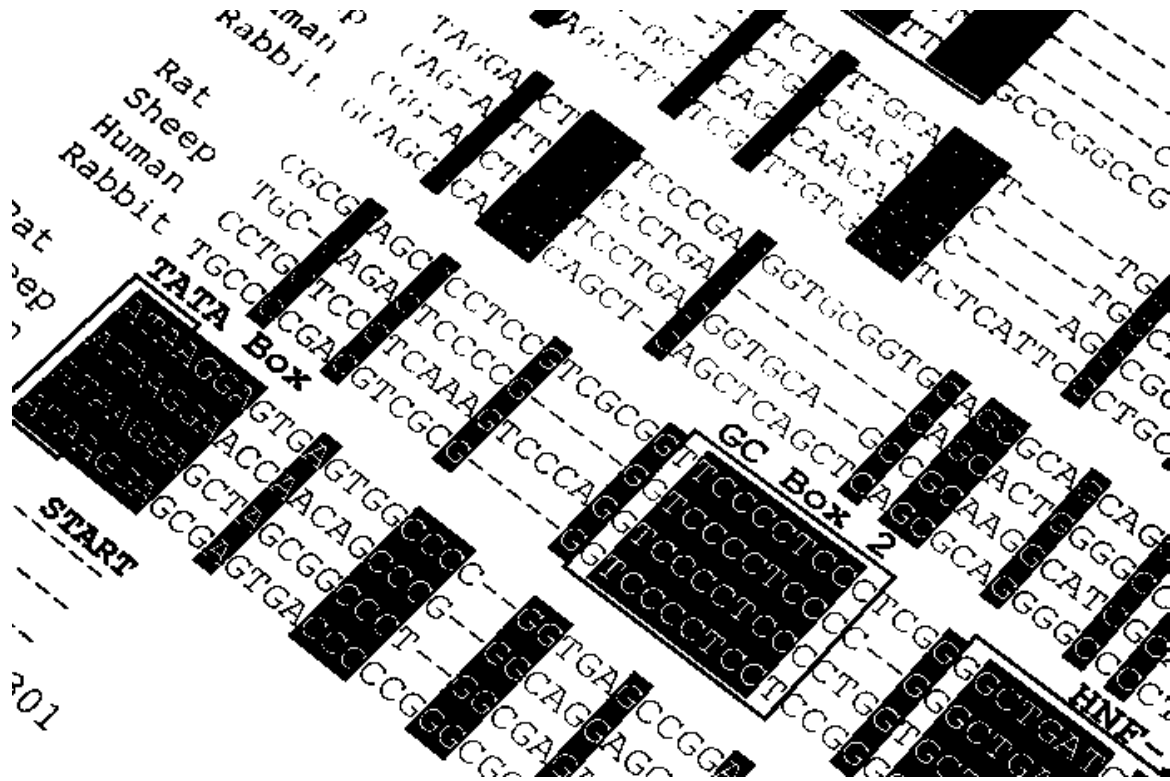
La aparición de un fenotipo clínico de tuberculosis no está determinada por un elemento genético exclusivo o único, sino por un conjunto de variaciones que pueden desencadenar diferencias en la respuesta inmune. Dichas variaciones se encuentran principalmente en genes de la familia PE/PPE/PGRS a los cuales se les ha descrito previamente como una de las principales fuentes de variación antigénica. Sin embargo, debido a que se desconoce la función de muchas de las proteínas codificadas por estos genes, es necesaria la realización de estudios que ayuden a dilucidar estas funciones y su asociación específica con los fenotipos pulmonar o extrapulmonar.

CAPÍTULO V
ARTÍCULOS DE DIVULGACIÓN

TECNOLOGÍA

Secuenciación genómica: Cuando el futuro nos alcance

Andrea Monserrat Negrete Paz y Ma. Soledad Vázquez Garcidueñas



Andrea Monserrat Negrete Paz, estudiante del Programa Institucional de Doctorado en Ciencias Biológicas, en la Opción de Biotecnología Molecular.
anegrete.pz@gmail.com

Ma. Soledad Vázquez Garcidueñas, profesora investigadora en la División de Estudios de Posgrado de la Facultad de Ciencias Médicas y Biológicas "Dr. Ignacio Chávez" de la Universidad Michoacana de San Nicolás de Hidalgo.
marisolvaz@yahoo.com

Actualmente se han secuenciado de manera parcial o completa, alrededor de 15 mil especies en nuestro planeta, debido al abaratamiento de los costos de secuenciación. Pero ¿qué tan cercana es la realidad de poder secuenciar nuestro genoma, no a nivel de especie sino de manera individual? ¿Qué beneficios tendríamos si pudiéramos llegar a hacerlo?

A caray... Eso si me interesa

Para responder estas preguntas es necesario conocer al protagonista: El genoma, éste es el con-

junto de toda la información genética contenida en el ADN e incluye las instrucciones necesarias para que los organismos desarrollen las características que los hacen únicos, sin importar si se trata de una bacteria, una planta, un animal o una persona. En los seres humanos, el genoma consta de 23 pares de cromosomas que se encuentran en el núcleo, así como un pequeño cromosoma que se encuentra en las mitocondrias de las células. Cada conjunto de 23 cromosomas contiene aproximadamente 3.2 mil millones de pares de bases, es decir dos nucleótidos opuestos y complementarios en la cadena de ADN. A pesar de ser secuencias tan largas, se trata de estructuras pequeñísimas e invisibles para los ojos. Para obtener esas pequeñas estructuras a partir de un organismo, los científicos utilizan técnicas tanto experimentales como de análisis computacional.

1. En primer lugar, se extrae el ADN de las células del organismo del cual nos interesa obtener su secuencia genómica en un tubo especial de laboratorio, mediante una serie de pasos que nos permiten eliminar todo lo que no se trate de ADN.
2. Después hay que cortarlo en pedazos, como si cortáramos una tira larga de papel en muchas partes y una vez que tenemos los pedacitos hay que hacer múltiples copias de cada uno.
3. Con este método de copiado, se obtienen conjuntos de millones de copias del mismo pedazo de ADN.

4. Después, cada una de estas pequeñas secuencias es leída -en un secuenciador- para saber el orden correcto que siguen las bases y así obtener lecturas de cada pedazo de ADN.

Si queremos darnos una idea de cuántas lecturas obtendríamos, tenemos que dividir el tamaño del genoma que hemos secuenciado entre el tamaño de la lectura que realiza el secuenciador. Por ejemplo, si utilizamos un secuenciador "Illumina MiSeq" que realiza lecturas de 300 pares de bases para secuenciar el genoma humano de 3.2 mil millones de pares de bases, obtendríamos cerca de 10.3 millones de lecturas. ¿Y... ahora qué sigue?

Armando el rompecabezas

Una vez que tenemos estas lecturas, necesitamos encontrar su orden, ya que si recuerdas, partimos la cadena de ADN en muchos pedazos; es como si tuviéramos que armar un rompecabezas de más de ¡10 millones de piezas! Suena bastante complicado y si además agregamos que puede haber muchas piezas repetidas en este rompecabezas, que es nuestro genoma, podríamos pensar que es una tarea imposible. Sin embargo, existen diversos programas de computadora que nos ayudan a resolver la tarea de encontrar la mejor opción en el armado de esta secuencia de ADN.

Al final de este proceso tenemos nuestro genoma secuenciado, pero si queremos saber a qué



“se dedica” cada porción de información en el genoma ¿cómo podemos hacerlo? Pues habría que “anotar” ese genoma, con anotar nos referimos a que debemos leer el genoma secuenciado, esto no sólo implica conocer el número de genes que posee, sino también muchas características específicas de cada gen, como por ejemplo su tamaño, su posición específica en el genoma, la función que desempeña, si se encuentra solo o quiénes son sus genes vecinos, de esta manera podemos tener toda la información descifrada.

¿Una herramienta para todos?

Si nos preguntamos ¿qué tienen en común la actriz Glenn Close, el cantante y vocalista de la banda Black Sabbath, Ozzy Osbourne y James D. Watson, uno de los considerados co-descubridores de la estructura de doble hélice del ADN? la respuesta podría sorprendernos un poco, ya que fueron de los primeros seres humanos en obtener su genoma secuenciado completamente. Mientras que el primer genoma humano costó cerca de 3 mil millones dólares y tardó más de una década en producirse, en 2010 la compañía de secuenciación estadounidense “Illumina” secuenció el genoma de la actriz Glenn Close por 300 mil dólares.

Actualmente, el Instituto Nacional de Investigación del Genoma Humano, ha estimado que obtener la secuencia completa de un genoma humano tiene un costo de aproximadamente 1000 dólares, aunque en noviembre de 2018, se anunció el desarrollo de un nuevo secuenciador por parte de la empresa “Illumina”, considerada como la empresa líder en servicios de secuenciación a nivel mundial, lo que significaría una reducción en el costo de secuenciación hasta casi el 10% del costo actual, es decir ¡100 dólares!.

Con el desarrollo de esta tecnología de secuenciación a bajo costo, podríamos obtener nuestra secuencia genómica por el precio de un teléfono celular.

Pero si ya se secuenció el genoma de James Watson, ¿para qué secuenciar el mío?

No existe un genoma humano único, cada individuo tiene sus propias particularidades y cambios frente a lo que podríamos considerar una secuencia “consenso”, es decir, si comparáramos todos nuestros genomas, habría bases en la secuencia que aparecerían con mayor frecuencia en cada posición, pero algunos de nosotros podríamos tener una base distinta –por ejemplo en la posición 16 de nuestro genoma, la mayoría tendrá una adenina pero habrá algunos otros que tengan una citosina- por lo que podríamos obtener una secuencia “consenso” o representativa del genoma humano basada en las frecuencias.

Estas pequeñas diferencias representan el 0.01% de nuestro genoma, ya que se calcula que todos los seres humanos compartimos más de un 99.9 % de similitud en el mismo. Son esas variantes las que dan sentido al estudio de un genoma personal. La forma en que actúan dichas variantes puede ser muy diversa ya que un cambio en una base de la secuencia de ADN puede inducir una modificación en la función de la proteína, pueden cambiar la expresión de ciertos genes o incluso pueden afectar a otros mecanismos aún desconocidos. Esto tiene múltiples consecuencias, como la predisposición a contraer determinadas enfermedades, como el cáncer, el Parkinson o la diabetes: también indican qué tratamientos aceptará mejor nuestro organismo, lo que ayuda a diseñar fármacos más efectivos, y qué dietas nos convienen. Incluso reflejan si somos o no propensos al alcoholismo o a tener mellizos. Qué interesante ¿verdad?

«Podemos estar seguros que la secuenciación de genomas personales, es la base de la medicina personalizada: podría ayudarnos a definir riesgos, predecir respuestas, escoger o mejorar tasas de éxito en el tratamiento, entre muchas otras aplicaciones, pero al menos por el momento tendremos que esperar un poco hasta que secuenciar nuestro genoma sea tan común como sacarse una radiografía o tomarse una resonancia magnética»



Estudios de asociación en todo el genoma. (2019). National Human Genome Research Institute. <https://www.genome.gov/27562846/estudios-de-asociacion-en-todo-el-genoma/>
Reguero-Reza, M.T. (2013). 10 años de la secuenciación del genoma humano: Encuentro entre el imaginario y la realidad. Revista Colombiana de Biotecnología. 15(1), 5-7.

Bermudez-Santana, C. I. (2016). Aplicaciones de la bioinformática en la medicina: el genoma humano. ¿Cómo podemos ver tanto detalle?. Acta Biológica Colombiana. 21(1), 249-258.

Herránz, A. S., Escrig, A. J., Gobernado, I. (2012). Secuenciación de genoma completo: un salto cualitativo en los estudios genéticos. Revista de neurología. 54(11), 692-698.

Las enfermedades desatendidas y el círculo vicioso de la pobreza

Andrea Montserrat Negrete Paz

Centro Multidisciplinario de Estudios en Biotecnología, Facultad de Medicina Veterinaria y Zootecnia, Universidad Michoacana de San Nicolás de Hidalgo. Morelia, Mich., México.
Contacto: anegrete.pz@gmail.com

Resumen. Casi una sexta parte de la población mundial enferma y a menudo fallece todos los años a causa de enfermedades que pueden tratarse o prevenirse mediante una vacuna o la administración de un solo medicamento, pero en poblaciones rurales y agrícolas o en situación de pobreza el poco acceso a los servicios de salud junto con otros factores socioeconómicos dificulta esta tarea. Afecciones milenarias como la lepra, tan comunes como la enfermedad de Chagas o prevenibles como las lombrices intestinales pertenecen a este grupo de enfermedades conocidas como desatendidas. Estas enfermedades representan una carga enorme en las comunidades y las personas en situación de pobreza porque generan una pérdida de ingresos y el estigma crónico y la discriminación asociados con sus secuelas. La pobreza no es solo una cuestión socioeconómica, es también un problema de salud de primer orden.

Palabras clave: Chagas, lepra, geohelmintiasis.

La ciencia es un sistema de conceptos sobre fenómenos objetivos o subjetivos que puede transformar la realidad en beneficio de la sociedad. Esta última, a su vez, influye en la selección de los objetos de estudio de la ciencia y en la forma en que se utilizan los resultados de la investigación científica (Kredov, 2004). Esta relación recíproca ciencia-sociedad se manifiesta también en la investigación sobre enfermedades que aquejan a la humanidad. En el siglo XXI, época de la clonación, la robótica y la inteligencia artificial, existen enfermedades infecciosas aún no resueltas debido a dificultades de tipo científico y social. En el contexto social encontramos a las enfermedades desatendidas, un conjunto de enfermedades infecciosas que durante muchos años han recibido atención insuficiente y son postergadas en las prioridades de salud pública debido a que atacan a poblaciones que carecen de influencia política, paralizadas por la pobreza y la inequidad (OPS, 2019).

Estas poblaciones se encuentran mayoritariamente en países subdesarrollados y en zonas de conflicto por lo que además persiste una inexistencia de estadísticas fiables provocando que millones de personas no tengan acceso al tratamiento y atención de salud a pesar de que a menudo tienen un costo extremadamente bajo. Es de suma importancia no dejar de lado la difusión en el conocimiento de estas enfermedades debido a que representan un grave problema de salud pública en países en vías de desarrollo, como nuestro país.

Las enfermedades desatendidas figuran entre las principales causas de discapacidad y son la causa principal de muerte prevenible y prematura en América Latina (Legetic et al., 2017). Hay que recordar que la pobreza va más allá de la falta de ingresos y recursos para garantizar medios de vida sostenibles. La pobreza es un problema de derechos humanos. Entre las distintas manifestaciones de la pobreza figuran el hambre, la malnutrición, la falta de vivienda digna y el acceso limitado a otros servicios básicos como la educación o la salud.

Actualmente, en el mundo más de 780 millones de personas, equivalentes 11% de la población mundial, viven por debajo del umbral de pobreza internacional (1,90 dólares al día) (ONU, 2020). En México, esta situación se ve reflejada, ya que dos de cada cinco

personas son pobres, y dos más son vulnerables de serlo por sus bajos ingresos (Figura 1). De acuerdo con la última medición de pobreza en el país, elaborada por el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL, 2020), en México se registraron 52.4 millones de personas pobres y 9.3 millones de ellas vivían en pobreza extrema.

La salud es considerada como uno de los principales determinantes de la pobreza y de su persistencia en el tiempo: las trampas de pobreza. Las trampas de pobreza asociadas a deficiencias en la salud ocurren porque los niños que presentan malnutrición son más susceptibles de padecer enfermedades y tienen un desarrollo cognitivo menor, lo cual resulta en menor productividad e ingreso cuando son adultos, dando origen al mismo ciclo en la generación siguiente (Lusting, 2007). Las poblaciones sometidas a un nivel socioeconómico bajo viven en condiciones que determinan la aparición de enfermedades, proporcionando a diversos microorganismos el ambiente ideal para su propagación, ocasionando altas tasas de incidencia de las enfermedades desatendidas. Estas afectan a más de 1000 millones de personas, es decir, a casi 1/6 de la población mundial principalmente en Asia, África y América del Sur, no es una sorpresa que más del 83% de las personas en situación de pobreza se concentren en estas regiones.

Las enfermedades desatendidas comprenden un grupo de entidades nosológicas que incluyen la rabia, dengue, ceguera por tracoma, úlcera de Buruli, treponematosi, lepra, enfermedad de



Figura 1. Medición de pobreza 2008–2018 y clasificación por poblaciones en México. (Modificado de CONEVAL, 2020)



Figura 2. Población en riesgo, número de casos, muertes por año y países afectados debido a enfermedades infecciosas desatendidas, en las Américas, 2014. (OPS, 2014)

Chagas, leishmaniasis, cisticercosis, filariasis, y las helmintiasis, entre otras (Figura 2).

La enfermedad de Chagas es una enfermedad parasitaria sistémica causada por el protozoo *Trypanosoma cruzi* que infecta alrededor de 10 millones de personas al año (OPS, 2019). Se transmite a los seres humanos y otros mamíferos por insectos chupadores de sangre, conocidos popularmente como chinche besucona. La infección casi siempre pasa desapercibida, ya que la mayoría de sus síntomas clínicos iniciales son inespecíficos, sin embargo, a largo plazo puede ocasionar insuficiencia cardíaca y trastornos digestivos graves (Pérez-Molina y Molina, 2018). Se asocia a múltiples factores sociales-ambientales entre los que destacan: residir en áreas de pobreza con inestabilidad social y económica, y muchas veces con altas tasas de migración, así como pertenecer a grupos vinculados con el trabajo agrícola. En los países endémicos de América Latina, se han dedicado esfuerzos para el control del insecto vector. Estos programas han sido exitosos en muchas zonas. Sin embargo, menos del 1% de los afectados recibe atención sanitaria y tratamiento para esta enfermedad (OPS, 2019).

El dengue es otro importante problema de salud pública en el mundo pues constituye una amenaza para un tercio de la población mundial. Se trata de un padecimiento viral, sistémico, agudo, transmitido por el mosquito *Aedes aegypti*. Tiene inicio súbito y se caracteriza por tener diferentes consecuencias físicas, dependiendo de la

forma en que se manifieste (OPS, 1994). Entre los factores económico-sociales asociados al dengue están el crecimiento poblacional sin precedente, la urbanización no controlada ni planificada, el aumento de la pobreza y el inadecuado ordenamiento ambiental (Gubler, 2002). Más del 80% de la población en América Latina vive en zonas urbanas; sin embargo, en la mayoría de las ciudades la urbanización se caracteriza por asentamientos informales sin las adecuadas condiciones de saneamiento. Existe una acumulación de utensilios a la intemperie, como llantas, latas y otros en los que se acumula el agua, que favorecen la proliferación del vector del dengue. Casi todas, son áreas densamente pobladas, y la pobreza es el factor predominante (Pérez, 2006).

Dentro del grupo de enfermedades desatendidas causadas por bacterias tenemos a una enfermedad milenaria: la lepra. Es causada por *Mycobacterium leprae* que afecta principalmente la piel, las mucosas del sistema respiratorio y sistema nervioso periférico de manera crónica (Blanco-Córdova y Claxton Louit, 2018). Las personas con esta enfermedad han sido objeto de estigma y rechazo desde hace cientos de años, y aunque actualmente puede creerse que ha sido eliminada, aún es un problema de salud pública en países como Brasil e India. En el año 2000, se registró un descenso aparente de la prevalencia global de la lepra, pero a pesar de los esfuerzos realizados se continuaron registrando casos nuevos en México y otros países (OPS, 2019). No es una coincidencia que se trate de países con una gran desigualdad económica. La pobreza es un

factor central para que se presente, ya que la hacinación propicia su desarrollo.

En el caso de las helmintiasis los agentes causales son los nematodos *Ascaris lumbricoides*, *Trichuris trichiura* y las uncinarias (*Necator americanus* y *Ancylostoma duodenale*). La infección se produce por la ingestión de huevos infectantes procedentes de tierra contaminada con heces humanas, de productos agrícolas crudos contaminados con tierra que contenga huevos infectantes o por la penetración de larvas desde el suelo a través de la piel (uncinarias) (OPS, 2019). En nuestro país las helmintiasis representan la enfermedad desatendida con mayor número de casos por año con 116,328 casos hasta la semana epidemiológica número 47 (SINAVE, 2020). Estas infecciones tienen frecuentemente repercusión sobre el crecimiento y desarrollo físico y cognitivo en los niños lo que desemboca en una disminución de su capacidad de generar ingresos adecuados que conlleva a su vez un deterioro en su nivel de vida perpetuando así el círculo pobreza-enfermedad. Aunque las helmintiasis son de tratamiento relativamente sencillo mediante la administración masiva de medicamentos antihelmínticos no han podido ser erradicadas debido a la población a la que afectan.

Todo esto nos deja ver que estas enfermedades representan una carga enorme en las comunidades y las personas en situación de pobreza, no solo debido al dolor y el sufrimiento que causan, sino también porque generan una pérdida de ingresos y el estigma crónico y la discriminación asociados con sus

Genética, epigenética o crianza: Determinantes de la personalidad en la vida adulta

POR M.C.S ANDREA MONSERRAT NEGRETE PAZ

Desde hace siglos profesionales de la filosofía y la psicología han debatido sobre la importancia de la naturaleza y la crianza en el desarrollo de las infancias. Actualmente, se sabe que el ADN o material genético de cada ser humano puede ser modificado mediante alteraciones epigenéticas ocasionadas por el entorno durante la niñez, conduciendo a la aparición de rasgos de la personalidad en la vida adulta; entonces, qué influye más en lo que somos, ¿la genética, la epigenética o la crianza?

Del ADN al gen y al genoma

El gen es la unidad básica de la herencia; los genes se transmiten de los padres a los hijos y contienen la información necesaria para precisar sus rasgos físicos y biológicos. Los genes están dispuestos uno tras otro en estructuras llamadas cromosomas (National Genome Research Institute, 2022) que forman el genoma de un organismo. Un cromosoma contiene una única molécula larga de ADN o ácido desoxirribonucleico; esta es considerada la molécula de la vida, ya que posee toda la información necesaria para el desarrollo de un organismo con rasgos y características definidas, incluyendo los de la personalidad.

Pero ¿qué es un rasgo de la personalidad? Los doctores en psicología Brent Roberts y Daniel Mroczec los definen como

"los patrones de pensamientos, sentimientos y comportamientos que distinguen a los individuos entre sí" (2008). Los rasgos de personalidad son complejos y la evidencia sugiere que nuestros rasgos están formados por factores hereditarios y ambientales, siendo uno de ellos la crianza; estos dos factores interactúan en una amplia variedad de formas para modelar nuestras personalidades individuales en la vida adulta.

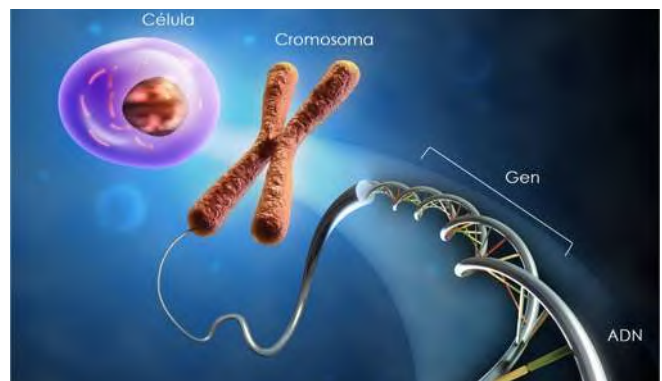
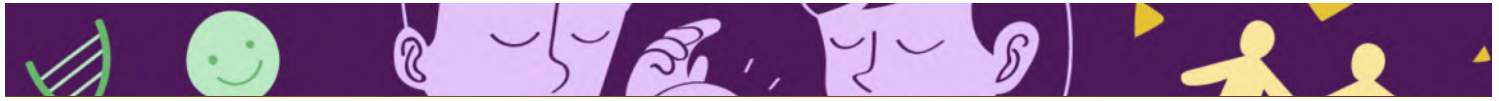


Imagen 1. El ADN forma los genes
<https://bit.ly/3J6QWEQ>



¿Somos el producto del entorno en el que nos desarrollamos o de la información contenida en nuestros genes?

Si bien siempre ha habido una aceptación generalizada de que los genes determinan nuestra fisiología para bien y para mal, ha habido una controversia mucho mayor en torno al tema de nuestra psicología: nuestro comportamiento y rasgos de personalidad.

Una cosa es afirmar que los genes determinan en gran medida qué tan altos y altas somos, qué tan rápido corremos y qué tanta predisposición tenemos, por ejemplo, a distintos tipos de cáncer.

En cambio, otra es argumentar que los genes también determinan en gran medida cuán inteligentes, simpáticos o antisociales somos; preferimos pensar en tales rasgos como construcciones sociales provocadas por los entornos familiares y sociales en los que nacemos y nos criamos, principalmente porque sabemos que las infancias aprenden de lo que ven a su alrededor. Sin embargo, debemos tener en cuenta que los genes definen nuestro potencial, pero es primordialmente el entorno lo que determina quiénes seremos, como consecuencia de modificaciones denominadas epigenéticas.

Aunque la secuencia de ADN es básicamente inalterable, una amplia variedad de factores del entorno puede conferir diferencias epigenéticas. La epigenética define cuáles genes en el genoma de una persona pueden activarse o desactivarse mediante determinadas "marcas", las que pueden ser diversas moléculas o grupos funcionales, como grupos metilo en las metilaciones (CH₃-) o grupos acetilo (CH₃CO-)

en las acetilaciones, estos grupos se añaden o remueven de algún gen y ocasionan modificaciones de la cromatina.

La cromatina es una forma empaquetada del ADN; para conocer cómo se forma hay que entender que en el núcleo de nuestras células hay una cantidad enorme de ADN -si lo viéramos estirado serían cerca de 2 metros-, éste tiene que compactarse para caber en el núcleo celular y esto se logra gracias a proteínas llamadas histonas, ya que el ADN se "superenrolla" sobre estas, empaquetándose en el genoma.

Las modificaciones epigenéticas sobre la cromatina pueden ser inserciones de grupos funcionales que van a cambiar la expresión del gen aumentando la síntesis de la proteína para la que codifican, es decir, van a activar al gen; o también pueden ser deleciones de grupos funcionales que van a disminuir la síntesis de la proteína, es decir, lo van a desactivar. De manera general, estas marcas hacen que determinados genes sean más o menos accesibles a las enzimas que los convierten en proteínas.

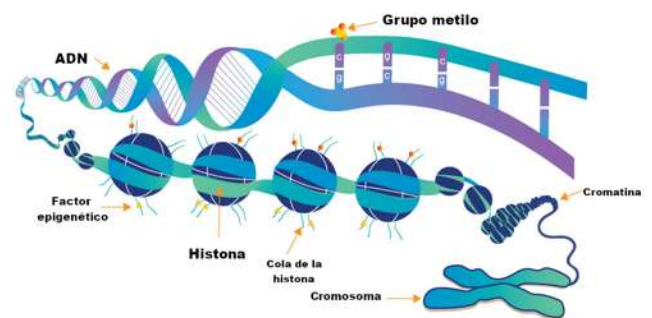
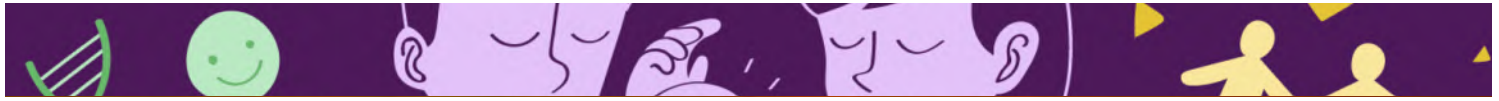


Imagen 2. Compactación del ADN.
<https://bit.ly/3zsSeqz>



Una investigación de décadas

Investigadores de distintos países han pasado décadas estudiando a la familia, los gemelos, los niños adoptados y las familias de acogida para comprender mejor qué parte de la personalidad es genética y qué parte es ambiental. Estos estudios se llevan a cabo en gemelos idénticos o gemelos homocigotos en distintas condiciones de crianza, debido a que se trata de individuos que son resultado de la fertilización de un solo óvulo, quiere decir que comparten todos sus genes.


¿Cómo pueden los genes idénticos producir personalidades tan diferentes? Un estudio en Minnesota sobre gemelos criados por separado, analizó a 350 pares de gemelos entre 1979 y 1999. Los participantes incluyeron gemelos idénticos y mellizos (gemelos no idénticos, dicigotos, que se producen por la fertilización de dos óvulos en el mismo embarazo y comparten la mitad de sus genes, como cualquier otro tipo de hermanos) que fueron criados juntos o separados. Los resultados revelaron que las personalidades de los gemelos idénticos eran similares si se criaron en el mismo hogar o separados, lo que sugiere que algunos aspectos de la personalidad están influenciados por la genética (Ridley y Pierpoint, 2003).

Se cree que muchos genes contribuyen al desarrollo de cada rasgo de la personalidad y que cada gen afecta a una gran cantidad de rasgos. En un trabajo realizado con individuos finlandeses se llevó a cabo una investigación que utiliza técnicas de inteligencia artificial para hallar agrupaciones de genes que interactúan entre sí y con el entorno, intentando analizar la herencia fisiológica de la personalidad. Estos genes identificados modulan procesos moleculares específicos en el cerebro para el establecimiento de objetivos intencionales, la autorreflexión, la empatía, el aprendizaje y la memoria. Los resultados muestran que el peso de la herencia en los principales rasgos de nuestra personalidad oscila entre el 30% y el 60% (Zwir et al., 2020)

Esto no significa que el entorno de la crianza no desempeñe un papel determinante en la formación de la personalidad. En un estudio realizado por Kaminsky y colaboradores (2008) se utilizó un análisis con micromatrices para la identificación de marcas epigenéticas en gemelos idénticos con personalidades diferentes en la vida adulta. Las micromatrices para el análisis de marcas epigenéticas (metilaciones) son técnicas moleculares ampliamente utilizadas para determinar el estado de metilación de múltiples genes de forma simultánea (Sánchez-Alvarez & Acevedo-Toro, 2015). Los resultados mostraron que, si bien los gemelos homocigotos son genéticamente idénticos a nivel de secuencia de ADN, una divergencia de los perfiles de metilación del



Imagen 3. Gemelos.
<https://bit.ly/3PtI0wH>



ADN producida durante el desarrollo, con el tiempo podría conducir a fenotipos diferentes, es decir, a la aparición de rasgos de la personalidad distintos.

Con base en los estudios anteriormente mencionados, se ha determinado que los patrones epigenéticos sirven como un puente importante entre la crianza y el desarrollo de la personalidad, debido a que también pueden verse afectados por el entorno del niño y cambiar a lo largo de la vida por una experiencia en etapa temprana, exposición ambiental o estado nutricional. Las firmas epigenéticas influenciadas por el entorno pueden determinar nuestra apariencia, comportamiento, respuesta al estrés, susceptibilidad a enfermedades e incluso longevidad, por lo que, en conjunto con nuestros genes, contribuyen al desarrollo de la personalidad en la vida adulta.

Referencias

- Kaminsky, Z., Petronis, A., Wang, S. C., Levine, B., Ghaffar, O., Floden, D., & Feinstein, A. (2008). Epigenetics of personality traits: an illustrative study of identical twins discordant for risk-taking behavior. *Twin Research and Human Genetics*, 11(1), 1-11.
- National Human Research Genome Institute. (2022). *Gen*. <https://www.genome.gov/es/genetics-glossary/Gen>
- Ridley, M., & Pierpoint, G. (2003). *Nature via nurture: Genes, experience, and what makes us human* (Vol. 19). New York: HarperCollins.
- Roberts, B. W., & Mroczek, D. (2008). Personality trait change in adulthood. *Current directions in psychological science*, 17(1), 31-35.
- Sánchez Álvarez, J. P., & Acevedo Toro, P. A. (2015). Principales herramientas epigenéticas para el diagnóstico y seguimiento de neoplasias hematológicas.
- Zwir, I., Mishra, P., Del-Val, C., Gu, C. C., de Erausquin, G. A., Lehtimäki, T., & Cloninger, C. R. (2020). Uncovering the complex genetics of human personality: response from authors on the PGMRA Model. *Molecular psychiatry*, 25(10), 2210-2213.



ANEXOS

Anexo 1. Cepas obtenidas de la base de datos SRA del NCBI

No.	Cepa	Sitio de infección	Fenotipo	SRA	País
1	DCTB0011	EPTB	EPTB	SRR998860	Rusia
2	Index 78-34	Sistema nervioso central	EPTB	SRR6480455	Indonesia
3	ETB0119	Huesos y articulaciones	EPTB	SRR6257091	Rusia
4	48-10477	Sistema nervioso central	EPTB	SRR5709859	Tailandia
5	Index 4-42	Sistema nervioso central	EPTB	SRR6480537	Indonesia
6	Index 28-66	Sistema nervioso central	EPTB	SRR6480585	Indonesia
7	Index 60-88	Sistema nervioso central	EPTB	SRR6480481	Indonesia
8	Index 66-88	Sistema nervioso central	EPTB	SRR6480501	Indonesia
9	TB0002	Huesos y articulaciones	EPTB	SRR6256980	Rusia
10	TB0010	Huesos y articulaciones	EPTB	SRR6256978	Rusia
11	L-92954	Sistema nervioso central	EPTB	SRR5709831	Tailandia
12	Index 4-66	Sistema nervioso central	EPTB	SRR6480445	Indonesia
13	DS-21277	Sistema nervioso central	EPTB	SRR5709931	Tailandia
14	TB0068	Huesos y articulaciones	EPTB	SRR6257006	Rusia
15	CSF-3367	Sistema nervioso central	EPTB	SRR5709804	Tailandia
16	Index 60-19	Sistema nervioso central	EPTB	SRR6480609	Indonesia
17	4-12339	Sistema nervioso central	EPTB	SRR5709998	Tailandia
18	TB0059	Huesos y articulaciones	EPTB	SRR6256986	Rusia
19	43-12700	Sistema nervioso central	EPTB	SRR5710006	Tailandia
20	Index 4-88	Sistema nervioso central	EPTB	SRR6480528	Indonesia

21	42-19187	Sistema nervioso central	EPTB	SRR5709980	Tailandia
22	DCTB0012	EPTB	EPTB	SRR998861	Rusia
23	Index 45-88	Sistema nervioso central	EPTB	SRR6480525	Indonesia
24	Index 18-66	Sistema nervioso central	EPTB	SRR6480582	Indonesia
25	TB0079	Huesos y articulaciones	EPTB	SRR6257013	Rusia
26	Mtb-manila-23	Sistema genitourinario	EPTB	SRR1510058	Canadá
27	CSF3943	Sistema nervioso central	EPTB	SRR5709999	Tailandia
28	CSF-4438	Sistema nervioso central	EPTB	SRR5710029	Tailandia
29	CSF-2894	Sistema nervioso central	EPTB	SRR5709977	Tailandia
30	46-18357	Sistema nervioso central	EPTB	SRR5709997	Tailandia
31	46-19260	Sistema nervioso central	EPTB	SRR5709971	Tailandia
32	43-19186	Sistema nervioso central	EPTB	SRR5710001	Tailandia
33	47-17522	Sistema nervioso central	EPTB	SRR5709742	Tailandia
34	Index 55-19	Sistema nervioso central	EPTB	SRR6480620	Indonesia
35	DCTB0004	EPTB	EPTB	SRR998140	Rusia
36	TB0036	Huesos y articulaciones	EPTB	SRR6257086	Rusia
37	CSF-4129	Sistema nervioso central	EPTB	SRR5709834	Tailandia
38	Index 30-88	Sistema nervioso central	EPTB	SRR6480394	Indonesia
39	Index 73-19	Sistema nervioso central	EPTB	SRR6480381	Indonesia
40	Index 12-88	Sistema nervioso central	EPTB	SRR6480407	Indonesia
41	DCTB0006	EPTB	EPTB	SRR998856	Rusia
42	Index 45-42	Sistema nervioso central	EPTB	SRR6480500	Indonesia
43	Index 21-42	Sistema nervioso central	EPTB	SRR6480416	Indonesia

44	Index 40-34	Sistema nervioso central	EPTB	SRR6480472	Tailandia
45	DKC2-0499	Ganglios linfáticos	EPTB	ERR2229374	Dinamarca
46	CSF-4374	Sistema nervioso central	EPTB	SRR5709861	Tailandia
47	46-10698	Sistema nervioso central	EPTB	SRR5709739	Tailandia
48	CSF4676	Sistema nervioso central	EPTB	SRR5709975	Tailandia
49	CSF-3518	Sistema nervioso central	EPTB	SRR5709788	Tailandia
50	Index 24-66	Sistema nervioso central	EPTB	SRR6480584	Indonesia
51	TB0092	Huesos y articulaciones	EPTB	SRR6257024	Rusia
52	Index 30-19	Sistema nervioso central	EPTB	SRR6480624	Indonesia
53	Index 21-34	Sistema nervioso central	EPTB	SRR6480435	Indonesia
54	Index 6-101	Sistema nervioso central	EPTB	SRR6480557	Indonesia
55	TB0039	Huesos y articulaciones	EPTB	SRR6257085	Rusia
56	PGI_IJOB_EPTB5	Ganglios linfáticos	EPTB	SRR5125077	India
57	CSF-3382	Sistema nervioso central	EPTB	SRR5709783	Tailandia
58	Index 73-88	Sistema nervioso central	EPTB	SRR6480477	Indonesia
59	43-08441	Sistema nervioso central	EPTB	SRR5709985	Tailandia
60	Index 6-19	Sistema nervioso central	EPTB	SRR6480373	Indonesia
61	Index 34-88	Sistema nervioso central	EPTB	SRR6480521	Indonesia
62	Index 73-34	Sistema nervioso central	EPTB	SRR6480457	Indonesia
63	CSF-4049	Sistema nervioso central	EPTB	SRR5709857	Tailandia
64	43-12683	Sistema nervioso central	EPTB	SRR5709973	Tailandia
65	CSF-4229	Sistema nervioso central	EPTB	SRR5709836	Tailandia

66	DKC2-0T97	Sistema genitourinario	EPTB	ERR1950087	Dinamarca
67	CSF-3541	Sistema nervioso central	EPTB	SRR5709779	Tailandia
68	Index 61-19	Sistema nervioso central	EPTB	SRR6480626	Indonesia
69	L-78532	Sistema nervioso central	EPTB	SRR5709800	Tailandia
70	CSF-3538	Sistema nervioso central	EPTB	SRR5709789	Tailandia
71	Index 8-34	Sistema nervioso central	EPTB	SRR6480454	Indonesia
72	Index 28-88	Sistema nervioso central	EPTB	SRR6480398	Indonesia
73	DKC2-0843	Sistema genitourinario	EPTB	ERR2229774	Dinamarca
74	Index 37-19	Sistema nervioso central	EPTB	SRR6480631	Indonesia
75	Index 12-20	Sistema nervioso central	EPTB	SRR6480452	Indonesia
76	43-15038	Sistema nervioso central	EPTB	SRR5710003	Tailandia
77	CSF-3681	Sistema nervioso central	EPTB	SRR5709746	Tailandia
78	15-1149-0143	Sistema genitourinario	EPTB	SRR6367399	Australia
79	47-5556	Sistema nervioso central	EPTB	SRR5709784	Tailandia
80	TB0055	Huesos y articulaciones	EPTB	SRR6256990	Rusia
81	PGI_JOB_EPTB1	Sistema nervioso central	EPTB	SRR5125076	India
82	DKC2-0244	Sistema genitourinario	EPTB	ERR2229064	Dinamarca
83	Index 34-34	Sistema nervioso central	EPTB	SRR6480424	Indonesia
84	Index 45-19	Sistema nervioso central	EPTB	SRR6480606	Indonesia
85	43-13037	Sistema nervioso central	EPTB	SRR5710005	Tailandia
86	Index 12-66	Sistema nervioso central	EPTB	SRR6480376	Indonesia
87	Index 15-20	Sistema nervioso central	EPTB	SRR6480363	Indonesia

88	17-3391-0148	Sistema genitourinario	EPTB	SRR6339646	Australia
89	Index 78-25	Sistema nervioso central	EPTB	SRR6480612	Indonesia
90	Index 32-19	Sistema nervioso central	EPTB	SRR6480385	Indonesia
91	TB0006	Huesos y articulaciones	EPTB	SRR6256975	Rusia
92	PGI_IOB_EPTB2	Ganglios linfáticos	EPTB	SRR5125078	India
93	Index 34-20	Sistema nervioso central	EPTB	SRR6480370	Indonesia
94	DKC2-0246	Huesos y articulaciones	EPTB	ERR2229066	Dinamarca
95	CSF-4056	Sistema nervioso central	EPTB	SRR5709856	Tailandia
96	Index 42-34	Sistema nervioso central	EPTB	SRR6480471	Indonesia
97	43-11791	Sistema nervioso central	EPTB	SRR5709974	Tailandia
98	Index 8-101	Sistema nervioso central	EPTB	SRR6480555	Indonesia
99	Index 6-88	Sistema nervioso central	EPTB	SRR6480508	Indonesia
100	TB0008	Huesos y articulaciones	EPTB	SRR6256976	Rusia
101	TB0048	Huesos y articulaciones	EPTB	SRR6257109	Rusia
102	Index 12-19	Sistema nervioso central	EPTB	SRR6480339	Indonesia
103	Index 8-19	Sistema nervioso central	EPTB	SRR6480608	Indonesia
104	Index 55-34	Sistema nervioso central	EPTB	SRR6480469	Indonesia
105	48-10563	Sistema nervioso central	EPTB	SRR5709853	Tailandia
106	PGI_IOB_EPTB4	Ganglios linfáticos	EPTB	SRR5125075	India
107	Index 8-20	Sistema nervioso central	EPTB	SRR6480483	Indonesia
108	Index 61-88	Sistema nervioso central	EPTB	SRR6480482	Indonesia
109	43-01910	Sistema nervioso central	EPTB	SRR5709979	Tailandia

110	Mtb-manila-1	Ganglios linfáticos	EPTB	SRR1510036	Canadá
111	Index 66-20	Sistema nervioso central	EPTB	SRR6480487	Indonesia
112	TB0157	Huesos y articulaciones	EPTB	SRR6257056	Rusia
113	Index 21-66	Sistema nervioso central	EPTB	SRR6480583	Indonesia
114	Index 18-20	Sistema nervioso central	EPTB	SRR6480364	Indonesia
115	Index 30-66	Sistema nervioso central	EPTB	SRR6480397	Indonesia
116	TB0040	Huesos y articulaciones	EPTB	SRR6257082	Rusia
117	48-10231	Sistema nervioso central	EPTB	SRR5709852	Tailandia
118	CSF-4666	Sistema nervioso central	EPTB	SRR5709745	Tailandia
119	CSF-3419	Sistema nervioso central	EPTB	SRR5709995	Tailandia
120	Index 55-20	Sistema nervioso central	EPTB	SRR6480491	Indonesia
121	Index 45-20	Sistema nervioso central	EPTB	SRR6480492	Indonesia
122	CSF-4549	Sistema nervioso central	EPTB	SRR5709790	Tailandia
123	TB0080	Huesos y articulaciones	EPTB	SRR6257027	Rusia
124	CSF-4261	Sistema nervioso central	EPTB	SRR5709830	Tailandia
125	TB0085	Huesos y articulaciones	EPTB	SRR6257031	Rusia
126	Index 21-20	Sistema nervioso central	EPTB	SRR6480365	Indonesia
127	TB0053	Huesos y articulaciones	EPTB	SRR6257100	Rusia
128	Index 40-19	Sistema nervioso central	EPTB	SRR6480627	Indonesia
129	43-11467	Sistema nervioso central	EPTB	SRR5709983	Tailandia
130	TB0042	Huesos y articulaciones	EPTB	SRR6257080	Rusia
131	Index 15-66	Sistema nervioso central	EPTB	SRR6480378	Indonesia

132	TB0054	Huesos y articulaciones	EPTB	SRR6256989	Rusia
133	Index 24-19	Sistema nervioso central	EPTB	SRR6480623	Indonesia
134	Mtb40	Ganglios linfáticos	EPTB	SRR5817466	Australia
135	Index 12-34	Sistema nervioso central	EPTB	SRR6480432	Indonesia
136	CSF-3542	Sistema nervioso central	EPTB	SRR5709778	Tailandia
137	CSF-4620	Sistema nervioso central	EPTB	SRR5709832	Tailandia
138	Index 32-34	Sistema nervioso central	EPTB	SRR6480439	Indonesia
139	CSF-4396	Sistema nervioso central	EPTB	SRR5709828	Tailandia
140	TB0064	Huesos y articulaciones	EPTB	SRR6256993	Rusia
141	Index 4-34	Sistema nervioso central	EPTB	SRR6480473	Indonesia
142	Index 18-19	Sistema nervioso central	EPTB	SRR6480565	Indonesia
143	TB0088	Huesos y articulaciones	EPTB	SRR6257030	Rusia
144	Index 75-34	Sistema nervioso central	EPTB	SRR6480456	Indonesia
145	Index 6-20	Sistema nervioso central	EPTB	SRR6480490	Indonesia
146	Index 55-88	Sistema nervioso central	EPTB	SRR6480430	Indonesia
147	Index 8-88	Sistema nervioso central	EPTB	SRR6480480	Indonesia
148	TB0044	Huesos y articulaciones	EPTB	SRR6257105	Rusia
149	Index 78-88	Sistema nervioso central	EPTB	SRR6480479	Indonesia
150	Index 37-42	Sistema nervioso central	EPTB	SRR6480543	Indonesia
151	ETB0090	Huesos y articulaciones	EPTB	SRR6257034	Rusia
152	DCTB0009	EPTB	EPTB	SRR998858	Rusia
153	Index 69-88	Sistema nervioso central	EPTB	SRR6480605	Indonesia

154	Index 15-34	Sistema nervioso central	EPTB	SRR6480433	Indonesia
155	TB0113	Huesos y articulaciones	EPTB	SRR6257097	Rusia
156	42-20464	Sistema nervioso central	EPTB	SRR5709741	Tailandia
157	Index 78-19	Sistema nervioso central	EPTB	SRR6480566	Indonesia
158	Index 40-20	Sistema nervioso central	EPTB	SRR6480371	Indonesia
159	DCTB0008	EPTB	EPTB	SRR998857	Rusia
160	CSF-3922	Sistema nervioso central	EPTB	SRR5709854	Tailandia
161	CSF-4202	Sistema nervioso central	EPTB	SRR5709835	Tailandia
162	Index 4-101	Sistema nervioso central	EPTB	SRR6480562	Indonesia
163	Index 60-20	Sistema nervioso central	EPTB	SRR6480489	Indonesia
164	TB0012	Huesos y articulaciones	EPTB	SRR6256984	Rusia
165	CSF-4105	Sistema nervioso central	EPTB	SRR5709833	Tailandia
166	CSF-4069	Sistema nervioso central	EPTB	SRR5709851	Tailandia
167	Mtb-manila-27	Ganglios linfáticos	EPTB	SRR1510062	Canadá
168	Index 42-40	Sistema nervioso central	EPTB	SRR6480372	Indonesia
169	CSF-1449	Sistema nervioso central	EPTB	SRR5709972	Tailandia
170	Index 6-34	Sistema nervioso central	EPTB	SRR6480468	Indonesia
171	Index 4-19	Sistema nervioso central	EPTB	SRR6480632	Indonesia
172	CSF-4480	Sistema nervioso central	EPTB	SRR5709838	Tailandia
173	Mtb-33	Ganglios linfáticos	EPTB	SRR5817467	Australia
174	TB0058	Huesos y articulaciones	EPTB	SRR6256985	Rusia
175	TB0086	Huesos y articulaciones	EPTB	SRR6257032	Rusia

176	TB0057	Huesos y articulaciones	EPTB	SRR6256992	Rusia
177	CSF-3384	Sistema nervioso central	EPTB	SRR5709786	Tailandia
178	Index 21-88	Sistema nervioso central	EPTB	SRR6480395	Indonesia
179	TB0011	Huesos y articulaciones	EPTB	SRR6256983	Rusia
180	43-05286	Sistema nervioso central	EPTB	SRR5709986	Tailandia
181	47-10904	Sistema nervioso central	EPTB	SRR5709785	Tailandia
182	Index 75-88	Sistema nervioso central	EPTB	SRR6480478	Indonesia
183	TB0037	Huesos y articulaciones	EPTB	SRR6257087	Rusia
184	Index 32-88	Sistema nervioso central	EPTB	SRR6480522	Indonesia
185	TB0047	Huesos y articulaciones	EPTB	SRR6257102	Rusia
186	43-13718	Sistema nervioso central	EPTB	SRR5710004	Tailandia
187	Index 61-34	Sistema nervioso central	EPTB	SRR6480466	Indonesia
188	14-14745	Sistema nervioso central	EPTB	SRR5709740	Tailandia
189	Index 42-88	Sistema nervioso central	EPTB	SRR6480526	Indonesia
190	Index 30-42	Sistema nervioso central	EPTB	SRR6480414	Indonesia
191	Index 28-20	Sistema nervioso central	EPTB	SRR6480367	Indonesia
192	Index 66-34	Sistema nervioso central	EPTB	SRR6480476	Indonesia
193	Index 61-20	Sistema nervioso central	EPTB	SRR6480488	Indonesia
194	48-4979	Sistema nervioso central	EPTB	SRR5709855	Tailandia
195	Mtb-manila-14	Sistema nervioso central	EPTB	SRR1510049	Canadá
196	Index 15-88	Sistema nervioso central	EPTB	SRR6480404	Indonesia
197	CSF-4115	Sistema nervioso central	EPTB	SRR5709850	Tailandia

198	Mtb-manila-22	Huesos y articulaciones	EPTB	SRR1510057	Canadá
199	TB0038	Huesos y articulaciones	EPTB	SRR6257084	Rusia
200	CSF-4014	Sistema nervioso central	EPTB	SRR5709858	Tailandia
201	TB0081	Huesos y articulaciones	EPTB	SRR6257028	Rusia
202	Index 66-19	Sistema nervioso central	EPTB	SRR6480331	Indonesia
203	CSF-3711	Sistema nervioso central	EPTB	SRR5709747	Tailandia
204	DCTB0005	EPTB	EPTB	SRR998855	Rusia
205	Index 28-34	Sistema nervioso central	EPTB	SRR6480437	Indonesia
206	Index 24-20	Sistema nervioso central	EPTB	SRR6480366	Indonesia
207	CSF2498	Sistema nervioso central	EPTB	SRR5709976	Tailandia
208	CSF-3346	Sistema nervioso central	EPTB	SRR5709791	Tailandia
209	CSF-3718	Sistema nervioso central	EPTB	SRR5709744	Tailandia
210	DKC2-0T40	Ganglios linfáticos	EPTB	ERR2229808	Dinamarca
211	72-125842	Sistema nervioso central	EPTB	SRR5710000	Tailandia
212	Index 24-88	Sistema nervioso central	EPTB	SRR6480396	Indonesia
213	Index 34-19	Sistema nervioso central	EPTB	SRR6480619	Indonesia
214	Index 69-19	Sistema nervioso central	EPTB	SRR6480333	Indonesia
215	TB0051	Huesos y articulaciones	EPTB	SRR6257106	Rusia
216	Index 12-42	Sistema nervioso central	EPTB	SRR6480460	Indonesia
217	Index 78-20	Sistema nervioso central	EPTB	SRR6480484	Indonesia
218	PGI_IOB_EPTB3	Ganglios linfáticos	EPTB	SRR5125074	India
219	Index 40-88	Sistema nervioso central	EPTB	SRR6480527	Indonesia

220	TB0083	Huesos y articulaciones	EPTB	SRR6257025	Rusia
221	CSF-4248	Sistema nervioso central	EPTB	SRR5709829	Tailandia
222	Index 32-20	Sistema nervioso central	EPTB	SRR6480369	Indonesia
223	Index 75-20	Sistema nervioso central	EPTB	SRR6480485	Indonesia
224	Index 18-34	Sistema nervioso central	EPTB	SRR6480434	Indonesia
225	TB0043	Huesos y articulaciones	EPTB	SRR6257081	Rusia
226	4317298	Sistema nervioso central	EPTB	SRR5710002	Tailandia
227	Index 6-66	Sistema nervioso central	EPTB	SRR6480448	Indonesia
228	CSF-3053	Sistema nervioso central	EPTB	SRR5709978	Tailandia
229	Index 69-34	Sistema nervioso central	EPTB	SRR6480475	Indonesia
230	43-11359	Sistema nervioso central	EPTB	SRR5709984	Tailandia
231	Index 73-20	Sistema nervioso central	EPTB	SRR6480486	Indonesia
232	DCTB0002	EPTB	EPTB	SRR993140	Rusia
233	Mtb-manila-25	Sistema genitourinario	EPTB	SRR1510060	Canadá
234	CSF-3520	Sistema nervioso central	EPTB	SRR5709787	Tailandia
235	Index 8-66	Sistema nervioso central	EPTB	SRR6480406	Indonesia
236	Index 42-19	Sistema nervioso central	EPTB	SRR6480629	Indonesia
237	CSF-4409	Sistema nervioso central	EPTB	SRR5709939	Tailandia
238	TB0050	Huesos y articulaciones	EPTB	SRR6257107	Rusia
239	Index 18-88	Sistema nervioso central	EPTB	SRR6480405	Indonesia
240	TB0072	Huesos y articulaciones	EPTB	SRR6257007	Rusia
241	TB0098	Huesos y articulaciones	EPTB	SRR6257018	Rusia

242	46-5069	Sistema nervioso central	EPTB	SRR5709738	Tailandia
243	Index 12-101	Sistema nervioso central	EPTB	SRR6480595	Indonesia
244	Index 24-34	Sistema nervioso central	EPTB	SRR6480436	Indonesia
245	Mtb-manila-36	Ganglios linfáticos	EPTB	SRR1510071	Canadá
246	Index 12-109	Pulmones	Pulmonar	ERR2229367	Indonesia
247	Index 32-66	Pulmones	Pulmonar	SRR6480408	Indonesia
248	Index 6-109	Pulmones	Pulmonar	SRR6480382	Indonesia
249	Index 60-42	Pulmones	Pulmonar	SRR6480494	Indonesia
250	Index 69-25	Pulmones	Pulmonar	SRR6480617	Indonesia
251	DS-25878	Pulmones	Pulmonar	SRR5709875	Tailandia
252	c37_2008	Pulmones	Pulmonar	ERR1144995	Australia
253	Index 37-133	Pulmones	Pulmonar	SRR6480326	Indonesia
254	DS-27524	Pulmones	Pulmonar	SRR5709753	Tailandia
255	Index 73-100	Pulmones	Pulmonar	SRR6480601	Indonesia
256	Mtb_ON-A_strain54	Pulmones	Pulmonar	SRR1573733	Canadá
257	DS-10110	Pulmones	Pulmonar	SRR5709802	Tailandia
258	DS-14191	Pulmones	Pulmonar	SRR5709823	Tailandia
259	DS-20257	Pulmones	Pulmonar	SRR5709943	Tailandia
260	DS-13949	Pulmones	Pulmonar	SRR5709820	Tailandia
261	DS-20915	Pulmones	Pulmonar	SRR5709770	Tailandia
262	c32_2012	Pulmones	Pulmonar	ERR1144994	Australia
263	DS-11745	Pulmones	Pulmonar	SRR5709801	Tailandia
264	Index 42-21	Pulmones	Pulmonar	SRR6480573	Indonesia
265	Index 42-100	Pulmones	Pulmonar	SRR6480343	Indonesia
266	DS-19861	Pulmones	Pulmonar	SRR5709940	Tailandia
267	DS-13671	Pulmones	Pulmonar	SRR5709799	Tailandia
268	Index 15-25	Pulmones	Pulmonar	SRR6480545	Indonesia
269	DS-18501	Pulmones	Pulmonar	SRR5709953	Tailandia
270	DS-16630	Pulmones	Pulmonar	SRR5709798	Tailandia
271	Index 28-142	Pulmones	Pulmonar	SRR6480411	Indonesia
272	Index 21-106	Pulmones	Pulmonar	SRR6480509	Indonesia
273	Index 66-109	Pulmones	Pulmonar	SRR6480345	Indonesia
274	DCTB021	Pulmones	Pulmonar	SRR1002689	Rusia
275	Index 60-21	Pulmones	Pulmonar	SRR6480574	Indonesia
276	DS-17016	Pulmones	Pulmonar	SRR5709892	Tailandia
277	RTS-3	Pulmones	Pulmonar	SRR6914120	Rusia
278	Index 15-100	Pulmones	Pulmonar	SRR6480401	Indonesia

279	Index 32-90	Pulmones	Pulmonar	SRR6480318	Indonesia
280	Index 55-109	Pulmones	Pulmonar	SRR6480628	Indonesia
281	Index 12-100	Pulmones	Pulmonar	SRR6480402	Indonesia
282	DS-14490	Pulmones	Pulmonar	SRR5709864	Tailandia
283	EKB43	Pulmones	Pulmonar	SRR8357275	Rusia
284	DS-20512	Pulmones	Pulmonar	SRR5710016	Tailandia
285	EKB25	Pulmones	Pulmonar	SRR8348943	Rusia
286	DS-33051	Pulmones	Pulmonar	SRR5709862	Tailandia
287	Index 40-133	Pulmones	Pulmonar	SRR6480328	Indonesia
288	Index 12-106	Pulmones	Pulmonar	SRR6480512	Indonesia
289	c30_2012	Pulmones	Pulmonar	ERR1144992	Australia
290	EKB9	Pulmones	Pulmonar	SRR8335024	Rusia
291	DS-32353	Pulmones	Pulmonar	SRR5709911	Tailandia
292	Index 75-109	Pulmones	Pulmonar	SRR6480342	Indonesia
293	Index 32-21	Pulmones	Pulmonar	SRR6480534	Indonesia
294	Index 37-109	Pulmones	Pulmonar	SRR6480389	Indonesia
295	EKB33	Pulmones	Pulmonar	SRR8439315	Rusia
296	RTS-6	Pulmones	Pulmonar	SRR6914123	Rusia
297	Index 24-90	Pulmones	Pulmonar	SRR6480634	Indonesia
298	DS-16496	Pulmones	Pulmonar	SRR5709912	Tailandia
299	EKB83	Pulmones	Pulmonar	SRR8375802	Rusia
300	Index 6-21	Pulmones	Pulmonar	SRR6480576	Indonesia
301	EKB24	Pulmones	Pulmonar	SRR8348431	Rusia
302	Index32-106	Pulmones	Pulmonar	SRR6480513	Indonesia
303	Index 55-42	Pulmones	Pulmonar	SRR6480498	Indonesia
304	EKB1	Pulmones	Pulmonar	SRR8327216	Rusia
305	Index 18-42	Pulmones	Pulmonar	SRR6480450	Indonesia
306	DS-15504	Pulmones	Pulmonar	SRR5710009	Tailandia
307	DS-20092	Pulmones	Pulmonar	SRR5709961	Tailandia
308	JAL_IOB_199	Pulmones	Pulmonar	SRR5341273	India
309	Index 45-142	Pulmones	Pulmonar	SRR6480499	Indonesia
310	Index 6-106	Pulmones	Pulmonar	SRR6480421	Indonesia
311	Index 45-66	Pulmones	Pulmonar	SRR6480446	Indonesia
312	Index 18-25	Pulmones	Pulmonar	SRR6480546	Indonesia
313	DS-30312	Pulmones	Pulmonar	SRR5710011	Tailandia
314	EKB7	Pulmones	Pulmonar	SRR8330159	Rusia
315	RTS-5	Pulmones	Pulmonar	SRR6914122	Rusia
316	DS-17092	Pulmones	Pulmonar	SRR5709889	Tailandia
317	DS-10471	Pulmones	Pulmonar	SRR5709948	Tailandia
318	c31_2012	Pulmones	Pulmonar	ERR1144993	Australia
319	DS-29147	Pulmones	Pulmonar	SRR5709763	Tailandia

320	DS-17290	Pulmones	Pulmonar	SRR5709794	Tailandia
321	Index 73-142	Pulmones	Pulmonar	SRR6480358	Indonesia
322	JAL_IOB_198	Pulmones	Pulmonar	SRR5341274	India
323	Index 30-21	Pulmones	Pulmonar	SRR6480533	Indonesia
324	Index 45-100	Pulmones	Pulmonar	SRR6480462	Indonesia
325	Index 28-25	Pulmones	Pulmonar	SRR6480548	Indonesia
326	DS-17984	Pulmones	Pulmonar	SRR5709888	Tailandia
327	DS-6882	Pulmones	Pulmonar	SRR5709933	Tailandia
328	Index 12-25	Pulmones	Pulmonar	SRR6480552	Indonesia
329	EKB20	Pulmones	Pulmonar	SRR8346382	Rusia
330	DS-14607	Pulmones	Pulmonar	SRR5709954	Tailandia
331	DS-21363	Pulmones	Pulmonar	SRR5709991	Tailandia
332	Index 6-133	Pulmones	Pulmonar	SRR6480321	Indonesia
333	Index 13-133	Pulmones	Pulmonar	SRR6480610	Indonesia
334	EKB27	Pulmones	Pulmonar	SRR8352208	Rusia
335	RTS-2	Pulmones	Pulmonar	SRR6914119	Rusia
336	Index 42-66	Pulmones	Pulmonar	SRR6480447	Indonesia
337	Index 40-42	Pulmones	Pulmonar	SRR6480539	Indonesia
338	DS-19190	Pulmones	Pulmonar	SRR5709921	Tailandia
339	Index 73-21	Pulmones	Pulmonar	SRR6480567	Indonesia
340	DS-15177	Pulmones	Pulmonar	SRR5709956	Tailandia
341	Index 78-106	Pulmones	Pulmonar	SRR6480431	Indonesia
342	JAL_IOB_197	Pulmones	Pulmonar	SRR5341275	India
343	Index 66-21	Pulmones	Pulmonar	SRR6480575	Indonesia
344	Index 40-106	Pulmones	Pulmonar	SRR6480463	Indonesia
345	DCTB018	Pulmones	Pulmonar	SRR1002684	Rusia
346	Index 30-34	Pulmones	Pulmonar	SRR6480438	Indonesia
347	Index 61-66	Pulmones	Pulmonar	SRR6480440	Indonesia
348	DS-19200	Pulmones	Pulmonar	SRR5709922	Tailandia
349	DS-12236	Pulmones	Pulmonar	SRR5709761	Tailandia
350	DKC2-0485	Pulmones	Pulmonar	ERR2229361	Dinamarca
351	EKB36	Pulmones	Pulmonar	SRR8354461	Rusia
352	Index 6-142	Pulmones	Pulmonar	SRR6480495	Indonesia
353	Index 40-109	Pulmones	Pulmonar	SRR6480390	Indonesia
354	DS-16282	Pulmones	Pulmonar	SRR5709837	Tailandia
355	Index 34-100	Pulmones	Pulmonar	SRR6480336	Indonesia
356	DS-23682	Pulmones	Pulmonar	SRR5709805	Tailandia
357	Index 55-142	Pulmones	Pulmonar	SRR6480497	Indonesia
358	DS-20503	Pulmones	Pulmonar	SRR5709777	Tailandia
359	DS-31799	Pulmones	Pulmonar	SRR5709910	Tailandia
360	DS-23522	Pulmones	Pulmonar	SRR5709806	Tailandia

361	EKB28	Pulmones	Pulmonar	SRR8351921	Rusia
362	EKB38	Pulmones	Pulmonar	SRR8354716	Rusia
363	DS-31224	Pulmones	Pulmonar	SRR5709907	Tailandia
364	DS-20316	Pulmones	Pulmonar	SRR5709936	Tailandia
365	Index 42-109	Pulmones	Pulmonar	SRR6480391	Indonesia
366	Index 69-42	Pulmones	Pulmonar	SRR6480359	Indonesia
367	DS-31213	Pulmones	Pulmonar	SRR5710026	Tailandia
368	Index 37-142	Pulmones	Pulmonar	SRR6480544	Indonesia
369	Index 42-142	Pulmones	Pulmonar	SRR6480507	Indonesia
370	DKC2-0491	Pulmones	Pulmonar	ERR2229366	Dinamarca
371	DS-29456	Pulmones	Pulmonar	SRR5709903	Tailandia
372	WT9	Pulmones	Pulmonar	SRR1573725	Canadá
373	Index 69-106	Pulmones	Pulmonar	SRR6480344	Indonesia
374	EKB100	Pulmones	Pulmonar	SRR8434639	Rusia
375	DS-16280	Pulmones	Pulmonar	SRR5709946	Tailandia
376	Index 78-133	Pulmones	Pulmonar	SRR6480329	Indonesia
377	WT17	Pulmones	Pulmonar	SRR1573730	Canadá
378	JAL_IOB_195	Pulmones	Pulmonar	SRR5341277	India
379	Index 4-106	Pulmones	Pulmonar	SRR6480607	Indonesia
380	Index 4_19	Pulmones	Pulmonar	SRR6480362	Indonesia
381	Index 69-21	Pulmones	Pulmonar	SRR6480568	Indonesia
382	DS-26102	Pulmones	Pulmonar	SRR5709749	Tailandia
383	WT16	Pulmones	Pulmonar	SRR1573729	Canadá
384	DS-17016	Pulmones	Pulmonar	SRR5709967	Tailandia
385	Index 45-133	Pulmones	Pulmonar	SRR6480327	Indonesia
386	Index 37-106	Pulmones	Pulmonar	SRR6480518	Indonesia
387	Index 21-21	Pulmones	Pulmonar	SRR6480530	Indonesia
388	Index 75-106	Pulmones	Pulmonar	SRR6480428	Indonesia
389	DS-19290	Pulmones	Pulmonar	SRR5709929	Tailandia
390	EKB15	Pulmones	Pulmonar	SRR8335313	Rusia
391	Index 37-90	Pulmones	Pulmonar	SRR6480316	Indonesia
392	Index 55-133	Pulmones	Pulmonar	SRR6480322	Indonesia
393	Index 40-101	Pulmones	Pulmonar	SRR6480563	Indonesia
394	Index 55-100	Pulmones	Pulmonar	SRR6480399	Indonesia
395	EKB19	Pulmones	Pulmonar	SRR8344409	Rusia
396	Index 75-100	Pulmones	Pulmonar	SRR6480604	Indonesia
397	JAL_IOB_200	Pulmones	Pulmonar	SRR5341272	India
398	Index 60-106	Pulmones	Pulmonar	SRR6480377	Indonesia
399	DS-19383	Pulmones	Pulmonar	SRR5709962	Tailandia
400	Index 21-133	Pulmones	Pulmonar	SRR6480611	Indonesia
401	DS-30442	Pulmones	Pulmonar	SRR5710012	Tailandia

402	2184	Pulmones	Pulmonar	SRR8662666	Rusia
403	EKB18	Pulmones	Pulmonar	SRR8427096	Rusia
404	DS-21517	Pulmones	Pulmonar	SRR5709812	Tailandia
405	DKC2-0448	Pulmones	Pulmonar	ERR2229324	Dinamarca
406	EKB42	Pulmones	Pulmonar	SRR8434867	Rusia
407	Index 24-21	Pulmones	Pulmonar	SRR6480535	Indonesia
408	L2	Pulmones	Pulmonar	SRR6480540	Indonesia
409	DS-18810	Pulmones	Pulmonar	SRR5709885	Tailandia
410	EKB99	Pulmones	Pulmonar	SRR8433589	Rusia
411	WT1	Pulmones	Pulmonar	SRR1573728	Canadá
412	Index 6-90	Pulmones	Pulmonar	SRR6480453	Indonesia
413	DS-30951	Pulmones	Pulmonar	SRR5709881	Tailandia
414	EKB32	Pulmones	Pulmonar	SRR8353274	Rusia
415	Index 32-100	Pulmones	Pulmonar	SRR6480403	Indonesia
416	Index12-21	Pulmones	Pulmonar	SRR6480531	Indonesia
417	Index 12-109	Pulmones	Pulmonar	SRR6480502	Indonesia
418	DS-10501	Pulmones	Pulmonar	SRR5710030	Tailandia
419	EKB65	Pulmones	Pulmonar	SRR8366603	Rusia
420	JAL_IOB_196	Pulmones	Pulmonar	SRR5341276	India
421	Index 4-133	Pulmones	Pulmonar	SRR6480622	Indonesia
422	EKB73	Pulmones	Pulmonar	SRR8369849	Rusia
423	Index 32-55	Pulmones	Pulmonar	SRR6480542	Indonesia
424	Index15-42	Pulmones	Pulmonar	SRR6480458	Indonesia
425	DS-25553	Pulmones	Pulmonar	SRR5709877	Tailandia
426	Index 32-133	Pulmones	Pulmonar	SRR6480621	Indonesia
427	Index 12-90	Pulmones	Pulmonar	SRR6480520	Indonesia
428	EKB17	Pulmones	Pulmonar	SRR8345980	Rusia
429	Index 78-21	Pulmones	Pulmonar	SRR6480550	Indonesia
430	EKB31	Pulmones	Pulmonar	SRR8353479	Rusia
431	Index 24-106	Pulmones	Pulmonar	SRR6480516	Indonesia
432	DS-32315	Pulmones	Pulmonar	SRR5709896	Tailandia
433	4702	Pulmones	Pulmonar	SRR8662668	Rusia
434	DS-30056	Pulmones	Pulmonar	SRR5709960	Tailandia
435	DS-19155	Pulmones	Pulmonar	SRR5709927	Tailandia
436	EKB63	Pulmones	Pulmonar	SRR8366172	Rusia
437	Index 30-25	Pulmones	Pulmonar	SRR6480541	Indonesia
438	EKB26	Pulmones	Pulmonar	SRR8351971	Rusia
439	DKC2-0446	Pulmones	Pulmonar	ERR2229322	Dinamarca
440	Index 8-90	Pulmones	Pulmonar	SRR6480596	Indonesia
441	EKB48	Pulmones	Pulmonar	SRR8358462	Rusia
442	DCTB003	Pulmones	Pulmonar	SRR993002	Rusia

443	WT15	Pulmones	Pulmonar	SRR1573727	Canadá
444	DCTB020	Pulmones	Pulmonar	SRR1002686	Rusia
445	DS-32512	Pulmones	Pulmonar	SRR5710028	Tailandia
446	Index 21-90	Pulmones	Pulmonar	SRR6480314	Indonesia
447	DCTB001	Pulmones	Pulmonar	SRR974700	Rusia
448	Index 78-100	Pulmones	Pulmonar	SRR6480603	Indonesia
449	Index 24-42	Pulmones	Pulmonar	SRR6480418	Indonesia
450	EKB76	Pulmones	Pulmonar	SRR8370129	Rusia
451	DS-11221	Pulmones	Pulmonar	SRR5709796	Tailandia
452	Index 15-142	Pulmones	Pulmonar	SRR6480459	Indonesia
453	BC14-Mtb515	Pulmones	Pulmonar	SRR6397999	Canadá
454	DCTB022	Pulmones	Pulmonar	SRR1002690	Rusia
455	EKB74	Pulmones	Pulmonar	SRR8369893	Rusia
456	DKC2-0483	Pulmones	Pulmonar	ERR2229359	Dinamarca
457	DS-21644	Pulmones	Pulmonar	SRR5709928	Tailandia
458	DS-08775	Pulmones	Pulmonar	SRR5709793	Tailandia
459	Index 40-66	Pulmones	Pulmonar	SRR6480444	Indonesia
460	Index 8-106	Pulmones	Pulmonar	SRR6480510	Indonesia
461	EKB47	Pulmones	Pulmonar	SRR8358501	Rusia
462	DCTB014	Pulmones	Pulmonar	SRR1002679	Rusia
463	Index 73-42	Pulmones	Pulmonar	SRR6480357	Indonesia
464	DS-20815	Pulmones	Pulmonar	SRR5709773	Tailandia
465	DS-6265	Pulmones	Pulmonar	SRR5709938	Tailandia
466	DKC2-0489	Pulmones	Pulmonar	ERR2229365	Dinamarca
467	Index 75-21	Pulmones	Pulmonar	SRR6480549	Indonesia
468	Index 69-100	Pulmones	Pulmonar	SRR6480602	Indonesia
469	DS-20500	Pulmones	Pulmonar	SRR5709776	Tailandia
470	DS-29188	Pulmones	Pulmonar	SRR5709884	Tailandia
471	EKB22	Pulmones	Pulmonar	SRR8347539	Rusia
472	Index 28-42	Pulmones	Pulmonar	SRR6480412	Indonesia
473	Index 18-21	Pulmones	Pulmonar	SRR6480529	Indonesia
474	Index 75-90	Pulmones	Pulmonar	SRR6480420	Indonesia
475	DS-15203	Pulmones	Pulmonar	SRR5709959	Tailandia
476	DS-19351	Pulmones	Pulmonar	SRR5709964	Tailandia
477	Index 60-66	Pulmones	Pulmonar	SRR6480441	Indonesia
478	EKB21	Pulmones	Pulmonar	SRR8346061	Rusia
479	Index 4-90	Pulmones	Pulmonar	SRR6480315	Indonesia
480	DS-18613	Pulmones	Pulmonar	SRR5709952	Tailandia
481	DS-11380	Pulmones	Pulmonar	SRR5709792	Tailandia
482	DS-17658	Pulmones	Pulmonar	SRR5709759	Tailandia
483	RTS-4	Pulmones	Pulmonar	SRR6914121	Rusia

484	EKB34	Pulmones	Pulmonar	SRR8439237	Rusia
485	DS-34062	Pulmones	Pulmonar	SRR5709860	Tailandia
486	DKC2-0487	Pulmones	Pulmonar	ERR2229363	Dinamarca
487	Index 78-142	Pulmones	Pulmonar	SRR6480354	Indonesia
488	DS-21639	Pulmones	Pulmonar	SRR5709813	Tailandia
489	Index 18-109	Pulmones	Pulmonar	SRR6480427	Indonesia
490	Index 21-142	Pulmones	Pulmonar	SRR6480415	Indonesia