Variabilidad genética en aislados clínicos de *Mycobacterium tuberculosis* resistentes a medicamentos antituberculosos de primera y segunda línea.

Estudiante: Gabriel Morey León
Directora Académica: Dra. Luisa Berná Zanotta
Co-director: Dr. Carlos Robello Porto

Tribunal: Dra. Rosario Durán, Dr. Andrés Iriarte, Dra Andrea Villarino

AGRADECIMIENTOS

A Dios por brindarnos una oportunidad de vida, luego de una pandemia sufrida que tuvimos que sobrellevar todos.

Al Programa PEDECIBA, por haberme brindado la oportunidad de pertenecer al mismo como estudiante y lograr la formación requerida.

A Luisa Berná, mi tutora académica, por darme confianza, ánimos, consejos y su amistad, además de la libertad de pensamiento para el desarrollo de mi trabajo de tesis.

A la Unidad de Bioinformática del Institut Pasteur de Montevideo, por su amabilidad, apertura y confianza brindada durante mi estancia formativa.

A Gulnara Borja, como Directora Ejecutiva en su momento del Instituto Nacional de Investigación en Salud Pública "INSPI-LIP"por su confianza y apoyo oportuno otorgando el acceso a aislados para la continuación de este trabajo de tesis.

A Juan Carlos Fernandez, Derly Andrade y los demás colegas del Laboratorio de Ciencias Ómicas, por su apoyo y equipamiento proporcionado para la realización de esta tesis.

A la Universidad Espíritu Santo, por el acceso al equipamiento y espacio para desarrollo de esta tesis.

A la Universidad de Guayaquil, por el financiamiento brindado como parte de los Fondos Concursables de Investigación FCI.

En especial mis padres, Pedro y Flor por siempre darme ánimos y no dejarme desfallecer cuando las situaciones se pusieron complicadas.

A todas las personas que de una u otra forma siempre han estado ahí apoyando para la culminación de esta etapa formativa.

Gracias, totales.

TABLA DE CONTENIDOS

RESUMEN **4**

ABREVIACIONES **5**

INTRODUCCIÓN **8**

Genoma de Mycobacterium tuberculosis 9

Tratamiento 12

Mecanismos de Resistencia 14

Diagnóstico 17

Distribución de la tuberculosis 20

Tuberculosis en Ecuador 22

HIPÓTESIS **25**

OBJETIVOS **26**

Objetivo general 26

Objetivos Específicos 26

METODOLOGÍA **28**

Artículo 1: Comparative genomics of drug-resistant strains of Mycobacterium tuberculosis in Ecuador. BMC genomics (https://doi.org/10.1186/s12864-022-09042-1) 31

Artículo 2: Transmission network and genomic diversity of Mycobacterium tuberculosis L4 lineage in Ecuadorian isolates provided clues to transoceanic epidemiology 48

Artículo 3: A precision overview of genomic resistance screening in Ecuadorian isolates of Mycobacterium tuberculosis using web-based bioinformatics tools. PlosOne (https://doi.org/10.1371/journal.pone.0294670). 71

Artículo 4: Decoding Ecuadorian Mycobacterium tuberculosis Isolates: Unveiling Lineage-Associated Signatures in Beta-Lactamase Resistance via Pangenome Analysis. 92

DISCUSIÓN **113**

CONCLUSIONES **120**

BIBLIOGRAFÍA **121**

RESUMEN

A pesar de la extensa inversión mundial anual destinada al control de la tuberculosis, esta enfermedad continúa siendo una de las principales causas de muerte a nivel global, ocupando el segundo lugar después de la COVID-19 en términos de enfermedades infecciosas. La identificación temprana y precisa de los aislados de *Mycobacterium tuberculosis* es fundamental para garantizar un tratamiento adecuado a los pacientes y reducir el impacto en la población, contribuyendo así a alcanzar los objetivos de eliminación de la tuberculosis para el año 2030.

En el caso de Ecuador, a pesar de contar con un programa de control funcional, el país sigue enfrentando una carga media de casos anuales (48/100.000). Se sabe muy poco sobre los perfiles genéticos asociados a la resistencia que presentan los aislados de *Mycobacterium tuberculosis* en el país. Esto resalta la necesidad de fortalecer los procesos de vigilancia epidemiológica mediante la implementación de técnicas rápidas y eficientes que permitan obtener un perfil completo de *Mycobacterium tuberculosis*.

En este contexto, el objetivo de este trabajo fue ampliar el conocimiento sobre los perfiles de resistencia a drogas de primera y segunda línea presentes en los aislados de *Mycobacterium tuberculosis* circulantes en el país. Para lograr esto, se utilizó el secuenciamiento del genoma completo, lo que contribuirá a la implementación de esta técnica como una estrategia de vigilancia nacional.

En primer lugar, se investigó la diversidad genética de los aislados analizados, lo que permitió identificar un predominio de aislados pertenecientes al linaje Euroamericano, especialmente de las subfamilias LAM y X-type. Estos aislados presentan una composición genética específica que sugiere una mayor adaptabilidad a las condiciones asociadas a la enfermedad, así como una capacidad para evadir la respuesta inmune, una mayor transmisibilidad, latencia y virulencia.

En segundo lugar, se identificaron patrones de mutaciones típicas asociadas a la resistencia a fármacos de primera y segunda línea. En algunos casos, estas mutaciones se encontraron fuera de las regiones habitualmente asociadas a la resistencia y utilizadas en los métodos moleculares estandarizados actualmente recomendados por la OMS. Además, se pudo clasificar como aislados pre-XDR aquellos

casos que previamente habían sido categorizados como MDR utilizando estrategias microbiológicas convencionales.

En tercer lugar, se destacó la importancia del uso de herramientas bioinformáticas para analizar los datos de secuenciación dentro de los programas de vigilancia de la enfermedad. Sin embargo, es crucial fortalecer las capacidades técnicas del personal y dotar de equipamiento de análisis adecuado para garantizar un uso eficiente de la tecnología, tanto en el laboratorio como en la interpretación de los resultados.

Finalmente, a través de la secuenciación del genoma completo, se logró identificar y rastrear posibles rutas de transmisión del bacilo. Se observó que la movilidad migratoria desde y hacia el país contribuye a la introducción de aislados con características de virulencia y resistencia más complejas. Si no se controlan de manera oportuna, estos aislados pueden tener un impacto negativo en el sistema de salud y en la población en general.

Esperamos que este trabajo contribuya al conocimiento de la diversidad genética de *Mycobacterium tuberculosis* circulante en el país y que, a su vez, promueva la implementación de estrategias de vigilancia fortalecidas. Asimismo, buscamos generar conciencia sobre el comportamiento de la tuberculosis en Ecuador y resaltar la importancia de abordarla de manera efectiva.

ABREVIACIONES

TB:	Tuberculosis
Mtb:	*Mycobacterium tuberculosis*
MTBC:	*Mycobacterium tuberculosis* complex
OMS:	Organización Mundial de la Salud
OPS:	Organización Panamericana de la Salud (OPS)
LAM:	Latin Americana y Mediterraneo
LTB:	Tuberculosis Latente
RR:	Rifampicina resistencia
MDR:	Multi-drogo resistencia
XDR:	Extremo-drogo resistencia
DST:	Test de susceptibilidad a drogas
SNP:	Polimorfismo de un solo nucleótido
InDels:	Inserciones y Deleciones
WGS:	Secuenciamiento de Genoma Completo
CDS:	Secuencias codificantes
VFDB:	Base de datos de Factores de Virulencia
TCDB:	Base de datos de clasificación de transportadores
cgMLST:	Tipificación de secuencias multi locus del genoma central
COG:	Clusters of Orthologous Groups
PATRIC:	Pathosystems Resource Integration Center
STPK:	Serina/treonina proteína quinasas de tipo eucariota.

MPTR:	Secuencias de repeticiones polimórficas en tándem
MIRU:	Unidades repetitivas interespecíficas micobacterianas
VNTR:	Repeticiones en tándem de número variable
ORF:	Marcos de lectura abierta
PE	Prolina-Ácido Glutámico
PPE	Prolina-Prolina-Ácido Glutámico
ESAT-6	Blanco 6 antigénico secretorio temprano (early secretory antigenic target 6)
CFP-10	Proteína de cultivo filtrado 10, tipo ESAT-6
Mce	Proteína de ingreso a células de mamífero
ESX	Proteínas del sistema de exportación
Mpt64	Proteínas de transición de permeabilidad mitocondrial 64
Mpt32	Proteínas de transición de permeabilidad mitocondrial 32
PhoP	Proteína reguladora transcripcional
RRDR:	Región Determinante de Resistencia a Rifampicina
NTM:	Micobacterias No tuberculosas
PCR:	Reacción en cadena de la polimerasa
PGRS:	Secuencias repetitivas polimorficas ricas en GC

INTRODUCCIÓN

La Tuberculosis (TB) es considerada una enfermedad re-emergente debido a la prevalencia de variantes genéticas con resistencia a los medicamentos antituberculosos. La Organización Mundial de la Salud (OMS) reportó en el año 2021 que cerca de 10,6 millones de personas contrajeron la enfermedad a nivel mundial, dentro de estas 450.000 fueron RR/MDR-TB (multidrogo resistencia), y acorde a estimaciones el 6,2% de los casos derivan en XDR-TB (extremo-drogo resistencia)(1). De las ocurrencias en 2021 se estima que 1.6 millones de personas fallecieron, 10% de estas muertes son debido a coinfecciones TB+ HIV(2–4).

Con la aplicación de estrategias y esfuerzos de los organismos internacionales, la incidencia de la tuberculosis a nivel mundial ha disminuido en término medio 1,5% anual, pasando de 184 casos en el 2000 a 134 casos en 2021 por cada 100.000 habitantes(5); sin embargo, la pandemia del COVID-19 mermó fuertemente los avances obtenidos, especialmente en la capacidad diagnóstica e incremento de los casos de resistencia a fármacos(6–8). Por otro lado, se estima que casi un cuarto de la población mundial presenta tuberculosis latente (LTB) y que alrededor de 170 millones de personas podrían volverse casos de tuberculosis activa debido a la disminución en la calidad de atención de pacientes con tuberculosis latente, especialmente en lograr el cumplimiento del tratamiento(9,10), lo cual amenaza el cumplimiento de objetivos en los principales programas mundiales asociados con su control(11,12).

Las micobacterias son bacilos aeróbicos, no moviles, las cuales presentan características ácido-alcohol resistente debido al alto contenido de ácido micólico en su pared celular(13,14). Las especies dentro del género *Mycobacterium*, el cual incluye un número creciente de integrantes, se distinguen tradicionalmente en base a características fenotípicas, entre ellas la capacidad de metabolizar diferentes componentes bioquímicos, la velocidad de crecimiento, y la presencia de pigmentos(15,16); identificándose principalmente que las cepas de crecimiento lento son las causantes de enfermedades relevantes en humanos y animales(17–19).

La tuberculosis es causada por un grupo de micobacterias estrechamente relacionadas denominadas Complejo *Mycobacterium tuberculosis* (MTBC), integrado por especies tales como: *M. tuberculosis, M. africanum, M. canettii, M. microti, M. caprae, M. bovis,* y *M. pinnipedii*(20), las cuales se encuentran distribuidas en 9 linajes, divididos entre los "linajes modernos - clado I" conformado por: linaje 2 con cepas del Este de Asia, linaje 3 con cepas de India y África del Este (EAI, por sus siglas en inglés) y linaje 4 con cepas de Europa, América y África y los "linajes ancestrales - clado II" que incluyen cepas de África del Oeste 1 (linaje 5) y 2 (linaje 6) y linaje 1 cepas de Filipinas y Borde del Océano Índico (Figura 1). El linaje 7 ha sido identificado de forma restringida en Etiopía(21,22), mientras que aislados asociados con los nuevos los linajes 8 y 9 recientemente han sido clasificados (23,24). Los integrantes del linaje 2 son los de mayor preocupación epidemiológica debido que están caracterizados por presentar hiper virulencia, alta transmisibilidad y por estar asociados con resistencia a medicamentos, consecuente con la alta tasa de mutación. Uno de sus principales representantes es el sublinaje Beijing(25–28).

Genoma de *Mycobacterium tuberculosis*

El genoma de *M. tuberculosis* H37Rv (NC_000962.3), desde su descifrado por secuenciamiento ha proporcionado información de la fisiología, mecanismos de infección, patogénesis, virulencia y persistencia(29). Este genoma consta de 4.41 Mb la cual se encuentra conformada por 4173 genes (3924 proteínas) incluyendo 30 pseudogenes, 45 tRNA, 30 ncRNA y 2 miscRNA, con un contenido de GC de 65.6% en casi todo el genoma, excepto en las regiones PGRS las cuales contienen un mayor porcentaje de GC. El genoma de *M. tuberculosis* consta de una gran cantidad de regiones repetitivas, las más estudiadas son las secuencias de inserción (IS6110 e IS1081)(30–33),y las secuencias de repeticiones polimórficas en tándem (MPTR), las cuales están asociadas con variaciones antigénicas(34). Estas MPTR en *M. tuberculosis* están presentes dentro de ciertos locus genéticos, tales como Unidades repetitivas interespecíficas micobacteriales (MIRU, por sus siglas en inglés) o repeticiones en tándem de número variable (VNTR, por sus siglas en inglés) los cuales han sido ampliamente utilizados para estudios de epidemiología molecular de TB(35,36). Entre las regiones codificantes, *M. tuberculosis* presenta una abundancia considerable de proteínas responsables de su

metabolismo, adaptabilidad, resistencia y virulencia. Estas proteínas son objeto constante de estudio con el objetivo de desarrollar estrategias de control eficiente dentro en su entorno de desarrollo.

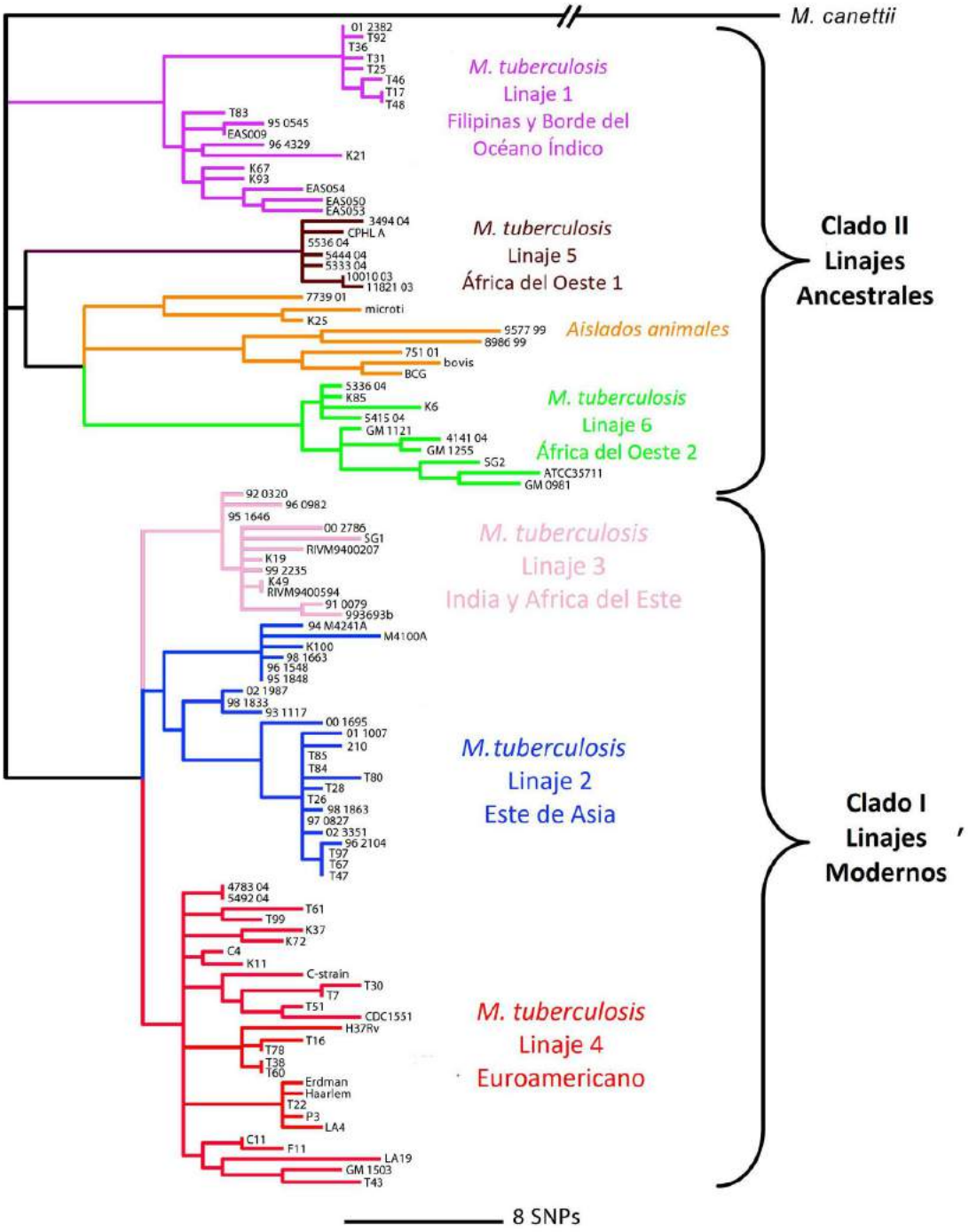

Figura 1. Relaciones filogenéticas entre las especies del Complejo *M. tuberculosis*. Análisis filogenético está basado en SNP de 108 secuencias en la secuencia concatenada de 89 genes y análisis de deleciones. Los colores representan los linajes congruentes. El Sublinaje Beijing dentro del Linaje 2 se encuentra de color azul. *M. canetti* fue utilizado como grupo externo. Figura adaptada de la referencia(20)

La composición proteica de *M. tuberculosis* es diversa y contribuye a su patogenicidad y supervivencia dentro del huésped. Se identificaron 3924 marcos de lectura abierta (ORF) en su genoma, de los cuales se pudieron identificar funciones para el 50% y se agruparon acorde a las actividades relevantes que desempeñan. En particular las proteínas de *M. tuberculosis* se asocian a: proteínas de metabolismo general, regulación y resistencia a drogas, proteínas del metabolismo de lípidos, proteínas que afectan la respuesta inmunológica del huesped y de patogenicidad. Entre las proteínas asociadas con metabolismos se ha podido identificar que *M. tuberculosis* es capaz de metabolizar una variedad de carbohidratos, hidrocarburos, alcoholes, cetonas y ácido carboxílico, además puede sintetizar más de 200 enzimas envueltas en los principales procesos metabólicos bajo condiciones aerobicas y anaerobicas(37–39).

Dada la complejidad del ambiente donde crece esta bacteria, un amplio repertorio de proteínas de regulación y transducción de señal son sintetizadas, entre ellas se han identificado trece factores sigma y más de 100 proteínas reguladoras. Además se han identificado 11 Serina/treonina proteína quinasas de tipo eucariota (STPK), caracterizadas como *PknA, PknB,* y de la *PknD-L*, las cuales juegan un papel importante en la transducción de señales, y pueden dirigir importantes procesos celulares como ingresar a un estado de dormancia o a división celular, entre otros(40–43), asi mismo han sido caracterizados igual número de sistemas de dos componente lo cual evidencia la importancia de estos dos mecanismos de transducción de señales para la adaptación del patógeno a su ambinete(43). En el caso de proteínas asociadas con el metabolismo de lípidos, *M.tuberculosis* es uno de los pocos microorganismos que puede generar una amplia gama de compuestos lipídicos, desde ácidos grasos simples, pasando por isoprenoides y llegando a moléculas complejas como ácidos micólicos entre otros, esto conlleva que existan más de 250 enzimas involucradas en estas vías(19,44,45).

Cerca del 10% de las secuencias codificantes se agrupan en dos grandes familias de proteínas ácidas ricas en glicina, las familia PE y PPE, las cuales están compuestas por secuencias repetitivas, PGRS y MPTR. Ambas familias poseen un amplio número de variantes (91 para PE y 68 para PPE). Varios estudios respaldan la hipótesis que estas familias están asociadas con la variación antigénica de la

bacteria, lo que evita la homogeneidad antigénica de la misma y conduce a una interferencia en la respuesta inmunológica del huésped al inhibir el procesamiento de los antígenos(34,46).

Por último, existe un grupo de proteínas que están asociadas a la virulencia de las micobacterias, las cuales están principalmente involucradas en la interacción de las especies de MTBC con los macrófagos del huésped. Un grupo de estos factores de virulencia está involucrado en el proceso de adaptabilidad a condiciones limitadas de nutrientes e iones, así como en la modificación del metabolismo de carbono dentro de las células del huésped. Otro grupo es responsable de la respuesta frente al ataque microbicida de las células del huésped. Un último grupo está vinculado con las familias PE y PPE que son responsables de la modulación de la respuesta inmune del huésped, especialmente los factores relacionados con la inhibición de la respuesta inflamatoria y la apoptosis. Entre los factores más estudiados se encuentran ESAT-6, CFP-10, Mce, ESX, PepX, Mpt64, Mpt32, PhoP, entre otras(47–51).

Tratamiento

El tratamiento de *M. tuberculosis*, generalmente implica una combinación de diferentes medicamentos que se toman durante un período específico. El tratamiento estándar para la tuberculosis sensible a los medicamentos consiste en un régimen de cuatro medicamentos conocidos como de "primera línea"(52). Los medicamentos específicos utilizados, sus dosis y la duración del tratamiento pueden variar según factores como la edad del paciente, el estado general de salud, los resultados de las pruebas de susceptibilidad a los medicamentos y la presencia de cualquier otra afección médica(53).

Los medicamentos de primera línea comúnmente utilizados en el tratamiento de la tuberculosis incluyen:

Isoniazida (INH): medicamento que inhibe el crecimiento de *M. tuberculosis* al interferir con la síntesis de ácidos micólicos, un componente crítico de la pared celular bacteriana. La actividad antibacteriana de la INH comprende la oxidación enzimática de la INH por la catalasa-peroxidasa bacteriana (katG) para formar radicales de isonicotinoilo, que a su vez forman un aducto químico

clave con nicotinamida adenina dinucleótido (INH-NAD+). Este aducto es un potente inhibidor de la proteína transportadora de enoil acil reductasa, una enzima esencial para la biosintesis del ácido micólico de la pared celular de *M. tuberculosis*(54,55). Rifampicina (RIF): la rifampicina es un agente bactericida que inhibe la síntesis de ARN, lo que interrumpe la producción de proteínas esenciales. RIF se une a la subunidad β de la ARN polimerasa dependiente de ADN (RNAP), cerca del canal de ARN/ADN, inhibiendo así la subunidad β de la RNAP bacteriana, bloqueando físicamente el alargamiento de la cadena de ARN en crecimiento después de que se hayan añadido 2 - 3 nucleótidos(56,57). Pirazinamida (PZA): es un profármaco que se convierte a su forma activa de ácido pirazinoico (POA), por la acción de la pirazinamidasa (PZasa)/nicotinamidasa, codificada por el gen pncA en *M. tuberculosis*. Se plantea que la pirazinamida actúa alterando el pH intracelular de *M. tuberculosis*, debido al ingreso de POA facilitada por ácido (HPOA ácido protonado sin carga) a través de la membrana, provocando una acumulación de POA en el interior, lo que hace que el ambiente sea menos favorable para el crecimiento bacteriano, llegando a desenergizar la membrana al colapsar la fuerza motriz de protones afectando el transporte de la membrana, además de inhibirla síntesis de proteínas y ARN(58). Etambutol (EMB): El etambutol se considera un fármaco bacteriostático que bloquea la polimerización de las subunidades de arabinosa en la capa de arabinogalactano (AG), un componente de la pared celular bacteriana, lo que conduce a la pérdida de la capa de ácidos micólicos de cadena larga (MAs), impidiendo el crecimiento bacteriano debido a la detención en los polos celulares(59).

El régimen de tratamiento estándar para la tuberculosis sensible a los medicamentos se conoce como "Terapia de observación directa, de ciclo corto" (DOTS, por sus siglas en inglés), y generalmente dura seis meses(60). En la fase inicial, tiene una duración de dos meses, donde se administrarán los cuatro fármacos (INH, RIF, PZA y EMB). Luego, en la fase de continuación, que dura cuatro meses adicionales, se continúa con INH y RIF, mientras que PZA y EMB pueden suspenderse si se descubre que las bacterias son susceptibles a INH y RIF.

Es importante tener en cuenta que el régimen de tratamiento exacto puede variar según las circunstancias individuales del paciente y los resultados de las pruebas de susceptibilidad a los

medicamentos. Si el tratamiento de primera línea fracasa, la tuberculosis resistente requiere otros regímenes de tratamiento que utilizan medicamentos de "segunda línea". Generalmente estos tratamientos son más potentes pero requieren mayor duración y pueden tener más efectos secundarios entre estos constan Levofloxacina y/o Moxifloxacina (Fluoroquinolona), Estreptomicina, Kanamicina, Capreomicina y/o Amikacina, Bedaquilina, Etionamida, Cicloserina, Linezolidos, Clofazimina(61).

Es fundamental que los pacientes completen el tratamiento según lo prescrito, incluso si comienzan a sentirse mejor antes de completarlo, esto ayuda a prevenir el desarrollo de cepas resistentes a los medicamentos y asegura la erradicación completa de la bacteria del cuerpo.

Mecanismos de Resistencia

Los mecanismos de resistencia pueden clasificarse como intrínsecos o adquiridos (62,63). En el caso de los intrínsecos están asociados a características fisiológicas o celulares inherentes del bacilo como son permeabilidad de la pared celular, dormancia, bombas de flujo, canales de porinas, modificaciones específicas de blancos de drogas, degradación enzimática, activación de reguladores de transcripción de genes asociado con resistencia entre otros(64,65). Mientras que la resistencia adquirida se genera a través de la transferencia horizontal de genes relacionados con resistencia a drogas o mutaciones espontáneas en genes que naturalmente son blanco de drogas(65–67). En el caso de *M. tuberculosis*, no se ha evidenciado que ocurra transferencia horizontal de manera significativa, o en caso de que ocurra, es relativamente rara debido al ambiente donde se desarrolla. Por lo tanto, la resistencia a los medicamentos está principalmente relacionada con el desarrollo de mutaciones espontáneas en los genes diana de los fármacos utilizados en el tratamiento.

La resistencia bacteriana debido a mutaciones en *M. tuberculosis* se produce a través de cuatro mecanismos distintos: modificación de blanco de la droga, disminución de la activación de la prodroga, sobreexpresión del blanco de la droga y sobreexpresión de enzimas modificadoras de la droga. En todos estos casos, la aparición de resistencia afecta la eficacia de las terapias medicamentosas(68,69). Dependiendo de los genes mutados y del número de fármacos afectados, se pueden identificar diferentes perfiles de resistencia en *M. tuberculosis*(70):

- Monorresistente: cuando las mutaciones sólo afectan a un fármaco del tratamiento.
- Poliresistente: implica resistencia a más de un medicamento antituberculoso, excepto a isoniazida y rifampicina.
- Multidrogorresistente (MDR-TB): se presenta cuando hay mutaciones que confieren resistencia a isoniazida y rifampicina.
- Pre-extremo drogorresistente (pre-XDR-TB): además de la multidrogorresistencia, presenta resistencia a cualquier fluoroquinolona.
- Extremadamente resistente o extremo drogorresistente (XDR-TB): incluye la multidrogorresistencia, resistencia a cualquier fluoroquinolona de última generación y al menos una de las tres drogas inyectables de segunda línea.
- Panresistente: cuando las mutaciones han conferido resistencia a todos los fármacos comúnmente empleados en el tratamiento de primera y segunda línea.

La resistencia de *M. tuberculosis* a la Isoniacida ha sido reportada en varios estudios y está relacionada con mutaciones en varios genes importantes en el ciclo de vida de *M. tuberculosis* tales como: *katG*, codifica para la enzima catalasa peroxidasa, mutaciones en este gen y/o su promotor se relacionan con el 60% de los casos de resistencia; el promotor del gen *inhA* que codifica la síntesis de la proteína enil ACP reductasa; el gen *ahpC*, codifica la enzima alquil hidroperóxido reductasa; el gen *nhd* que codifica la enzima NADH deshidrogenasa, esencial en la cadena respiratoria; *kasA* que codifica para la cetoacil-ACP sintasa; el gen *oxyR* regulador central en la respuesta al estrés oxidativo y de nitratos; *furA* es un regulador de la captación del hierro; *fabG1* codifica para una 3-oxoacyl-thioester reductasa(71–74).

Mientras la resistencia a la Rifampicina en *M. tuberculosis* está asociada principalmente con mutaciones en una región Determinante de Resistencia a Rifampicina (RRDR, por sus sigla en inglés) de 81 bases del gen rpoB(75–77) que codifica para la subunidad β de la RNA polimerasa. Las mutaciones asociadas a la región RRDR comprenden los codones 426 a 452, siendo los codones 435, 445 y 450 los que mayoritariamente se han identificado, en especial en cepas altamente transmisibles (MDR y XDR). Las mutaciones en RRDR que confieren resistencia a Rifampicina tienen un impacto

en la capacidad adaptativa del microorganismo(78,79). Sin embargo, en varios estudios se han identificado mutaciones compensatorias que se encuentran fuera de la región RRDR(79,80), así como en los genes adyacentes rpoA y rpoC(81–85), que contribuirían en la generación de resistencia a rifampicina.

La Pirazinamida (PZA) es una pro-droga, empleada para el tratamiento de tuberculosis en combinación con otras drogas, esta debe ser activada por la enzima pirazinamidasa (Pzasa) bacteriana codificada por el gen pncA, variaciones o mutaciones en este gen conducen a resistencia a PZA, por pérdida de actividad PZasa(86). Mutaciones tipo SNP o pequeñas deleciones son diversas y están altamente dispersas en el gen pncA y ha sido reportado que el 70% de las cepas MDR y 96% de las XDR de *M. tuberculosis*(86–91). Aunque mutaciones en el gen pncA han sido reportadas por diferentes autores, es necesario considerar la ocurrencia de mutaciones en otros genes como panD y rpsA identificadas en los aislados resistentes a PZA (92–96), así como la heterorresistencia debido a la discordancia en los resultados con las pruebas fenotípicas clásicas(97,98).

El Etambutol (EMB) afecta la arabinosil transferasa codificada por el locus del gen embCAB que está implicado en la biosíntesis de los componentes de la pared celular, arabinogalactano y lipoarabinomanano(99–103). Las cepas resistentes se caracterizan principalmente por mutaciones canónicas notificadas en los codones 306, 406 y 497 dentro del gen embB; sin embargo, aislados con mutaciones en los codones 320, 324, 397, 445, 1024 del gen embB y 13 de embC han sido asociadas con resistencia a EMB(104). Adicionalmente se han podido encontrar las mutaciones no sinónimas en los codones 188, 237, 240 y 249 del gen ubiA en aislados que han presentado una resistencia incrementada a EMB(105). A pesar de que la mutaciones M306V y M306I son la más frecuente en los aislados resistentes, esta ha sido detectada en aislados sensibles en diferentes estudios, esta discrepancia se puede haber dado debido a una no estandarización de concentraciones críticas de la droga durante las pruebas de sensibilidad.

La Estreptomicina (STR) es una de las alternativas de drogas antituberculosas de primera línea que han sido utilizadas en incremento para tratar MDR-TB(106,107). La estreptomicina como todos los

aminoglucósidos es un bactericida que interfiere con la síntesis de péptidos y proteínas ribosomales, mediante la union al 16S ARNr en la subunidad 30S, lo cual detiene la síntesis de proteina por inhibición de la formación de enlaces peptídicos(108). La resistencia ha sido asociada con mutaciones dentro del gen rrs que codifica para la subunidad pequeña del 16S rRNA y del gen rpsL que codifica para la proteína ribosomal S12. Las mutaciones más frecuentes están localizadas alrededor de dos regiones de rrs, denominadas bucle 530 y la región 912, o los codones 43 y 88 dentro de rpsL(106,109–112). Mutaciones dentro del gen gidB que codifican en una 7-metilguanosina metiltransferasa específica del 16S rRNA, podrían conferir una baja resistencia a STR en *M. tuberculosis*(113–116). Varias investigaciones han mostrado que el tipo y frecuencia de las mutaciones en los genes rrs, rpsL y gidB difieren marcadamente entre regiones geográficas(117–119), por lo cual identificar las variantes locales con el fin de desarrollar pruebas de diagnóstico específicas es de primordial importancia.

Diagnóstico

El diagnóstico de la tuberculosis presenta un panorama complejo basado en las manifestaciones variadas de la enfermedad (latente o activa), además de las localizaciones en las cuales se tome la muestra (pulmonar o extrapulmonar). Además se debe tener en cuenta la posibilidad de la presencia de otras micobacterias denominadas No tuberculosas (NTM). Aunque es muy poco común detectar estas micobacterias, pueden afectar a ciertos pacientes, en especial a aquellos con sistemas inmunológicos debilitados. Entre estas micobacterias, se han identificado casos de *M. fortuitum*(120), *M. kansasii*(121), *M. scrofulaceum*(122), *M. abscessus*(123) entre otras.

El diagnóstico y diferenciación oportuna, junto con una efectiva caracterización de la resistencia a los medicamentos antituberculosos, especialmente a isoniacida y rifampicina, en el complejo *M. tuberculosis*, son fundamentales para el control de esta enfermedad. Estos pasos permiten una intervención epidemiológica y terapéutica temprana, lo que resulta crucial en la gestión exitosa de la tuberculosis(124). Los métodos radiológicos, así como de diagnóstico de laboratorio basado en la observación directa por microscopía de bacilos ácido-alcohol resistentes y los métodos clásicos de

cultivo, se mantienen vigentes como métodos estandarizados para la detección de la tuberculosis. Además de esto, la detección de resistencia a fármacos a través de métodos fenotípicos plantea desafíos logísticos significativos para lograr una detección precisa y exhaustiva del tipo de micobacteria que presenta el paciente(52). El test de susceptibilidad a drogas (DST, por sus siglas en ingles) en *M. tuberculosis* es generalmente desarrollado después de que una muestra clínica es cultivada, lo cual lo hace un método que toma bastante tiempo y se basa principalmente en medir la inhibición del crecimiento inducido por la droga, comparado con el crecimiento en un medio sin la droga evaluada(125,126). En la actualidad la aplicación de métodos de cultivo rápido automatizado ha permitido al DST analizar diferentes niveles de concentraciones de fármacos a primera y segunda línea del tratamiento(71,127,128), basado en las características fenotípicas de crecimiento.

Los métodos moleculares implementan estrategias basadas en la identificación de regiones específicas que aportan información relacionada con el tipo de micobacteria o el perfil de resistencia a fármacos que posee. Por un lado, estas estrategias permiten la tipificación genética de los aislados acorde a regiones de inserción, como la IS6110 (considerada su estándar de oro). Esta técnica se basa en la amplificación de las copias en número variable del elemento de inserción 6110 y su detección mediante la digestión del producto, a través de su captura por sondas específicas de ADN, generando un perfil específico de cada aislado(30,31,33). Otra estrategia es el Espoligotipificación (Spoligotyping) la cual se basa en la amplificación por PCR y análisis del polimorfismos en secuencias espaciadores presentes en regiones específicas del genoma de *M. tuberculosis* denominadas locus repetido directo (direct repeat o DR por su siglas en inglés). Esta corresponde a un región del cromosoma de *M. tuberculosis* que contiene de 10 a 50 copias de un espaciador de ADN que varía en tamaño y secuencia (entre 37 a 41 pb)(35). Finalmente, la técnica ampliamente utilizada para tipificación genética, considerada como el nuevo estándar de oro, es la amplificación de las regiones MIRU-VNTR (mycobacterial interspersed repetitive units-variable number tandem repeats). Las regiones MIRU consisten en secuencias repetitivas de 40 a 100 pb, y se han identificado 41 de ellas dentro del genoma de *M. tuberculosis*. Considerando se conoce la longitud de cada secuencia repetida, se puede discernir el número de copias acorde al tamaño del amplicón generado, lo cual

permite la diferenciación de los aislados; el poder discriminatorio de la técnica se basa en la combinación de MIRU-VNTR que sea utilizado. Esta técnica de amplificación es muy útil para la tipificación y la caracterización de cepas de *M. tuberculosis*, permitiendo un mayor entendimiento de la epidemiología y la transmisión de la enfermedad

Por otro lado, los métodos moleculares determinan la resistencia a drogas acorde a la composición genética en vez de la resistencia fenotípica. Estos métodos se han aplicado al entorno clínico por su versatilidad y rapidez, incluidas variantes de Xpert MTB/RIF y GenoType MTBDRplus(129–132) que determinan las principales variantes genéticas asociadas a resistencia a drogas de primera línea. Aunque estos métodos son rápidos y sencillos, por un lado *M. tuberculosis* evoluciona continuamente a través de la adquisición de polimorfismos de un solo nucleótido (SNP), inserciones y deleciones (InDels) en su genoma, lo que afecta la capacidad de discriminar cepas. Por otro lado, estos métodos se basan en buscar cambios conocidos y no es capaz de identificar mutaciones de novo(133). En este sentido, la secuenciación del genoma completo (WGS, por sus siglas en inglés) se ha convertido en una herramienta de diagnóstico molecular muy útil en la investigación de la tuberculosis. Es clínicamente relevante para predecir la resistencia a los medicamentos, tipificar los linajes de la bacteria, rastrear la transmisión y definir brotes en forma simultánea utilizando la información obtenida del aislado(134–139). Además, el WGS tiene el potencial de determinar la resistencia a los medicamentos de manera más rápida que las pruebas de sensibilidad a medicamento (DST) tradicionales(140) y no requiere una infraestructura de seguridad biológica(141,142) una vez obtenido el material genético de forma segura.

A pesar de la existencia de diferentes estrategias para el diagnóstico y la caracterización de los aislados de tuberculosis, existen desafíos que deben superarse en los diferentes países. Estos desafíos incluyen la necesidad de contar con una infraestructura mínima adecuada, capacitar y desarrollar las habilidades del personal sanitario y técnico, así como incorporar recursos y tecnología avanzada para una aplicación eficiente. Además, no se deben olvidar los procesos de estandarización necesarios que deben implementarse para generar resultados validados dentro de programas de vigilancia de esta enfermedad.

Distribución de la tuberculosis

Según el Reporte Global de la Tuberculosis, 2022 elaborado por la Organización Mundial de la Salud, se evidencia que a nivel mundial se reportaron 10.6 millones de personas infectadas con tuberculosis en 2021, lo cual representa un incremento del 4.5% en comparación con 2020. Estos casos están distribuidos en 202 países, con la mayoría concentrada en el Sudeste Asiático (45%) y África (23%). América, por su parte, representa solo el 2.9% de la carga global de tuberculosis (Figura 2). Treinta países son responsables del 87% de los casos, siendo India (28%), Indonesia (9.2%) y China (7.4%) los que acumulan la mayor cantidad de casos. Junto con Filipinas, Pakistán, Nigeria, Bangladesh y la República Democrática del Congo, estos países representan más de dos tercios de la incidencia global (Figura 2). Por otro lado, existen 47 países con una baja carga de casos de tuberculosis, con menos de 10 casos por 100,000 habitantes por año. Estos países se encuentran distribuidos en América, Europa y algunos países de las regiones del Mediterráneo Oriental y el Pacífico Occidental.

Del total de casos reportados en 2021, se determinó que el 6.7% correspondía a personas que viven con VIH. Los países africanos presentaron la mayor cantidad de casos, llegando a superar el 50% en algunos países del sudeste africano. En cuanto a la resistencia a los medicamentos, se observó un incremento del 3.1% en los casos de tuberculosis resistente en comparación con 2020, especialmente en personas que desarrollaron tuberculosis resistente a rifampicina y/o a múltiples fármacos (RR/MDR), lo cual se atribuye al impacto de la pandemia de la COVID-19. India, Federación Rusa y Pakistán representaron el 42% del total de casos de pacientes con resistencia.

En 2021, el diagnóstico de casos de tuberculosis se vio afectado adversamente por la pandemia, con un total de 6.4 millones de casos reportados, alcanzando niveles similares a los obtenidos en 2017, pero aún lejos de los 7.1 millones alcanzados en 2019. A pesar de la recomendación de la OMS de utilizar pruebas moleculares rápidas para una detección temprana, solo el 38% de los diagnósticos se realizaron mediante estas técnicas. A nivel mundial, 10 países representaron el 75% de la brecha existente entre la incidencia estimada y los casos reportados de tuberculosis. India (24%), Indonesia (13%), Filipinas (10%), Pakistán (6.6%) y Nigeria (6.3%) son los países que presentan las mayores diferencias, principalmente debido a subregistro o subdiagnóstico de casos.

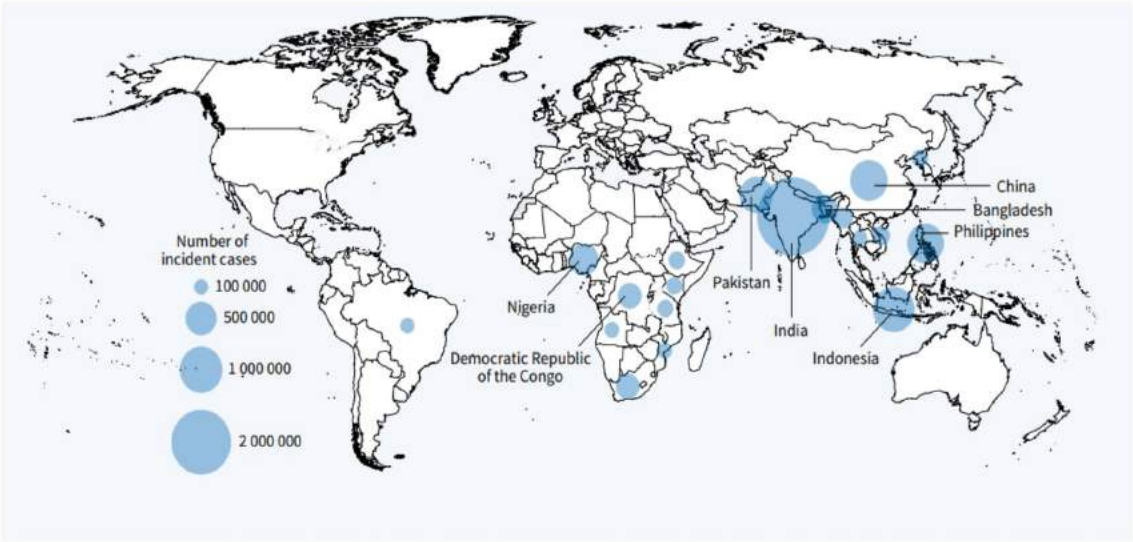

Figura 2. Estimación de la incidencia de Tuberculosis en 2021, en países con al menos 100.000 casos. Los primeros ocho países que conforman las ⅔ partes de los casos globales, son marcados, el tamaño de los círculos representa la incidencia. Figura tomada de la referencia (3).

A nivel mundial, debido a la migración especialmente a países con altos ingresos, la transmisión de la tuberculosis se ha visto incrementada, especialmente enfocado en los linajes más ampliamente distribuidos (L2: Euroasiática y L4: Euroamericano), porque se ha demostrado que diferentes linajes pueden correlacionarse con diferentes resultados epidemiológicos y potenciales de enfermedades(143,144). Este monitoreo de linaje circulantes con el desarrollo de herramientas moleculares como Espoligotipado(35), MIRU-VNTR(145) y Secuenciamiento de Genoma completo (en las dos primeras basadas en la diferenciación de regiones repetidas en tándem específicas dentro del genoma y/o locus CRISPR-Cas locus, mientras que la secuenciación emplea todo el genoma analizando sus variaciones), han permitido identificar entre los aislados circulantes investigados la presencia de agrupaciones clonales "Genomic clusters" acorde con las diferencias nucleotidicas entre dos o más casos activos de TB, en especial los polimorfismo de un solo nucelótido (SNP), adicionalmente si a la información genética entre las agrupaciones clonales se le incluye los datos epidemiológicos, clínicos y/o de contactos existentes entre los huéspedes de la micobacteria, se pueden establecer redes de transmisión "transmission networks", las cuales han permitido comprender de mejor manera los diferentes resultados epidemiológicos y dinámica de la enfermedad asociado a regiones geográficas específicas(146).

Tuberculosis en Ecuador

La tuberculosis a nivel mundial según la OMS es una de las 10 principales causas de mortalidad, siendo el principal agente infeccioso que causa mortalidad por sobre VIH/SIDA. En el Ecuador, luego de la evaluación por parte de la Organización Panamericana de la Salud (OPS) en 1999, se recomienda la implementación de la Estrategia DOTS, y a partir de 2001 se inicia la aplicación de esta estrategia en tres provincias del país (Azuay, Guayas y Pichincha), bajo el Proyecto de Fortalecimiento del Programa Nacional de Prevención y Control de la Tuberculosis, pero no es hasta 2006 que se expande el proyecto a todas las dependencias a nivel nacional(147). Ecuador presenta una carga media de tuberculosis (48/100.000 habitantes) y aunque se llevan a cabo programas activos de control, los casos de tuberculosis se han incrementado en los últimos años, pasando de 5474 casos en 2013 a 5973 casos en 2021 (Tabla 1). Por otro lado, los casos de resistencia se han mantenido variables, alcanzando los 419 casos en 2021. Esto muestra la importancia de implementar pruebas de detección de la resistencia rápida, eficientes, oportunas y de bajo costo.

Tabla 1. Casos de tuberculosis en el Ecuador durante los años 2013 a 2021.

Año	N° de casos totales de tuberculosis	Casos resistentes notificados MDR/RR	Inician tratamiento contra MDR
2013	5474	407	95
2014	5352	451	179
2015	5346	635	135
2016	5374	160	155
2017	5815	249	233
2018	6094	252	252
2019	6542	317	317
2020	5062	256	256
2021	5973	419	342

Fuente: Ministerio de Salud Pública, OMS.
(https://app.powerbi.com/view?r=eyJrIjoiZDhjNDM0YmMtOGExOS00ODIxLWEzMjktZDk0NmI4YTAwODgwIiwidCI6ImY2MtbMGI3LWJkMjQtNGIzOS04MTBLTNkYzI4MGFmYjU5MCIsImMiOjh9).

El tratamiento de la tuberculosis se basa en la administración de medicamentos combinados de forma que afectan procesos vitales en el desarrollo de la micobacteria. En Ecuador, la cobertura de tratamiento a los pacientes está en promedio al 62% y se mantiene en esos niveles desde 2012, estando conformado por medicamentos de primera línea como: Isoniacida, Rifampicina, Etambutol,

Pirazinamida, y en caso de presentar resistencia, se emplean medicamentos de segunda línea como: Levofloxacina y/o Moxifloxacina (Fluoroquinolona), Estreptomicina, Kanamicina, Capreomicina y/o Amikacina(52).

El tratamiento subóptimo y mala adherencia a las terapias de drogas son factores que contribuyen al incremento de los casos de resistencia en la tuberculosis, en el contexto local los casos resistentes totales detectados se mantienen alrededor del 10%, a pesar de que exista un incremento en el 2021 de un 34% de casos resistente detectados como impacto de la pandemia debido al confinamiento y disminución en las capacidades diagnósticas. Además, la administración de un tratamiento inadecuado aumenta el riesgo de transmisión de cepas resistentes y la producción de infecciones primarias, ya que puede haber desconocimiento de los patrones de resistencia presentes en la comunidad. La prueba gold estándar a nivel país para la detección de resistencia antimicrobiana sigue siendo el cultivo bacteriano, la cual toma varias semanas (6 a 8) en el caso de *M. tuberculosis*. Pruebas moleculares, tales como, GeneXpert (MTB/RIF), GenoType MTBDR plus, ensayos de sonda lineales, y actualmente el empleo del antigeno LAM han sido utilizados parcialmente para mejorar la detección de *M. tuberculosis* MDR.

En el contexto de la investigación, muy pocos estudios se han realizado a nivel nacional que aborden el origen de los mecanismos que interfieren en la acción de las drogas antituberculosas, así como las mutaciones presentes en los genes relacionados con la resistencia a fármacos. En su mayoría, estos estudios se han limitado a la detección de linajes basada en perfiles de restricción que generan patrones específicos para ciertos tipos de cepas circulantes. Principalmente los esfuerzos se centran en caracterizar los linajes circulantes mediante análisis de espoligotipos y unidades repetidas interespecíficas micobacterianas - repeticiones en tándem de número variable (MIRU-VNTR)(148–151). Estos métodos se complementan con perfiles de resistencia fenotípica y, en casos limitados, con las técnicas moleculares estandarizadas como GeneXpert (MTB/RIF), GenoType MTBDR plus, y SNP-PCR. Estas técnicas se utilizan para investigar el comportamiento epidemiológico del bacilo de la tuberculosis en diferentes contextos, como poblaciones cautivas como prisiones (151) así como en la población general.

Son escasos los estudios que utilizan la secuenciación de genoma completo para analizar la variabilidad genética, los linajes, los patrones de resistencia y los modos de transmisión en *M. tuberculosis*. Hasta ahora, sólo se ha realizado un estudio en el cual se utilizó WGS para dilucidar la asociación de linajes en 8 aislados que presentaban resultados ambiguos generados mediante MIRU-VNTR(152). Con base a lo antes expuesto, y en ausencia de información que permita conocer el estatus real de las variantes genéticas de *M. tuberculosis* surgen las siguientes interrogantes: cuáles son los linajes circulantes de *M.tuberculosis* en Ecuador? ¿Cuáles son las mutaciones circulantes presentes en los genes relacionados a la resistencia de primera y segunda línea en aislados de *M. tuberculosis* circulantes en el Ecuador?, ¿cuáles son los linajes circulantes de *M. tuberculosis* que presentan mayor número de mutaciones?, ¿El método de las proporciones considerando el tiempo de obtención de sus resultados sería la mejor estrategia para la administración del tratamiento antituberculoso a los pacientes?, ¿Al no conocer las variantes genéticas circulantes, es pertinente que se implemente alguna prueba molecular basada en PCR como son los kits comerciales?

HIPÓTESIS

La detección oportuna del *M. tuberculosis*, el agente causal de la tuberculosis, y una caracterización completa son fundamentales para un tratamiento adecuado de la enfermedad. Es importante conocer la diversidad de aislados circulantes y su perfil de resistencia a las drogas de tratamiento, además de desarrollar capacidades en el personal de salud. Esto permitiría detectar rápidamente el patógeno y administrar, un tratamiento específico acorde al perfil global del patógeno en cada paciente, evitando el desarrollo de variantes más resistentes que puedan complicar el tratamiento de la enfermedad y su diseminación en la población.

En la actualidad la WGS, ha fortalecido las capacidades de diagnóstico en los sistemas de vigilancia de patógenos y enfermedades. Permite obtener más información sobre la virulencia, resistencia, transmisibilidad y adaptabilidad del patógeno al huésped. La vigilancia de *M. tuberculosis* ha seguido esta tendencia al aprovechar las ventajas de la tecnología para complementar a los métodos tradicionales de detección del patógeno. Aunque cada vez más países están fortaleciendo la vigilancia genómica de la tuberculosis, la diversidad de aislados identificados en función de su distribución geográfica, perfiles de resistencia y virulencia sigue siendo una limitación para su implementación completa. Por esta razón, la OMS actualizó en 2021 las guías de criterios para identificar mutaciones relacionadas con la resistencia en aislados mono, multi y extremo drogorresistentes, con el objetivo de fortalecer el uso de tecnologías genómicas en la lucha contra la tuberculosis.

Sin embargo, en el país, el uso de tecnología de secuenciación del genoma completo para la vigilancia de la tuberculosis es muy limitado. Se utilizan principalmente técnicas moleculares que se enfocan en mutaciones puntuales asociadas a la resistencia y clasificación filogenética, lo cual limita la toma de decisiones en la práctica clínica debido al tiempo requerido para obtener los resultados. Por lo tanto, nuestra hipótesis es que es posible estudiar la variabilidad genética de la tuberculosis en el país mediante la secuenciación de genoma completo. Esto permitirá identificar los linajes circulantes, las variantes genéticas presentes o asociadas a la resistencia a drogas de primera y segunda línea, y evaluar la confiabilidad de los estudios de casos. Así el presente estudio contribuirá la implementación de estrategias efectivas en la lucha contra la tuberculosis.

OBJETIVOS

Objetivo general

El presente trabajo tiene como objetivo identificar mediante secuenciación masiva las variantes genéticas relacionadas a la resistencia antituberculosa de primera y segunda línea en aislados clínicos de *M. tuberculosis* y su aplicabilidad en la detección oportuna de casos de resistencia. *M. tuberculosis* y aportar en la personalización de los tratamientos antituberculosos en la población afectada, que puedan tener un impacto en las incidencias de MDR.

Objetivos Específicos

Para cumplir con el objetivo general hemos propuesto los siguientes objetivos específicos.

Genotipificar los linajes circulantes en aislados sensibles y resistentes de *M. tuberculosis*. La caracterización de los aislados de *M. tuberculosis* se realizaría mediante el análisis y comparación de regiones blanco y mutaciones específicas de genes que permite la identificación de linajes circulantes para lo cual se empleará algoritmos y/o aplicativos informáticos disponibles en la web.

Determinar la variabilidad de las cepas circulantes en Ecuador a través de la secuenciación masiva de cepas *M. tuberculosis*. Se realizaría el secuenciamiento del genoma completo de aislados sensibles y resistentes de *M. tuberculosis* circulantes de diversas provincias del país e identificarían las variantes genéticas (polimorfismos de un sólo nucleótido y pequeñas inserciones y/o deleciones) así como posibles duplicaciones o pérdidas de genes. Se analizará la variabilidad mediante comparación de secuencias utilizando a *M. tuberculosis* H37Rv (NC_000962.3) como referencia.

Determinar las variantes genéticas circulantes de los aislados de *M. tuberculosis* resistentes a medicamentos de primera y segunda línea mediante secuenciación masiva. Se analizará la presencia de mutaciones en los genes relacionados a la resistencia a medicamentos de primera y segunda línea e

identificarán posibles nuevas mutaciones comparándolas con las secuencias obtenidas de aislados sensibles.

Evaluar la capacidad diagnóstica de las mutaciones identificadas en comparación al método de las proporciones (PSD) para la caracterización rápida de cepas resistentes de *M. tuberculosis*. Se realizará la comparación y validación estadística de los resultados obtenidos entre las mutaciones identificadas mediante secuenciación masiva y resistencia por PSD para la determinación de aislados resistentes a medicamentos, además se evaluaría la capacidad de identificación de rutas de transmisión de la infección por tuberculosis, mediante el análisis de las secuencias generadas.

METODOLOGÍA

Para poder cumplir con los objetivos planteados, en esta sección se detalla un breve resumen de los procesos realizados. Fueron evaluados 91 aislados de *M. tuberculosis* provenientes de diferentes provincias del Ecuador (Figura 3 y Tabla 2). Estos aislados correspondian a cepas colectadas disponibles las cuales fueron proporcionadas por laboratorios privados y del Centro Nacional de Referencia de Micobacterias del Instituto Nacional de Investigación en Salud Pública "Leopoldo Izquieta Pérez" (INSPI-LIP), durante los años 2019 a 2021.

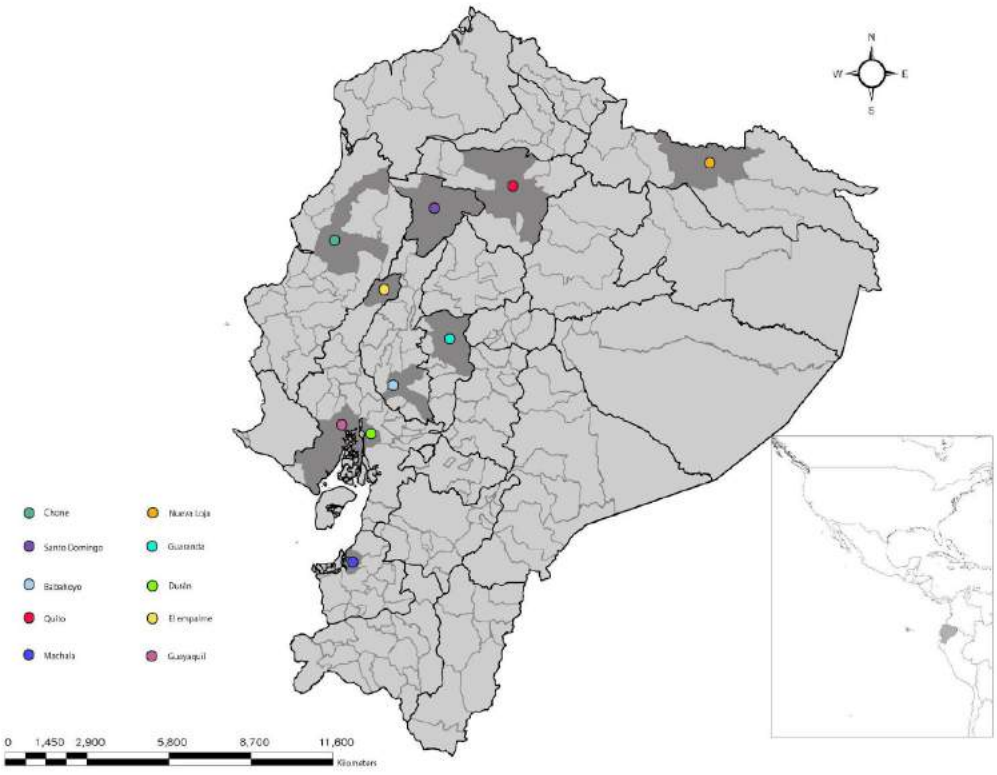

Figura 3. Distribución geográfica por ciudades de los diferentes aislados de *M. tuberculosis* utilizados en el presente estudio.

A todos los aislados se les realizó la evaluación de resistencia mediante método de las proporciones (DST assay) y Bactec MGIT 960 System con lo cual podrían clasificarse entre sensibles y resistentes acorde al patrón de crecimiento en el fármaco evaluado.

Tabla 2. Detalle sociodemográfico, y clínico de los aislados empleados en este estudio.

Muestra	Sexo	Edad	Ciudad	Resultado del tratamiento	Perfil fenotípico
S0001	Mujer	49	Guayaquil	Antes tratado	Resistente
S0007	Mujer	53	Guayaquil	En tratamiento	Resistente
S0008	Mujer	53	Guayaquil	En tratamiento	Resistente
S0013	Mujer	34	El Empalme	En tratamiento	Resistente
S0017	Hombre	34	Guayaquil	En tratamiento	Resistente
S0022	Hombre	38	Guayaquil	Antes tratado	Resistente
S0036	Mujer	27	Guayaquil	Antes tratado	Resistente
S0039	Mujer	26	Guayaquil	Antes tratado	Resistente
S0040	Mujer	29	Guayaquil	En tratamiento	Resistente
S0046	Hombre	71	Babahoyo	Antes tratado	Resistente
S0048*	Mujer	29	Guayaquil	Antes tratado	Resistente
S0059	Hombre	58	Guayaquil	Antes tratado	Resistente
S0062	Hombre	62	Guayaquil	Antes tratado	Resistente
S0066	Hombre	70	Guayaquil	Antes tratado	Resistente
S0067	Hombre	26	Guayaquil	No han recibido tratamiento	Resistente
S0070	Mujer	36	Guayaquil	Antes tratado	Resistente
S0091	Mujer	40	Guayaquil	Antes tratado	Resistente
S0105	Hombre	23	Guayaquil	No han recibido tratamiento	Sensible
S0106	Mujer	21	Guayaquil	No han recibido tratamiento	Sensible
S0106*	Mujer	58	Guayaquil	En tratamiento	Resistente
S0107	Mujer	27	Guayaquil	No han recibido tratamiento	Sensible
S0108	Mujer	64	Guayaquil	No han recibido tratamiento	Sensible
S0110	Hombre	48	Guayaquil	En tratamiento	Sensible
S0111	Mujer	45	Guayaquil	No han recibido tratamiento	Sensible
S0112	Hombre	52	Guayaquil	Antes tratado	Sensible
S0114	Hombre	45	Guayaquil	No han recibido tratamiento	Sensible
S0126	Hombre	62	Guayaquil	En tratamiento	Sensible
S0128	Hombre	23	Guayaquil	No han recibido tratamiento	Sensible
S0134	Hombre	29	Guayaquil	Antes tratado	Resistente
S0137	Hombre	23	Santo Domingo	No han recibido tratamiento	Sensible
S0139*	Mujer	46	Guayaquil	En tratamiento	Resistente
S0145	Hombre	51	Babahoyo	No han recibido tratamiento	Resistente
S0149	Hombre	45	Guayaquil	Antes tratado	Resistente
S0157	Hombre	23	Guayaquil	No han recibido tratamiento	Resistente
S0158	Hombre	27	Guayaquil	Antes tratado	Resistente
S0159	Hombre	53	Guayaquil	No han recibido tratamiento	Resistente
S0165	Mujer	49	Guayaquil	Antes tratado	Resistente
S0166	Mujer	35	Guayaquil	Antes tratado	Resistente
S0194	Mujer	34	El Empalme	En tratamiento	Resistente

S0196	Hombre	28	Chone	Antes tratado	Resistente
S0198	Mujer	52	Guayaquil	No han recibido tratamiento	Resistente
S0202	Mujer	34	El Empalme	Antes tratado	Resistente
S0204	Hombre	43	Guayaquil	Antes tratado	Resistente
S0205	Hombre	45	Guayaquil	Antes tratado	Resistente
S0285	Hombre	54	Guayaquil	En tratamiento	Resistente
S0464	Hombre	33	Guayaquil	En tratamiento	Resistente
S0467	Hombre	11	Guayaquil	No han recibido tratamiento	Resistente
S0468	Hombre	28	Guayaquil	Antes tratado	Resistente
S0498	Hombre	39	Guayaquil	Antes tratado	Resistente
S0506	Hombre	47	Guayaquil	Antes tratado	Resistente
S0516	Hombre	46	Guayaquil	En tratamiento	Resistente
S0518	Hombre	55	Durán	En tratamiento	Resistente
S0555	Hombre	39	Guayaquil	En tratamiento	Resistente
S0585	Hombre	59	Guayaquil	En tratamiento	Resistente
S0605	Hombre	34	Guayaquil	Antes tratado	Resistente
S0654	Hombre	46	Guayaquil	Antes tratado	Resistente
S0660	Hombre	43	Guayaquil	En tratamiento	Resistente
S0813	Hombre	54	Guayaquil	Antes tratado	Resistente
S0868	Hombre	18	Guayaquil	En tratamiento	Resistente
S0889	Hombre	55	Guayaquil	No han recibido tratamiento	Resistente
S0909	Hombre	28	Guayaquil	En tratamiento	Resistente
S1063	Hombre	58	Babahoyo	No han recibido tratamiento	Sensible
S1131	Hombre	37	Guayaquil	En tratamiento	Resistente
S1132	Mujer	37	Guayaquil	No han recibido tratamiento	Resistente
S1133	Hombre	23	Guayaquil	En tratamiento	Resistente
S1135	Hombre	61	Quito	Antes tratado	Resistente
S1136	Hombre	74	Machala	En tratamiento	Resistente
S1205	Hombre	27	Guayaquil	No han recibido tratamiento	Resistente
S1206	Hombre	47	Guayaquil	No han recibido tratamiento	Resistente
S1207	Hombre	27	Nueva Loja	Antes tratado	Resistente
S1208	Mujer	23	Guayaquil	No han recibido tratamiento	Resistente
S1209	Hombre	25	Guayaquil	En tratamiento	Resistente
S1210	Hombre	25	Guayaquil	No han recibido tratamiento	Resistente
S1212	Hombre	24	Guayaquil	No han recibido tratamiento	Resistente
S1451	Hombre	38	Guaranda	Antes tratado	Resistente
S1452	Hombre	37	Guayaquil	Antes tratado	Resistente
S1453	Hombre	39	Guayaquil	Antes tratado	Resistente
S1454	Hombre	38	Guayaquil	Antes tratado	Resistente
S1477	Hombre	33	Quito	No han recibido tratamiento	Resistente
S1478	Mujer	73	Babahoyo	No han recibido tratamiento	Resistente
S1479	Mujer	22	Babahoyo	Antes tratado	Resistente

S2186	Hombre	22	Guayaquil	No han recibido tratamiento	Sensible
S2190	Hombre	38	Guayaquil	No han recibido tratamiento	Sensible
S2191	Hombre	15	Guayaquil	No han recibido tratamiento	Sensible
S2192	Hombre	23	Guayaquil	No han recibido tratamiento	Sensible
S2193	Hombre	49	Guayaquil	No han recibido tratamiento	Sensible
S2202	Hombre	32	Guayaquil	No han recibido tratamiento	Sensible
S2203	Hombre	68	Guayaquil	Antes tratado	Sensible
S2205	Hombre	57	Guayaquil	No han recibido tratamiento	Sensible
S2206	Hombre	52	Guayaquil	No han recibido tratamiento	Sensible
S2209	Hombre	21	Guayaquil	No han recibido tratamiento	Sensible

*Muestas por criterio de calidad de secuenciamiento fueron eliminadas.

A partir de cada aislado, se extrajo el ADN genómico mediante dos metodologías CTAB (24 aislados) y PureLink DNA Mini Kit (67 aislados) (Fisher Scientific, Pennsylvania, USA). Posteriormente se evaluó la pureza, la calidad y la cantidad de ADN mediante electroforesis vertical y el fluorómetro Qubit 4.0 (Invitrogen, Carlsbad, CA, USA.). Para la preparación de las librerías genómicas se usó el kit Tagmentation-based library prep (Illumina Inc., San Diego, CA) acorde a las indicaciones del proveedor. La secuenciación de las muestras se realizó en la plataforma Illumina Miniseq con High Output Reagent Kit (Illumina Inc., San Diego, CA), produciendo lectura "paired-end" de 2x150 bp. Mediante análisis bioinformáticos en pipeline desarrollados por varios autores, y plataformas informáticas como PATRIC, Galaxy, popART, entre otros, se realizó la caracterización de linajes, identificación de variantes genéticas y patrones de resistencia. Se estudiaron las relaciones filogenéticas y se estimaron grupos de transmisión. Un mayor detalle de la metodología, así como los resultados obtenidos se encuentran organizados en cuatro artículos científicos publicados o remitidos para su respectiva evaluación en revistas científicas. A continuación se expone cada artículo con un breve resumen introductorio.

Artículo 1: Comparative genomics of drug-resistant strains of *Mycobacterium tuberculosis* in Ecuador. BMC genomics (https://doi.org/10.1186/s12864-022-09042-1)

En este artículo se presenta la caracterización genómica de los primeros veinticuatro aislados clínicos de *Mycobacterium tuberculosis* resistentes. Esta caracterización se llevó a cabo mediante el análisis de perfiles de resistencia molecular y filogenia utilizando información obtenida del secuenciamiento del

genoma completo de aislados provenientes de diferentes provincias. Como resultado, se identificó que del total de aislados, en 21 aislamientos que cumplieron los criterios de secuenciamiento, 15 fueron identificados como MDR, 4 como pre-XDR (resistentes a una amplia gama de fármacos) y 2 como XDR (extensamente resistentes a fármacos), mientras que tres aislamientos fueron descartados debido a baja calidad. Los sublinajes principales fueron LAM (61,9%) y Haarlem (19%), pero también se identificaron los clados X, T y S. De especial interés es el hecho de que de las seis cepas pre-XDR y XDR, cinco se encontraron en pacientes de género femeninoAdemás, cuatro de estas cepas pertenecían al sublinaje LAM y dos al sublinaje X-type. Se pudo determinar un core genoma de 3.750 genes, distribuidos en 295 subsistemas. Entre estos genes, se identificaron un total de 64 que codifican para proteínas relacionadas con la virulencia y que desempeñan un papel importante en la patogenicidad de *M. tuberculosis*. Además, se encontraron 66 posibles dianas farmacológicas. La mayoría de las variantes observadas resultaron en cambios de aminoácidos no sinónimos y se determinó que los genotipos más comunes confieren resistencia a rifampicina, isoniazida, etambutol, ácido para-aminosalicílico y estreptomicina. Sin embargo, se detectó un aumento de la resistencia a las fluoroquinolonas. Es importante destacar que este estudio reveló por primera vez la variabilidad de cepas resistentes circulantes entre hombres y mujeres en Ecuador, lo que resalta la utilidad de la secuenciación genómica para identificar resistencias emergentes. Se enfatiza el hallazgo en el aumento de la resistencia a las fluoroquinolonas. Estos resultados subrayan la necesidad de aumentar los esfuerzos de muestreo para determinar la variabilidad total y establecer asociaciones con los metadatos recopilados, con el fin de generar políticas de salud más efectivas.

Morey-León *et al. BMC Genomics* (2022) 23:844
https://doi.org/10.1186/s12864-022-09042-1

BMC Genomics

RESEARCH

Open Access

Comparative genomics of drug-resistant strains of *Mycobacterium tuberculosis* in Ecuador

Gabriel Morey-León[1,2,3*], Derly Andrade-Molina[4], Juan Carlos Fernández-Cadena[4] and Luisa Berná[1,5*]

Abstract

Background: Tuberculosis is a serious infectious disease affecting millions of people. In spite of efforts to reduce the disease, increasing antibiotic resistance has contributed to persist in the top 10 causes of death worldwide. In fact, the increased cases of multi (MDR) and extreme drug resistance (XDR) worldwide remains the main challenge for tuberculosis control. Whole genome sequencing is a powerful tool for predicting drug resistance-related variants, studying lineages, tracking transmission, and defining outbreaks. This study presents the identification and characterization of resistant clinical isolates of *Mycobacterium tuberculosis* including a phylogenetic and molecular resistance profile study by sequencing the complete genome of 24 strains from different provinces of Ecuador.

Results: Genomic sequencing was used to identify the variants causing resistance. A total of 15/21 isolates were identified as MDR, 4/21 as pre-XDR and 2/21 as XDR, with three isolates discarded due to low quality; the main sub-lineage was LAM (61.9%) and Haarlem (19%) but clades X, T and S were identified. Of the six pre-XDR and XDR strains, it is noteworthy that five come from females; four come from the LAM sub-lineage and two correspond to the X-class sub-lineage. A core genome of 3,750 genes, distributed in 295 subsystems, was determined. Among these, 64 proteins related to virulence and implicated in the pathogenicity of *M. tuberculosis* and 66 possible pharmacological targets stand out.

Most variants result in nonsynonymous amino acid changes and the most frequent genotypes were identified as conferring resistance to rifampicin, isoniazid, ethambutol, para-aminosalicylic acid and streptomycin. However, an increase in the resistance to fluoroquinolones was detected.

Conclusion: This work shows for the first time the variability of circulating resistant strains between men and women in Ecuador, highlighting the usefulness of genomic sequencing for the identification of emerging resistance. In this regard, we found an increase in fluoroquinolone resistance. Further sampling effort is needed to determine the total variability and associations with the metadata obtained to generate better health policies.

Keywords: Tuberculosis, Drug-resistance, Lineages, Ecuador, Pan Genome

Introduction

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* (Mtb), is one of the top 10 causes of death in 2020. The World Health Organization estimated that 9.9 million people were infected, and 1.5 million died [1]. Despite constant efforts over the years to lessen the impact of this disease, several challenges, such as patient adherence, long duration of treatment, and late diagnosis, have reduced the effectiveness of TB therapy [2]. In addition, the frequent emergence of drug-resistant strains is a serious global threat and poses significant challenges to public health, especially in low- and middle-income countries.

*Correspondence: garielmorey@uees.edu.ec; lberna@pasteur.edu.uy

[1] Laboratorio de Interacciones Hospedero-Patógeno, Unidad de Biología Molecular, Institut Pasteur de Montevideo, Montevideo, Uruguay
Full list of author information is available at the end of the article

Drug resistance in bacteria, particularly in *M. tuberculosis*, negatively affects health programs worldwide due to the increase of resistant strains that cannot be treated with existing anti-TB drug therapy. Indeed, tuberculosis is considered to be one of the "Big three" infectious diseases worldwide [3]. Patients with resistant TB may develop it due to clinical factors such as late or inappropriate diagnosis, ineffective treatment, poor compliance to the regimen, and exposure to circulating multidrug-resistant and extensively drug-resistant tuberculosis (MDR/XDR-TB) strains [2]. The slow growth of Mtb slows diagnosis and consequently limits the timely detection of resistance to anti-TB drugs. This contributes to the rising incidence of MDR and XDR TB worldwide. Conversely, the accumulation of point mutations in coding regions for drug targets or drug-converting enzymes is an essential mechanism for acquiring resistance in Mtb [4].

The *Mycobacterium tuberculosis* complex is composed of the human-adapted members, the *Mycobacterium tuberculosis* sensu stricto (lineages 1, 2, 3, 4, and 7) and *M. tuberculosis var. africanum* (lineages 5 and 6; *M. africanum*), distinct phylogenetic lineages that have evolved over millennia [5–7] and lineage 8 and 9 are new lineages recently discovered in Africa [8]. Research on the distribution of lineages of Mtb in South America has shown that the Euro-American lineage (lineage 4) has a variable distribution of lineages/sub-lineages between and within these countries, establishing that the predominant strains in their America have evolved from the Euro-American lineage [9]. This lineage has been described as more transmissible than others [10]. From 10 sub-lineages, including in Lineage 4, LAM (L4.3), Haarlem (L4.1.2), X-type and T families are the most observed members of the Euro-American lineage in South and Central America and the Caribbean, a distribution profile shared with Europe and Middle Africa [11–14].

In 2015, due to drug resistance rising worldwide, the WHO proposed expanding rapid testing and detection of cases as one of the five high-priority actions to tackle the global DR-TB crisis [15]. As part of these actions, some molecular biology-based diagnostic methods have been applied in a clinical setting for their versatility and speed, including variants of Xpert MTB/RIF and GenoType MTBDRplus [16–19]. Although these methods are rapid and straightforward, Mtb continuously evolves through the genomic acquisition of single nucleotide polymorphisms (SNPs), insertions, and deletions (indels), impacting the ability to discriminate strains without the classic regions used for resistance detection [20]. Whole-genome sequencing (WGS), as a molecular diagnostic tool, has been greatly developed in TB research and is clinically useful for predicting drug resistance, lineage, tracing transmission, and defining outbreaks [21–26]. Furthermore, WGS has the potential to determine drug resistance much faster than traditional phenotypic drug susceptibility testing (DST) [27] and does not require biological safety infrastructure [28, 29]. These technologies are used differentially, with the highest concentration of sequenced strains found in 16 countries [30–32]. However, for some countries, particularly in South America, WGS information on the variability and distribution of circulating strains is unavailable.

In Latin America, are extensive information on *Mycobacterium tuberculosis*, highlighting its genetic diversity analyzed by molecular methods; *despite* this, very few studies have been carried out in Ecuador, in which the LAM and Haarlem lineages are described as the more prevalent, and a little for the Beijing family [33–35]. Recently Garzon-Chavez et al. [36] determined that the main circulating lineages in Ecuador were LAM, Haarlem, and S and only applied WGS for resolving the classification in 8 isolates with an ambiguous result from 373 strains assessed by 24 loci-MIRU-VNTR and DR analysis, previously. In this sense, it is essential to know the genetic composition of *M. tuberculosis*, especially the variations associated with drug resistance distributed within its genome. This study aims to assess the lineage and molecular resistance profile characterization in clinical isolates of *M. tuberculosis* through whole-genome sequencing.

Results

In order to assess the genetic variability of resistant M. tuberculosis strains, twenty-four culture-confirmed clinical isolates of M. tuberculosis with at least one antibiotic resistance were selected for whole genome sequencing. From them, the bacillary load test in the microscopy smear was greater than two crosses, and 66,7% of patients relapsed. The isolates come from four regions of western Ecuador (20, 83.3%, Guayaquil; 2, 8.3%, El Empalme, 1, 4.2% Babahoyo and Chone, respectively). The majority of the patients were female (64,5%); the mean age was 42.9 ± 12.9 years; comorbidity like HIV was present in three patients (Table 1). All isolates were only tested for phenotypic resistance to four first-line drugs: rifampicin (R), isoniazid (H), ethambutol (E), and pyrazinamide (Z); and one second-line drug: streptomycin (S). Thus, a wide range of drug resistance profiles were covered: 85.7% are polydrug resistant (50% resistant to all first-line drugs plus streptomycin), and 14.3% are MDR (Table 1 and Supplementary table 1).

The 24 strains isolated were sequenced using the Illumina platform with paired-end reads (2×150 bp). The sequences were filtered by the quality and trimmed when necessary. The depth obtained ranged from 22.7X

Table 1 Information on isolates and genomic statistics for analyzed samlples

Clinical isolate	Age	Phenotypic resistance profile	Phenotypic classification according to WHO	Coverage	Contigs	Assembly size (Mb)	CDS	tRNA	Total Mutations
S001_MTb_EC	49	H-R-S	Poly-resistance	35X	144	4,4	4.043	52	1.460
S007_MTb_EC	53	H-R	MDR	33X	152	4,3	4.086	53	1.357
S008_MTb_EC	53	H-R	MDR	41 X	149	4,3	4.054	53	1.378
S013_MTb_EC	34	H-R-S-E-Z	Poly-resistance	34X	159	4,3	4.050	53	1.206
S017_MTb_EC	34	H-R-S	Poly-resistance	33X	141	4,4	4.116	54	1.428
S022_MTb_EC	38	H-R-S-E-Z	Poly-resistance	41X	144	4,4	4.116	54	1.222
S036_MTb_EC	27	H-R-S-Z	Poly-resistance	38X	146	4,3	4.021	52	1.306
S039_MTb_EC	26	H-R-S	Poly-resistance	38X	150	4,3	4.051	53	1.430
S040_MTb_EC	29	H-R-Z	Poly-resistance	33X	162	4,3	4.056	53	1.230
S046_MTb_EC	71	H-R-S	Poly-resistance	38X	150	4,3	4.058	53	1.377
S059_MTb_EC	58	H-R-E	Poly-resistance	47X	190	4,3	4.055	52	1.224
S066_MTb_EC	70	H-R-S-Z	Poly-resistance	38X	150	4,3	4.056	53	853
S070_MTb_EC	36	H-R	MDR	36X	144	4,3	4.097	53	1.209
S091_MTb_EC	40	H-R-S-E-Z	Poly-resistance	54X	137	4,3	4.072	53	1.292
S149_MTb_EC	45	H-R-S-E-Z	Poly-resistance	31X	169	4,3	4.076	58	1.230
S165_MTb_EC	49	H-R-S	Poly-resistance	37X	132	4,3	4.048	52	1.445
S194_MTb_EC	34	H-R-S-E-Z	Poly-resistance	25X	160	4,3	4.060	53	1.182
S196_MTb_EC	28	H-R-S-E-Z	Poly-resistance	29X	138	4,3	4.036	53	1.187
S202_MTb_EC	34	H-R-S-E-Z	Poly-resistance	23X	168	4,3	4.070	53	1.176
S204_MTb_EC	43	H-R-S-E-Z	Poly-resistance	44X	140	4,3	4.094	53	1.326
S205_MTb_EC	45	H-R-S-E-Z	Poly-resistance	26X	159	4,3	4.045	52	1.225

Abbreviations MDR Multidrug resistance, *R* rifampicin, *H* isoniazid, *E* ethambutol, *Z* pyrazinamide, *S* streptomycin, *CDS* Coding sequence, *tRNA* transference RNA

to 53.6X (Table 1). The strains were mapped against the reference *M. tuberculosis* H37Rv (NC_000962.3) (4.4 million base pairs and 4008 protein-coding genes) to verify that they were free of contamination. The sequencing and mapping statistics are presented in Table 1. Most strains have an outstanding quality, depth, and percentage mapping against the reference; however, three strains presented an average depth below 20 and were discarded for further analysis. The 21 remaining strains were *the novo* assembled, and the assemblies obtained were checked for contamination. On average, we recovered draft genomes of 4.3 Mb distributed in 140 contigs per strain with at least an N50 of 56 Kb (Table 1 and Supplementary table 1). Although a complete genome was obtained for all strains, the fragmentation into several contigs is due primarily to regions of extreme base composition and low complexity. This can be seen in Fig. 1, which shows the reference *M. tuberculosis* H37Rv, the percentage of GC and GC-Skew, and the coverage plot of 21 genomes assembled and plotted according to drug resistance profile. Here, common breaks in coverage correspond to peaks in GC or GC-skew.

Prokka and RASTtk were used to determine the structure and function of genes. On average, the Prokka annotation showed the presence of $\sim$4,119 $\pm$ 26.28 genes; among them, 4,065 correspond to coding sequences (CDS), 53 are transfer RNAs, and 1 code for a tmRNAs (Table 1). Based on the annotations, the analysis of the pangenome was carried out with Roary [37]. Thus, the pangenome analysis of 21 genomes of *M. tuberculosis* revealed that there were 4735 gene families defined as the pangenome. The protein-coding genes identified for each strain have slight variation in number, ranging from a minimum of 4021 genes (strain S0036_Mtb_Ec) to a maximum of 4115 genes (strains S0017_Mtb_Ec and S0022_Mtb_Ec). A total of 3750 common genes (core genes) shared by all the genomes analyzed were identified. In addition, 99 soft-core genes (presented in 20 of the 21 strains), 353 accessory genes (shared between two and 19 strains), and 533 cloud genes (strain-specific genes) were identified (Fig. 2A). The core and pangenome size ratio were found to be 0.79; thus, the core forms 79% of the pangenome, representing a rather closed genome with less variability. Gene accumulation was calculated for the core and pangenome. It is observed that after including 21 strains, the increase in the total number of genes is significantly reduced, and on the other hand, the core genome increases minimally, indicating a pangenome close to closing. This can be observed in Figs. 2B and C; from this visualization, it can be suggested that

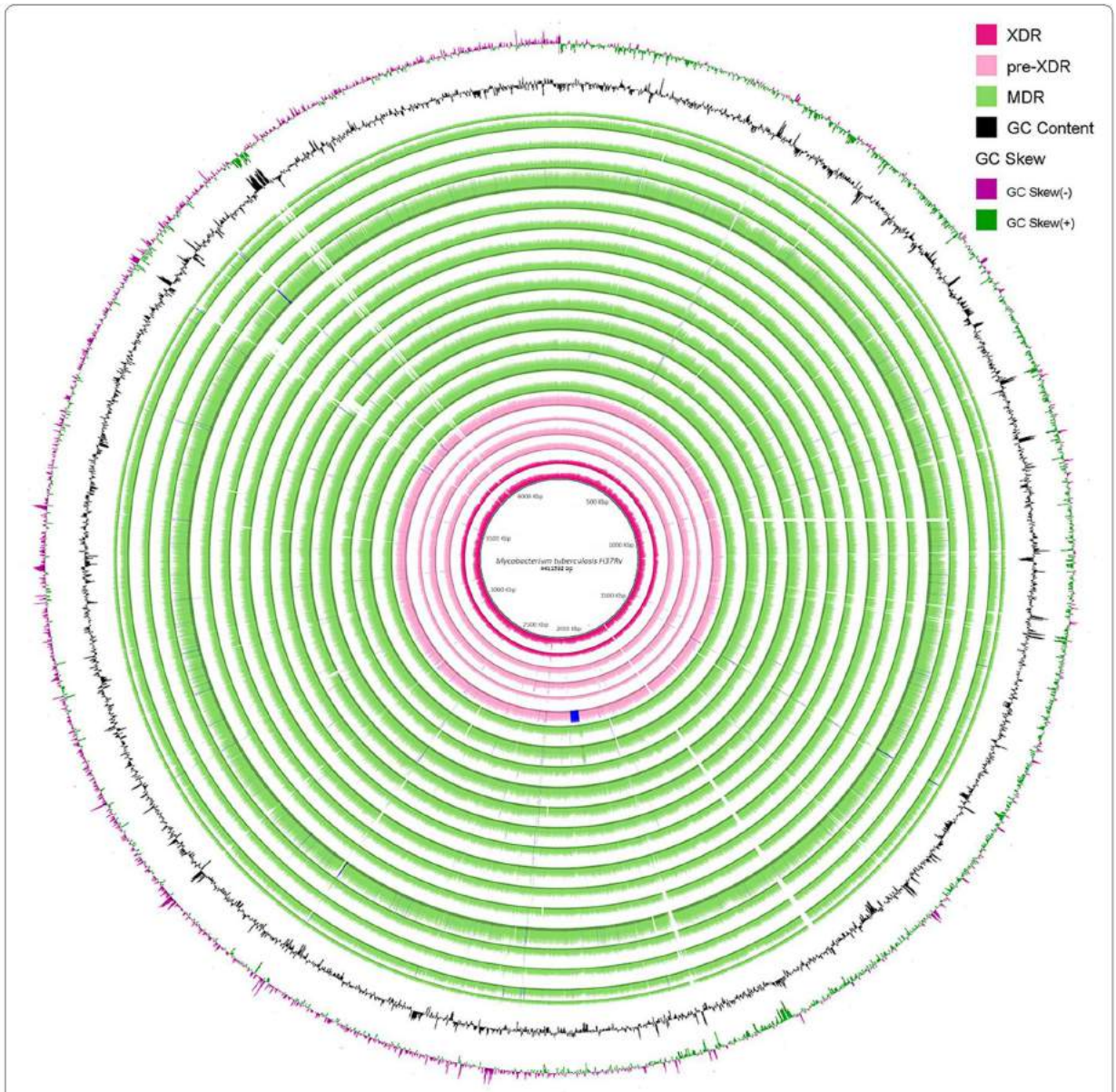

Fig. 1 Genome sequence comparison of 21 resistant *M. tuberculosis* genomes (15 MDR, four pre-XDR, and 2 XDR strains) shown by BRIG. Circles display (from the inside): 1. Reference genome of strain H37Rv; 2–21. Read mapping coverage of sequencing reads calculated from 21 genomes (2, 3. XDR strains (turquoise); 4–7. pre-XDR strains (rose) and 8–21 MDR strains (light green)); 23. GC content; 24. GC skew values positive (green) to negative (purple);. The coverage shown in rings 2 through 21 is from mapping the reads against the reference using BWA. The height of the plot in each ring is proportional to the number of reads mapped at each nucleotide position in the reference genome from 0 to 30 × coverage. The lighter area of the rings represents regions with less coverage. Regions with genome coverage greater than 30 × are shown as solid blue bands

genome characterization of approximately 21 strains provides the genetic repertoire of *M. tuberculosis*.

From the core genes and using Prokka and RAST-tk, subsystems were annotated. Using COG annotations, a function was assigned for a mean of 1,731 genes for each strain analyzed. These annotated genes were classified into 21 metabolic pathways within four primary categories, including Metabolism (49.2%); Information, storage, and processing (17.5%); Cellular processes, and signaling (16.8%), and poorly characterized (16.5%). As for many other organisms, almost half of the genes identified in all strains analyzed (~2022

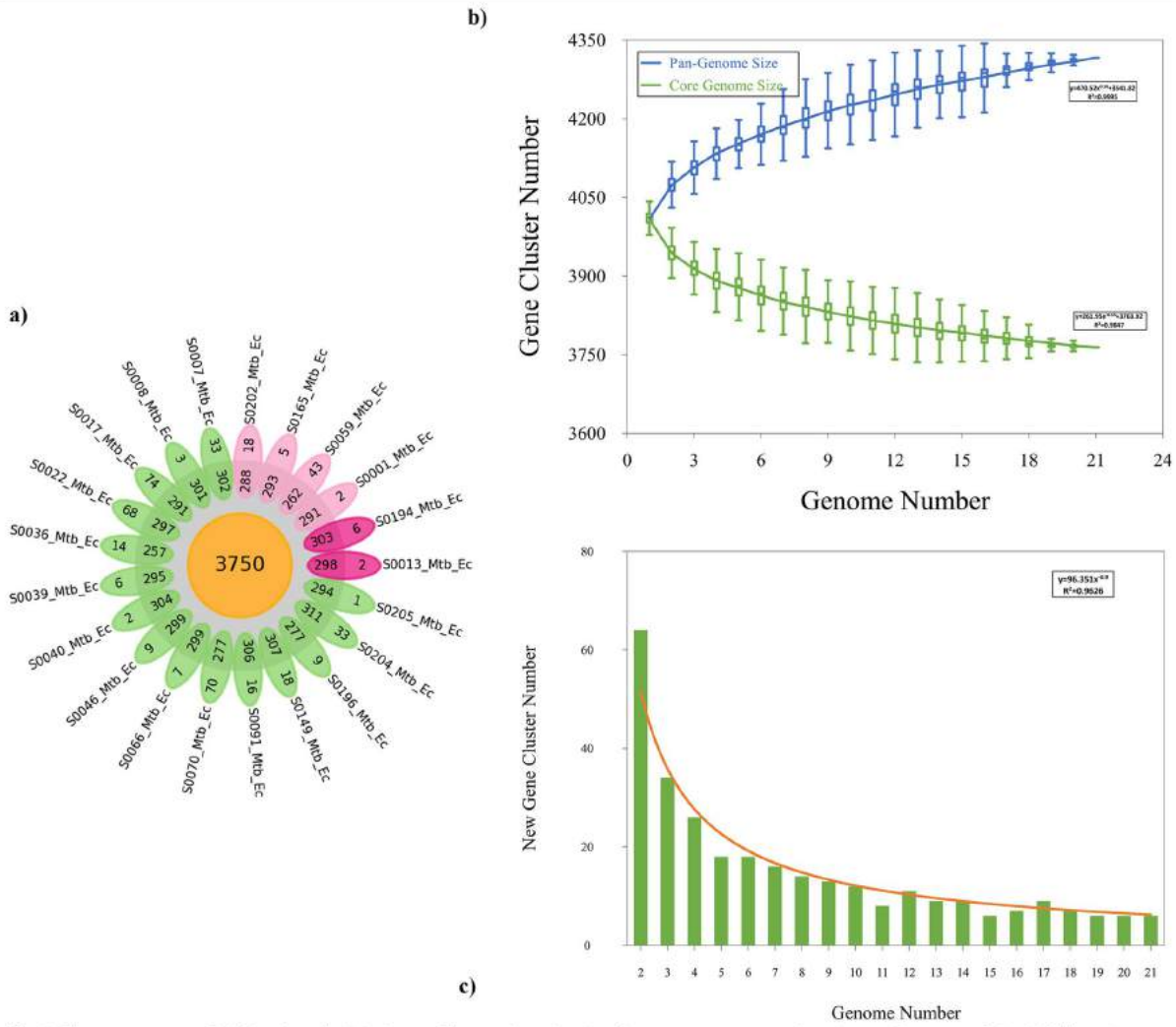

Fig. 2 The pangenome of 21 M. *tuberculosis* isolates. **a** Flower plots showing the core, accessory, and strain-specific genes of the 21 Mtb strains. The flower plot shows the core gene number (in the center), the accessory gene number (in the annulus), and the strain-specific gene number (in the petals) for the 21 Mtb strains. **b** Gene accumulation curves of the pangenome (blue) and the core-genome (green). The blue and green boxes denote the Mtb pangenome and core size for each genomic comparison. **c** The curve (red) shows the number of new genes against the increase in the number of Mtb genomes

genes, ~ 53.9% of the total) could not be assigned function (Supplementary table 2 and Fig. 3a).

In addition, using RASTtk [38], ~1,521 genes were assigned to the subsystems (a group of proteins that implement a specific biological process or structural complex). It was observed that almost 18% of the annotated protein-coding genes were associated with amino acids and their derivatives (267 ± 3.04), followed by genes related to carbohydrates (13.1%, 199 ± 1.74), protein metabolism (10.9%, 166 ± 0.68), cofactors, vitamins, prosthetic groups, and pigments (9.5%, 144 ± 4.65 genes). About 60% of the genes identified in subsystems are involved in the Metabolism of proteins, amino acids, carbohydrates, lipids, and cofactors (Fig. 3b). Cell regulation and signaling (6.6%, 101 ± 1.7) and DNA metabolism (6.1%, 93 ± 1.32) also appear as subsystems with a relevant role. Partly due to the number of strains, a higher diversity in the percentage of genes in the subsystems was found for MDR strains than for pre-XDR and XDR strains (Supplementary table 3). Most of the genes involved in amino acid biosynthesis were conserved and proved to have an essential function for pathogenicity in bacteria, including *M. tuberculosis* [39].

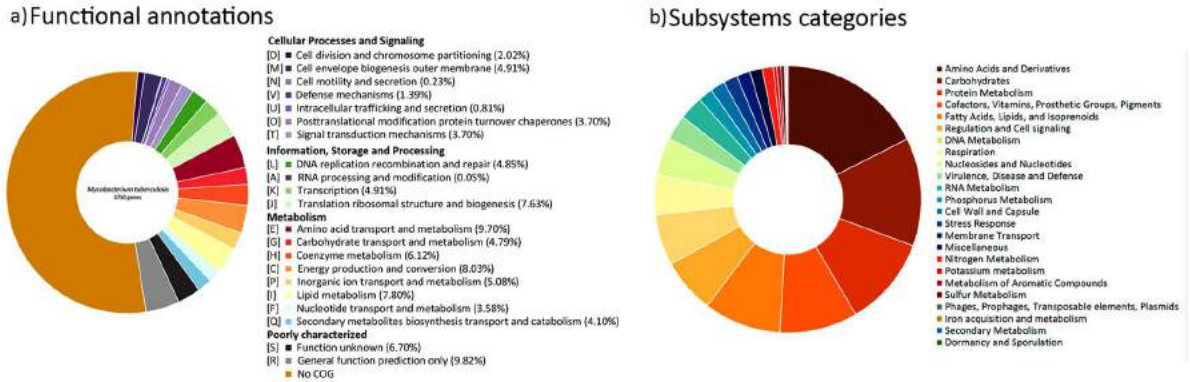

Fig. 3 Functional annotation of 21 M. *tuberculosis* genomes. **a** COG categories and subcategories of predicted genes within the core genomes of 21 M. *tuberculosis* genomes by eggNOG Mapper v2.1.7. Each subcategory is graphed as an average percentage of the total number of genes in the core genomes. **b** Subsystem category distribution of 21 M. *tuberculosis* genomes based on RASTtk annotation. Genes analyzed 4,119 ± 26.28; genes in subsystems 1,521 ± 7.60; subsystems identified 295

Of particular interest are the strain-specific genes that may be responsible for the variability and particularities of this set. Therefore, based on the identification and classification of the genes, we performed a characterization analysis of the functional groups of strain-specific genes. By analyzing the distribution in the 21 strains, we found that most of the identified unique genes involved in metabolism and transport. In particular in the metabolism of lipids, secondary metabolites, coenzymes, amino acids, and inorganic compounds. A smaller proportion of genes are linked to replication, transcription, and translation, and only a few unique genes are related to cellular processes and signaling (Fig. 4).

Since our analysis is based on resistant strains, particular attention was devoted to specific genes, like those genes related to virulence. First, using the Virulence Factor Database (VFDB), a total of 64 virulence-related proteins implicated in the pathogenicity of *M. tuberculosis* were identified, classified as follows: 58.96% related to Secretion system; 11.36% to iron

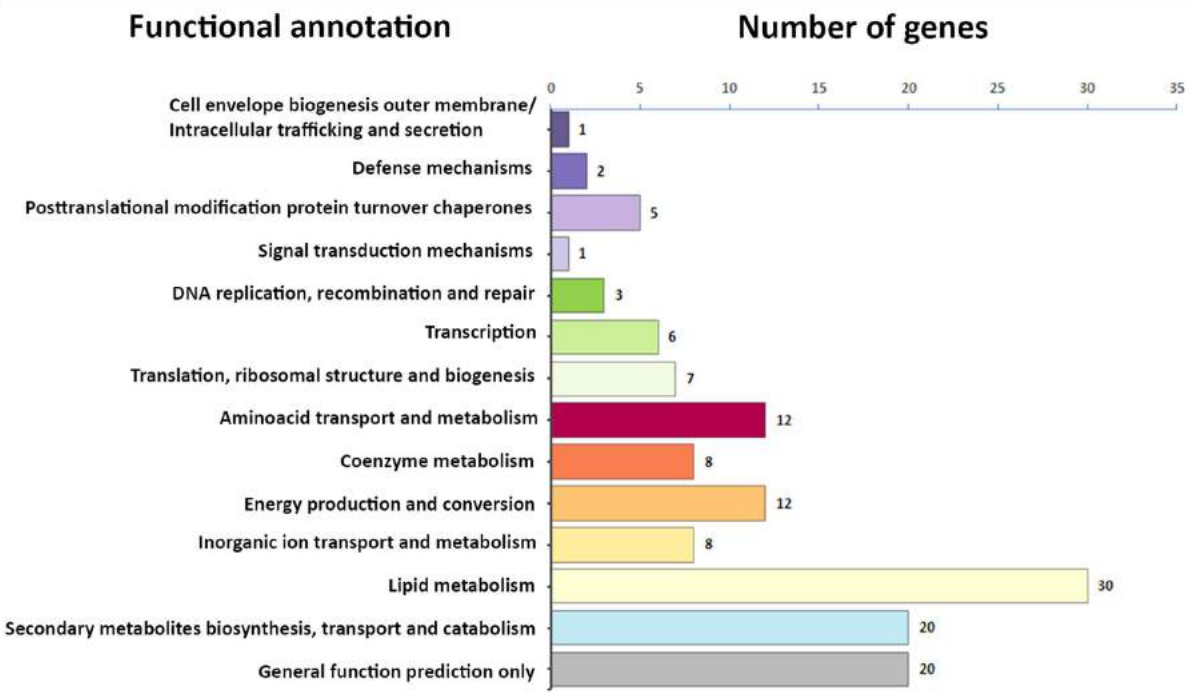

Fig. 4 Functional annotation of the unique genes into 21 M. *tuberculosis* genome by eggNOG Mapper v2.1.7

uptake, Siderophore; 4.26% to Regulation and Cellular Metabolism; 4.12% to the Cell wall and 1.42% to Heat-shock protein and Magnesium uptake (Supplementary table 4). Second, using DrugBank, 66 proteins were identified as potential drug targets (Supplementary table 5). Similarly, 38 families of Transporter proteins were identified by the Transporter Classification Database (TCDB), the most representative being the ATP-binding cassette (abc) superfamily (25.17%), followed by the type vii or esx protein secretion system (t7ss) family (23.65%), the general secretory pathway (sec) family (8.38%), the major facilitator superfamily (mfs) (5.86%) and the resistance-nodulation-cell division (rnd) superfamily (5.03%) (Supplementary table 6). Finally, the presence of CRISPR genes was analyzed using CRISPRfinder, and between 2 and 3 CRISPRs were detected in each genome (Supplementary table 1).

From the genomes, in silico lineage inference was performed using different web tools: TB-Profiler v2.8.13, PhyReSse v1.0, SNP-IT, Mykrobe v0.10.0, CASTB v1.5, TB-Lineage, PhyTB, GenTB and a comparison of their performance was made. Moreover, in silico spoligotyping was determined by KvarQ v0.12. All genomes in this study belonged to Euro-American Lineage (Lineage 4), being the sub-lineages LAM (61.9%, 13/21) and Haarlem (19.0%, 4/21) the most representative, and clades X, S, and T founded in lowest percentage (Supplementary table 7 and Fig. 5). Among all lineages characterized by

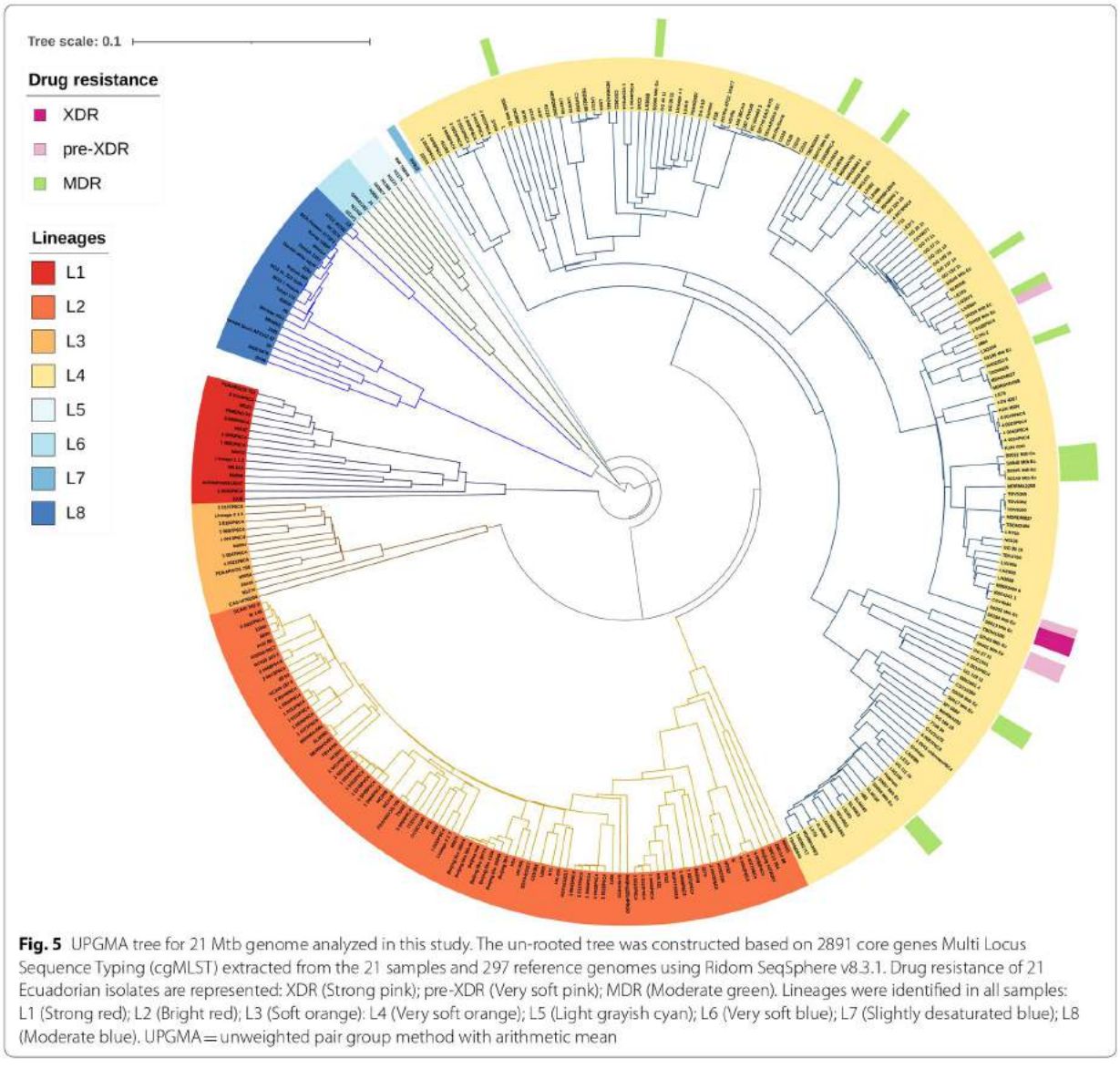

Fig. 5 UPGMA tree for 21 Mtb genome analyzed in this study. The un-rooted tree was constructed based on 2891 core genes Multi Locus Sequence Typing (cgMLST) extracted from the 21 samples and 297 reference genomes using Ridom SeqSphere v8.3.1. Drug resistance of 21 Ecuadorian isolates are represented: XDR (Strong pink); pre-XDR (Very soft pink); MDR (Moderate green). Lineages were identified in all samples: L1 (Strong red); L2 (Bright red); L3 (Soft orange): L4 (Very soft orange); L5 (Light grayish cyan); L6 (Very soft blue); L7 (Slightly desaturated blue); L8 (Moderate blue). UPGMA = unweighted pair group method with arithmetic mean

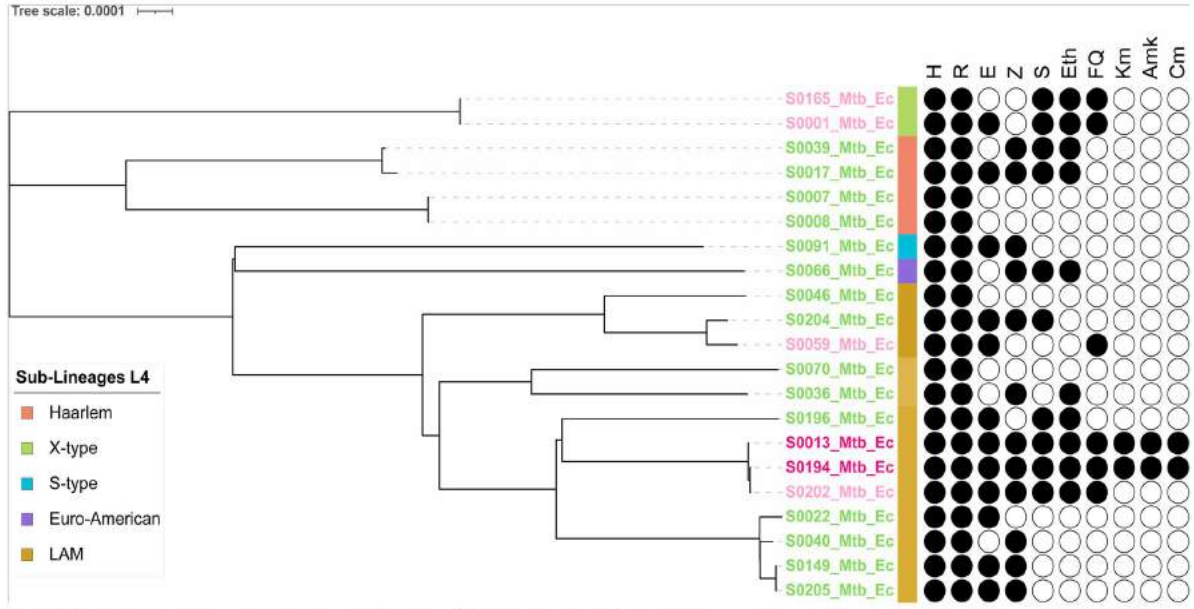

Fig. 6 ML phylogenetic tree showing the relationship of 21 M. *tuberculosis* drug resistance pattern and Lineages. Abbreviations: rifampicin (R), isoniazid (H), pyrazinamide (Z), and ethambutol (E), fluoroquinolone (FQ), streptomycin (S), ethionamide (Eth), and aminoglycosides (Amk, Km, Cm). Color strips represent the sub-lineages. ML: Maximum Likelihood

complete genomic sequencing, isolates classified as MDR accounted for 71.4% (15/21) and resistant (pre-XDR and XDR) for 28.6% (6/21). Notably, most of the latter belong to the LAM sub-lineage, and within the LAM sub-lineage, 30.8% (4/13) are resistant with great epidemiological importance (Fig. 6). TB-profiler, Mykrobe, PhyTB, and GenTB were the most informative web tools for lineages characterization. These tools detailed more genotypes with 100% concordance into *M. tuberculosis* sub-lineages. However, isolates from the T sub-lineage, including genotypes 4.8 to 4.10, could not be effectively resolved. PhyReSse, on the other hand, showed a lower ability to differentiate sub-genotypes within genotypes 4.1.2 and 4.3.4. TB-Lineage, SNP-IT, and CASTB could

Table 2 Frequency and distribution of *Mycobacterium tuberculosis* sub-lineages in 21 Ecuadorian isolates characterized by web-tools

L4 Sub-lineages	Genotype	Web-tools							
		PhyReSse	TB-profiler	Mykrobe	PhyTB	GenTB	TB-LINEAGE	SNP-IT	CASTB
X-type/X3	4.1.1	2	2	2	2	2	2	2	2
Haarlem/H3	4.1.2	4	2	2	2	2	4	4	4
	4.1.2.1	0	2	2	2	2	0	0	0
LAM	4.3.2	3	3	3	3	3	13	13	13
	4.3.3	8	8	8	8	8	0	0	0
	4.3.4	2	0	0	0	0	0	0	0
	4.3.4.1	0	1	1	1	1	0	0	0
	4.3.4.2	0	1	1	1	1	0	0	0
S-type/H1	4.4.1	1	0	0	0	0	1	1	1
	4.4.1.1	0	1	1	1	1	0	0	0
Mainly T	4,8	1	1	0	1	1	0	0	0
	4,10	0	0	1	0	0	0	0	0
Lineage 4	ND	0	0	0	0	0	0	1	0
T1;T2;T3;T4	ND	0	0	0	0	0	1	0	1

Abbreviations L4 Lineage 4, *ND* Not defined

not effectively differentiate sub-lineages with a wide variety of genotypes and subgenotypes, especially the LAM sublineage (Table 2).

To determine drug-resistance single nucleotide polymorphism (SNPs) and small insertions and deletion (indels), we performed a single nucleotide variant calling (SNV) comparing the samples with the H37Rv genome. From the 21 Mtb clinical isolates analyzed in this study, 21,596 high-quality variants were discovered (19,547 SNPs; 1154 insertions and 892 deletions) (Supplementary table 8). Overall, on average 1028 SNPs were found per sample (range 682–1152 SNPs), corresponding to a median SNP density of 1 SNP per 4.2 kb. Few variants were present in all samples, only 1.32% of SNPs ($n=289$) and 0.8% of indels ($n=5$). Most SNPs were found in coding regions (88%, 17,214) and the remaining 2,333 in intergenic regions. Of the SNPs in coding regions, the majority of the variants lead to non-synonymous (NS) changes in amino acids ($n=10{,}685$), including modification of the coding gene (97.5%), loss of start codon (0.1%); gaining of stop codon (1.82%); lost of stop codon (0.6%). The amino acid changes more frequently were *Thr25Ala*, *Ile245Thr*, *Lys212Asn*, *Gly13Asp*, *Ala182Val*, *Asn140Ser*, *Gln20Leu*, *Gly33Ser*, *Thr136Ala*, and *Val259Ala* observed 27, 25, 24 23 and 22 each time, respectively. Among synonymous changes (6529) including 99.35% of synonymous, 0.4% of missense, 0.2% of stop retained variants.

Applying the web-based tools TB-Profiler v2.8.13, PhyReSse v1.0 Mykrobe v0.10.0, CASTB, KvarQ v0.12, PhyTB, GenTB, SAM-TB for genomic analysis of 21 strains phenotypic characterized MDR and polydrug-resistance according to WHO guidelines we identified genotypically 71.43% of strains MDR (15/21), 19.05% pre-XDR (4/21) and 9.52% XDR (2/21). A total of 55 SNVs were identified and distributed into 18 genes known to confer resistance to first- and second-line drugs. Among these, 67.27% (37/55) SNVs were related to resistance to first-line drugs, and genetic regions associated with pyrazinamide resistance presented more frequent variation regions (30.9%, 17/55). Among all genomes, the most frequently genotype identified were *Ser315Thr* (76.2%,16/21), *Ser450Leu* (57.1%,12/21), and *Met306Ile* (33.3%, 7/21), encoding a substitution in *katG*, *rpoB* and *embB* gene that confers resistance to rifampicin, isoniazid, and ethambutol, respectively; whereas for second-line drugs were *Thr202Ala* in *thyA* which is associated with para-aminosalicylic acid resistance in 61.9% (13/21) and *Lys43Arg* in *rpsL* related with streptomycin resistance (28.5%,6/21) (Fig. 7).

We also evaluated the prevalence of resistance in the fluoroquinolone among isolates sequenced by looking at *gyrA* and *gyrB* gene mutation frequencies. In the *gyrA* gene, the SNV *Asp94Asn*, known to confer fluoroquinolone resistance, was found in all pre-XDR isolates, while SNV *Arg90Val* was found in XDR isolate and one pre-XDR. Additionally, the known SNV *Ser95Thr* in the *gyrA* gene was found in 20 isolates. For resistance characterization in tuberculosis, the web-based tools SAM-TB (76.4%) and Mykrobe (72.7%) approached the most frequently identified mutations, complemented by PhyReSse (50.9%), to determine genetic resistance by whole-genome sequencing in surveillance programs (Supplementary table 9 and 10).

Discussion

The present study describes for the first time the use of whole genome sequencing in the screening of TB drug susceptibility in Ecuador and provides information on the phylogenetic characteristics and the identification of the predominant mutations in circulating isolates. Despite its

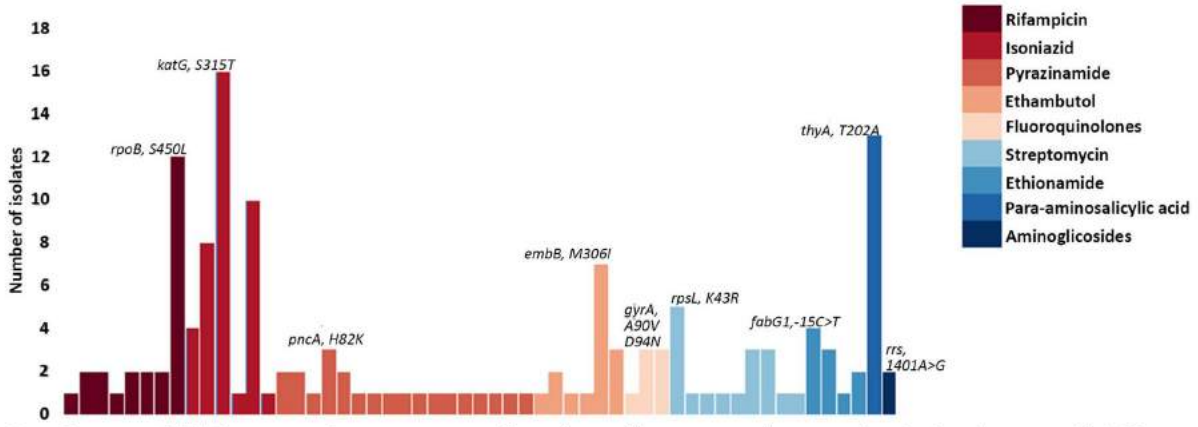

Fig. 7 Frequency of SNVs known to confer resistance against first and second-line treatment drugs in 21 clinical isolates characterized by WGS

limited number of samples including in this study (24 isolates), it represents the initial effort to determine genetic diversity using WGS. Most isolates in our study were from Guayaquil (79.17%), one of the most economically important cities in Ecuador; the high number of tuberculosis cases detected in this city is possibly due to the high mobility from other provinces for trade or work in addition to the location of the leading health centers for the monitoring of this pathogen.

The potential of WGS to provide a quick and comprehensive view of the genotype and reliable prediction of the lineages has been extensively reported [5, 7, 14, 40]; thus, its application in surveillance programs is necessary. The Euro-American lineage (lineage 4) presents a variable distribution in Argentina with a predominance of M and Ra strains [41], for Brazil, lineage 4 varies between 25–99% [13, 42, 43], for Colombia, between 85–99% [44–46], for Mexico between 69–98% [47, 48] and for Peru around 90% [12, 49, 50]. Ecuador is not an exception, in fact, all the samples analyzed by us belong to the Euro-American lineage. This lineage in South and Central America and the Caribbean is represented mainly by the LAM, Haarlem, and T families sub-lineages [11, 41, 51]. In our study, most of the cases analyzed also corresponded to LAM (61.9%), followed by the Haarlem lineage (19.0%). This result provides genomic information from Ecuadorian isolates of *Mycobacterium tuberculosis* that contribute to the knowledge of the distribution of Mtb lineages in South America and confirms the European origin of the circulating strains in Ecuador [9, 52, 53]. Similar results have been described in some, but not all, previous studies using genotyping by 24-loci MIRU-VNTR, Spoligotyping, and SNP-PCR strategies in Ecuador [36, 54].

The movement or conservation of genes in bacteria is a fundamental factor related to virulence, survival, and host adaptability, allowing them to co-evolve together [55]. Several studies in *M. tuberculosis* have shown that the ability of transmission within populations, latency periods, and drug resistance are due to genetic differences within the members of the *M. tuberculosis* complex and that many of its genes have been under positive selection at different periods [56]. Furthermore, the composition within its genome remains relatively stable due to the absence of horizontal gene transfer, and this superficial level of genetic variation is probably because these pathogens have a strongly clonal nature [57]. The highly conserved genetic content and extreme clonal nature suggest that *M. tuberculosis* should have a relatively high percentage of core genome, however, analysis of core and pangenome have shown more variability than expected [58–61]. A pangenome is a union of the entire genetic pool of several strains of a species under comparison, essentially consisting of a core genome containing genes and sequences shared in all strains and an accessory genome consisting of genes and sequences which may be absent from one or more strains and genes that are unique to each strain remaining. In our study, the genetic composition of 21 drug-resistant isolates was analyzed, and 3750 genes (79.2%) were identified in the core genome, considering it a relatively large core genome in line with previous studies in which few genomes were evaluated. Therefore, it is necessary to expand the number of isolates to determine exactly which are the essential genes in *M. tuberculosis* within those circulating in Ecuador.

Characterizing the protein's functionality in microorganisms of importance in public health is essential to understanding their pathogenicity, resistance to antibiotics, and virulence, which allow them to adapt or survive in their host. Our results showed that annotations of protein-encoding genes are mainly associated with metabolisms being Amino acids and their derivatives, Energy production and conversion, and lipid metabolism, the most representative. Similar studies in other *M. tuberculosis* populations also showed a high presence of genes related to energy production and conversion (C), amino acid transport and metabolism (E), and lipid transport and metabolism (I) [72–74]. This suggests that these genes are widely conserved to ensure the interactions of *M. tuberculosis* with its human host, especially in mycobacterial persistence, host pathogens struggle for nutrients and immune recognition. On the other hand, we were able to identify mainly virulence factors associated with the secretion system (58.96%), the most important being the transporter proteins of the ATP-binding cassette (abc) superfamily (25.17%) and the type vii or esx protein secretion system (t7ss) family (23.65%). It has been suggested that high levels of transporters are involved in cell detoxification, nutrient recycling, and antibiotics and drug efflux, significantly affecting the survival and development of multiple drug-resistant strains in Mtb [58, 59, 62, 63]. It should be noted that the unique genes identified 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase and Arylamine N-acetyltransferase were previously characterized as high-confidence drug targets [64–66].

All resistant isolates analyzed showed any variations in genes related to confer resistance, classified them in 71.43% of strains MDR (15/21), 19.05% pre-XDR (4/21), and 9.52% XDR (2/21). Both XDR strains presented similar mutations profile (8 non-synonyms and one deletion SNV in the coding region, 2 SNV in the intergenic region), but differing in additional Ser315Gly mutation in the katG gene arise to isolate S0013_Mtb_Ec, for the four pre-XDR isolates the S0202_Mtb_Ec was that

more mutations were identified (9 genomic regions). It is brought up that of these six isolates of concern, five belong to women within the same age range.

The study of the *M. tuberculosis* resistance is an important factor within its surveillance plan in all countries, for which the timely detection of strains resistant to treatment drugs allows the application of an efficient drug strategy, reducing or limiting the incidence of new cases, especially when applying culture-based methods to detect it. In this study, we found high concordance between WGS and conventional culture-based DST in predicting phenotypic drug resistance to anti-TB drugs; In addition, WGS was able to determine resistance patterns that DST does not evaluate. These findings are in agreement with previous studies [40, 67, 68]. Drug resistance is mediated through mutations in specific gene targets. The key, therefore, is to identify single nucleotide polymorphisms (SNPs) that are responsible for or strongly associated with resistance. In the case of the first-line drug in Mtb treatment, mutations in the 81-bp rifampicin resistance determining region (RRDR) of *the rpoB* gene, also known as a hotspot region, have been accurate predictors of rifampicin resistance in many studies [69–71]. On the other hand, it has been established that INH resistance, predominantly mediated through loss of catalase-peroxidase activity via mutations in katG, produces high-level resistant strains [18, 72, 73]. Finally, in ethambutol, most resistance-related genes are located on *the embB, embC*, and upstream of the embA [74, 75]. Meanwhile, previous studies have agreed that the genes *rrs*, *rpsL*, and *gid* are related to different levels of streptomycin resistance [76, 77], and the chromosomal mutations in the quinolone resistance determining region of *gyrA* or *gyrB* are the main mechanism of resistance to fluoroquinolones in *M. tuberculosis* [78, 79]. In our study, we found 57.1%, 70.0%, 47.6%, and 28.6% and 28.45% of mutations *S450L (rpoB), S315T/G (katG), M306V/I (embB), K43R* (*rpsL*) and c492t, a514c (*rrs* locus) related to rifampicin, isoniazid, ethambutol, streptomycin and fluoroquinolones resistance respectively. The increased resistance to fluoroquinolones, also recently reported in other studies [80], is of concern, probably caused by the increased administration of this drug without prescription [81]. It is therefore imperative to monitor the increase in these resistances in tuberculosis surveillance programs [82, 83].

Whole-genome sequencing (WGS) is a powerful method for detecting drug resistance, genetic diversity, and transmission dynamics of *M. tuberculosis*. Despite their advantages, the analysis of genomic sequencing data remains an obstacle to the routine use of WGS technology in clinical tuberculosis because it requires bioinformatics expertise, high-performance computing, funding, and training that are not readily available in most clinical laboratories [84, 85], which represents a significant challenge for TB control efforts. Over the last few years, many TB-specific genome browsers and WGS analysis tools such as TBProfiler, KvarQ, Mykrobe Predictor TB, CASTB, PhyTB, GenTB, PhyResSE, and others have been developed for genotyping and drug resistance identification, however, despite their use have not been widely used within antimicrobial resistance (AMR) surveillance program on low- and middle-income countries (LMICs) [86]. In our results, web-based tools enabled an effective and user-friendly identification of resistance-associated mutations in *M. tuberculosis*. The recently developed SAM-TB platform [21] is the most capable of determining the mutations present in the isolates of our study, which, when used in conjunction with web-based tools Mykrobe and PhyReSse, would allow rapid screening of *M tuberculosis* isolates. These tools could easily be implemented in surveillance programs based on microbiological procedures to obtain results efficiently and cheaply [84, 87, 88].

Given that the sample size used is small and could be underestimating the real frequencies of mutations in *Mycobacterium tuberculosis* present in this region, it becomes imperative to make greater efforts -and use the advantage of WGS- to analyze a larger number of samples to identify and study the mutations present in patients with tuberculosis including other regions of Ecuador.

Conclusion

The findings of this study demonstrate the usefulness of applying next-generation sequencing tools such as WGS to characterize mutations and describe the existing variability in tuberculosis strains that allow adequate monitoring to generate health policies. We identified the variability of resistant strains circulating among men and women in Ecuador and showed that mainly strains of the American-European lineage 4 circulate, with a higher proportion of the LAM sub-lineage. XDR/MDR strains are not associated with a specific lineage, region or other metacharacter analyzed. However, we found an association between sex and resistance, which should be verified with further sampling. Within the observed resistances, an increase in fluoroquinolone resistance is evident, which should be monitored. Additional sampling is needed to determine the total variability and associations with the metadata obtained to generate better health policies.

Although *M. tuberculosis* is considered a highly conserved clonal species, among the 21 strains analyzed, we found unique genes specialized in lipid metabolism that are particularly interesting for identifying potential new drug targets.

Methods

Twenty-four clinical isolates were sampled and collected from private laboratories between March to December 2020; these isolates had been previously characterized as multidrug-resistant through drug susceptibility testing. The resistance pattern were performed for first-line anti-tuberculosis drugs, isoniazid (H), rifampicin (R), pyrazinamide (Z), ethambutol (E), and one second-line drug (streptomycin (S)), the proportional method by [89] on Lowenstein-Jensen medium and the resistance profile was defined according to WHO recommendations. The critical concentrations of anti-TB drugs evaluated were as follows: rifampicin, 40.0 μg/ml; isoniazid, 0.2 μg/ml; ethambutol, 0.4 μg/ml; streptomycin, 4.0 μg/ml, and 200 μg/mL for pyrazinamide.

DNA extraction and sequencing

Genomic DNA was extracted from 24 isolates of *Mycobacterium tuberculosis* grown in a Lowenstein-Jensen medium using the CTAB method [90]. The quantity of isolated DNA was measured using a Qubit 4.0 fluorometer (Invitrogen, Carlsbad, CA, USA.). DNA samples that fulfilled the integrity, purity, and quantity standards were sequenced. Genomic DNA libraries were prepared for whole genome sequencing using the Tagmentation-based library prep kit according to the manufacturer's instructions (Illumina Inc., San Diego, CA) and sequenced on an Illumina MiniSeq platform with High Output Reagent Kit, producing 150 bp paired-end reads.

Bioinformatic analyses

Using the Galaxy platform (https://usegalaxy.org/), reads were classified by Kraken version 2 [91] to detect any contamination or the presence of other mycobacteria; In addition, FastQC version 0.11.9 [92] and Trimmomatic version 0.38 [93] were used to control the quality and trim the low-quality ends of the reads, respectively. In particular, a sliding window was used to trim sequences with an average quality value lower than 20.

Genome assembly

The final high-quality reads were assembled using Megahit version 1.1.3 [94]. The assembly was carried out with a kmer range from 29 to 141 at intervals of 20 and 800 bp as minimum contig size. To perform contig taxonomic assignment, we used Kaiju version 1.9.0 [95, 96]. For generating metrics and evaluating the quality of the assemblies using *M. tuberculosis* strain HR37v (NC_000962) as a reference genome, Quast version 5.0.2 was used [97].

Gene prediction and functional annotation

The structural and functional annotation was performed with Prokka version 1.12 [98] and the "Rapid Annotation using Subsystem Technology (RASTtk server)" pipeline online [38], and the predicted proteins were also annotated for the identification with the Clusters of Orthologous Groups (COG) by eggNOG Mapper v2.1.7 (http://eggnog-mapper.embl.de/). For the identification of CRISPRs, we used the webserver of CRISPRFinder (https://crispr.i2bc.paris-saclay.fr/Server/). Analysis of virulence-related proteins, potential drug targets, and Transporter proteins by Virulence Factor Database (VFDB), DrugBank, and Transporter Classification Database (TCDB) was performed in the Pathosystems Resource Integration Center (PATRIC) online (https://www.patricbrc.org/).

Variant calling

Processed reads were mapped with Bowtie2 [99] using H37Rv (NCBI ID: NC_000962.3) as a reference genome. Sequence alignment files were sorted and indexed with Samtools v0.1.19 [100]. Bcftools [101] were used for calling variants, and vcftools [102] were used to filter the raw variants (minimum quality score of 30, minimum depth 10). From high-quality variants annotated for 21 genomes, the effect and impact were determined by SnpEff [103].

Pangenome analyses

The pangenome was analyzed with Roary [37] and organized as follows: core, soft-core, shell, and cloud genes if the genes were presented in all, 20, 3—19, and 1—2 samples, respectively; the characteristic curves of the Mtb pangenome, the core-genome, and the new genes were depicted using the pangenome Profile Analyze Tool (PanGP) [58, 104, 105]. Alignments for the core genes Multi Locus Sequence Typing (cgMLST) were used for generating a phylogenetic tree using Ridom SeqSphere v8.3.1 [106]. The output tree was visualized and annotated using the online tool iTOL [107].

Phylogenetic analysis

From high-quality reads the *M. tuberculosis* complex (MTBC) lineages/sublineages were determined and compared using the user-friendly web tools TB-Profiler v2.8.13 (https://tbdr.lshtm.ac.uk/), PhyReSse v1.0 (The Phylo-Resistance-Search-Engine) (https://bioinf.fz-borstel.de/mchips/phyresse/), SNP-IT [108], Mykrobe v0.10.0 (https://www.mykrobe.com/), CASTB v1.5 (The Comprehensive Analysis Server for the *Mycobacterium tuberculosis* complex) (http://castb.ri.ncgm.go.jp/CASTB/), TB-Lineage (https://tbinsight.cs.rpi.edu/run_tb_lineage.html), PhyTB (http://pathogenseq.lshtm.ac.uk/phytblive/index.php), GenTB (https://gentb.hms.harvard.edu/) and in silico spoligotyping were determined by KvarQ v0.12 (https://gap.tbportals.niaid.nih.gov/#/dashboard/home), the different web-programs were working on default parameters in our on-line pipeline.

Predicting susceptibility and drug resistance

Genes related to resistance to rifampicin (R), isoniazid (H), pyrazinamide (Z), and ethambutol (E) [first-line drug], and fluoroquinolone (FQ), streptomycin (S), ethionamide (Eth), aminoglycosides (Amk, Km, Cm), para-aminosalicylic acid (PAS) [second-line drugs] were considered for the analysis. After variant calling and annotation, variants on each resistance gene were determined and compared from the variant call format (VCF) files using the web-based tools TB-Profiler v2.8.13, PhyReSse v1.0 Mykrobe v0.10.0, CASTB, KvarQ v0.12, PhyTB, GenTB, SAM-TB (https://samtb.szmbzx.com/index).

Fundings

This work was funded by grant FCI-016–2017 from University of Guayaquil. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. GM is a doctoral student in the PEDECIBA program. LB is a member of PEDECIBA and the Sistema Nacional de Investigadores (SNI) of ANII.

Abbreviations

TB: Tuberculosis; Mtb: *Mycobacterium tuberculosis*; MTBC: *M. tuberculosis* Complex; R: Rifampicin; H: Isoniazid; Z: Pyrazinamide; E: Ethambutol; FQ: Fluoroquinolone; S: Streptomycin; Eth: Ethionamide; Amk, Km, Cm: Aminoglycosides; PAS: Para-aminosalicylic acid; MDR: Multi-drug resistance; XDR: Extreme-drug resistance; DST: Drug susceptibility testing; SNPs: Single nucleotide polymorphisms; Indels: Insertions, and deletions; WGS: Whole-genome sequencing; CDS: Coding sequences; VFDB: Virulence Factor Database; TCDB: Transporter Classification Database; cgMLST: core-genome Multi Locus Sequence Typing; COG: Clusters of Orthologous Groups; PATRIC: Pathosystems Resource Integration Center.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-022-09042-1.

Additional file 1: Supplementary table 1. Sociodemographic, epidemiological and assembly quality data from 21 isolates of Mycobacterium tuberculosis.

Additional file 2: Supplementary table 2. Distribution of COG annotations on 21 samples of Mycobacterium tuberculosis.

Additional file 3: Supplementary table 3. Distribution of Subsystems on 21 samples of Mycobacterium tuberculosis by Rastk.

Additional file 4: Supplementary table 4. Distribution of Virulence Factor on 21 samples of Mycobacterium tuberculosis.

Additional file 5: Supplementary table 5. Distribution of potential drug target on 21 samples of Mycobacterium tuberculosis.

Additional file 6: Supplementary table 6. Distribution of Transporters on 21 samples of Mycobacterium tuberculosis.

Additional file 7: Supplementary table 7. Characterization of Mycobacterium tuberculosis sub-lineages in 21 Ecuadorian isolates characterized by web-tools and their correlation with phenotypic and genotypic drug resistance profile.

Additional file 8: Supplementary table 8. Genetic variants characterized in 21 Ecuadorian isolates of Mycobacterium tuberculosis.

Additional file 9: Supplementary table 9. Mutations identified in resistance related genes by web-tools on 21 samples of Mycobacterium tuberculosis.

Additional file 10: Supplementary table 10. Distribution of mutations identified in resistance related genes for web-tools on 21 samples of Mycobacterium tuberculosis.

Acknowledgements

We acknowledge the FCI Research program of University of Guayaquil for the partial financial support of the procurement of the reagents to MiniSeq sequencing machine that were used to run the samples. We thank the private laboratories for supporting the isolate analyzed in this study. We also thank the Omics Lab of Universidad Espiritu Santo for permitting the use of the MiniSeq sequencing machine employed to run the samples. We also thank the Bioinformatics Unit of the Institute Pasteur de Montevideo for their assistance and support.

Authors' contributions

GM conceived and designed the study, performed the collection of isolates, sample processing, data analysis (bioinformatic processing of the raw sequencing data) and wrote the main manuscript text, DAM and JCFC performed sample processing (library prep and sequencing), data analysis and review the draft manuscript, LB conceived and designed the study, performed a data analysis, validation, wrote the main manuscript text and review the manuscript; The funding acquisition was performed by GM. All authors read and approved the final manuscript.

Availability of data and materials

The *Mycobacterium tuberculosis* whole-genome sequencing data will be deposited in the public archive of NCBI under the BioProject ID PRJNA827129, BioSamples SAMN29877831—SAMN29877851.

Declarations

Ethics approval and consent to participate

The study was approved by the Ethics Committee of the University San Francisco de Quito (code P2019-188E), certified by the Ministry of Public Health from Ecuador, following guidelines from the Declaration of Helsinki. The samples were processed and anonymized by the clinical laboratory, hence samples are for the researchers in the study de-identified. We obtained a laboratory's permission to use the positive samples.

Consent of publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Laboratorio de Interacciones Hospedero-Patógeno, Unidad de Biología Molecular, Institut Pasteur de Montevideo, Montevideo, Uruguay. [2]Universidad de Guayaquil, Guayaquil, Ecuador. [3]Facultad de Ciencias de la Salud, Universidad Espíritu Santo, Samborondón, Ecuador. [4]Laboratorio de Ciencias Ómicas, Universidad Espíritu Santo, Samborondón, Ecuador. [5]Facultad de Ciencias, Unidad de Genómica Evolutiva, Universidad de La República, Montevideo, Uruguay.

Received: 18 August 2022 Accepted: 23 November 2022
Published online: 21 December 2022

References

1. Factsheet Global TB report 2021 [Internet]. [cited 2022 Jun 18]. Available from: https://www.who.int/publications/m/item/factsheet-global-tb-report-2021.

2. Singh R, Dwivedi SP, Gaharwar US, Meena R, Rajamani P, Prasad T. Recent updates on drug resistance in Mycobacterium tuberculosis. J Appl Microbiol. 2020;128(6):1547–67.
3. Makam P, Matsa R. "Big Three" Infectious Diseases: Tuberculosis, Malaria and HIV/AIDS. Curr Top Med Chem. 2021;21(31):2779–99.
4. Alame Emane AK, Guo X, Takiff HE, Liu S. Drug resistance, fitness and compensatory mutations in Mycobacterium tuberculosis. Tuberculosis. 2021;1(129):102091.
5. Napier G, Campino S, Merid Y, Abebe M, Woldeamanuel Y, Aseffa A, et al. Robust barcoding and identification of Mycobacterium tuberculosis lineages for epidemiological and clinical studies. Genome Med. 2020;12(1):114.
6. Coll F, McNerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, et al. A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. Nat Commun. 2014;5(1):4812.
7. Senghore M, Diarra B, Gehre F, Otu J, Worwui A, Muhammad AK, et al. Evolution of Mycobacterium tuberculosis complex lineages and their role in an emerging threat of multidrug resistant tuberculosis in Bamako, Mali. Sci Rep. 2020;10(1):327.
8. Ngabonziza JCS, Loiseau C, Marceau M, Jouet A, Menardo F, Tzfadia O, et al. A sister lineage of the Mycobacterium tuberculosis complex discovered in the African Great Lakes region. Nat Commun. 2020;11(1):2917.
9. Woodman M, Haeusler IL, Grandjean L. Tuberculosis Genetic Epidemiology: A Latin American Perspective. Genes. 2019;10(1):53.
10. Coscolla M, Gagneux S. Consequences of genomic diversity in Mycobacterium tuberculosis. Semin Immunol. 2014;26(6):431–44.
11. Demay C, Liens B, Burguière T, Hill V, Couvin D, Millet J, et al. SITVITWEB – A publicly available international multimarker database for studying Mycobacterium tuberculosis genetic diversity and molecular epidemiology. Infect Genet Evol. 2012;12(4):755–66.
12. Santos-Lazaro D, Gavilan RG, Solari L, Vigo AN, Puyen ZM. Whole genome analysis of extensively drug resistant Mycobacterium tuberculosis strains in Peru. Sci Rep. 2021;11(1):9493.
13. Tatara MB, Perdigão J, Viveiros M, Kritski A, da Silva KE, Sacchi FPC, et al. Genetic Diversity and Molecular Epidemiology of Mycobacterium tuberculosis in Roraima State. Brazil Am J Trop Med Hyg. 2019;101(4):774–9.
14. Lagos J, Couvin D, Arata L, Tognarelli J, Aguayo C, Leiva T, et al. Analysis of Mycobacterium tuberculosis Genotypic Lineage Distribution in Chile and Neighboring Countries. Manganelli R, editor. PLOS ONE. 2016 Aug 12;11(8):e0160434.
15. World Health Organization. Global tuberculosis report 2015 [Internet]. World Health Organization; 2015 [cited 2022 Jun 20]. 192 p. Available from: https://apps.who.int/iris/handle/10665/191102.
16. Thosani P, Agarwal A, Karade S, Sen S, Katoch CS, Jindamwar P, et al. Assessment of resistance to second line anti-tubercular drugs by line probe assay at a tertiary care hospital. 2022 Jun 19 [cited 2022 Jun 18]; Available from: https://www.mjdrdypv.org/preprintarticle.asp?id=336821;type=0.
17. Iem V, Chittamany P, Suthepmany S, Siphanthong S, Somphavong S, Kontogianni K, et al. Pooling sputum for Xpert MTB/RIF and Xpert Ultra testing during the Covid-19 pandemic in Lao People's Democratic Republic. PLOS Glob Public Health. 2022;2(4):e0000116.
18. Rahman SMM, Ather MdF, Nasrin R, Hoque MA, Khatun R, Rahman T, et al. Performance of WHO-Endorsed Rapid Tests for Detection of Susceptibility to First-Line Drugs in Patients with Pulmonary Tuberculosis in Bangladesh. Diagnostics. 2022;12(2):410.
19. Simieneh A, Tadesse M, Kebede W, Gashaw M, Abebe G. Combination of Xpert® MTB/RIF and DetermineTM TB-LAM Ag improves the diagnosis of extrapulmonary tuberculosis at Jimma University Medical Center, Oromia, Ethiopia. PLoS ONE. 2022;17(2):e0263172.
20. Tantivitayakul P, Ruangchai W, Juthayothin T, Smittipat N, Disratthakit A, Mahasirimongkol S, et al. Homoplastic single nucleotide polymorphisms contributed to phenotypic diversity in Mycobacterium tuberculosis. Sci Rep. 2020;10(1):8024.
21. Yang T, Gan M, Liu Q, Liang W, Tang Q, Luo G, et al. SAM-TB: a whole genome sequencing data analysis website for detection of Mycobacterium tuberculosis drug resistance and transmission. Brief Bioinform. 2022;23(2):bbac030.
22. Lam C, Martinez E, Crighton T, Furlong C, Donnan E, Marais BJ, et al. Value of routine whole genome sequencing for Mycobacterium tuberculosis drug resistance detection. Int J Infect Dis. 2021;113:S48-54.
23. Wang L, Yang J, Chen L, Wang W, Yu F, Xiong H. Whole-genome sequencing of *Mycobacterium tuberculosis* for prediction of drug resistance. Epidemiol Infect. 2022;150:e22.
24. Brown AC. Whole-Genome Sequencing of Mycobacterium tuberculosis Directly from Sputum Samples. In: Parish T, Kumar A, editors. Mycobacteria Protocols [Internet]. New York, NY: Springer US; 2021 [cited 2022 Jun 18]. p. 459–80. (Methods in Molecular Biology). Available from: https://doi.org/10.1007/978-1-0716-1460-0_20.
25. Torres Ortiz A, Coronel J, Vidal JR, Bonilla C, Moore DAJ, Gilman RH, et al. Genomic signatures of pre-resistance in Mycobacterium tuberculosis. Nat Commun. 2021;12(1):7312.
26. Freschi L, Vargas R, Husain A, Kamal SMM, Skrahina A, Tahseen S, et al. Population structure, biogeography and transmissibility of Mycobacterium tuberculosis. Nat Commun. 2021;12(1):6099.
27. Bogaerts B, Delcourt T, Soetaert K, Boarbi S, Ceyssens PJ, Winand R, et al. A Bioinformatics Whole-Genome Sequencing Workflow for Clinical Mycobacterium tuberculosis Complex Isolate Analysis, Validated Using a Reference Collection Extensively Characterized with Conventional Methods and In Silico Approaches. J Clin Microbiol. 59(6):e00202–21.
28. Shea J, Halse TA, Lapierre P, Shudt M, Kohlerschmidt D, Van Roey P, et al. Comprehensive Whole-Genome Sequencing and Reporting of Drug Resistance Profiles on Clinical Cases of Mycobacterium tuberculosis in New York State. J Clin Microbiol. 2017;55(6):1871–82.
29. Coll F, McNerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. Genome Med. 2015;7(1):51.
30. Cohen KA, Manson AL, Abeel T, Desjardins CA, Chapman SB, Hoffner S, et al. Extensive global movement of multidrug-resistant M. tuberculosis strains revealed by whole-genome analysis. Thorax. 2019;74(9):882–9.
31. Gygli SM, Keller PM, Ballif M, Blöchliger N, Hömke R, Reinhard M, et al. Whole-Genome Sequencing for Drug Resistance Profile Prediction in Mycobacterium tuberculosis. Antimicrob Agents Chemother. 2019;63(4):e02175-e2218.
32. The CRyPTIC Consortium and the 100,000 Genomes Project. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. N Engl J Med. 2018 Oct 11;379(15):1403–15.
33. Arias APJ, Lahiguera MJ, Borrás R, Cardona CG, Silva MG, López MJV, et al. Comparative study of the genetic diversity of Mycobacterium tuberculosis Complex by Simplified Amplified Fragment Length Polymorphism and Mycobacterial Interspersed Repetitive Unit Variable Number Tandem Repeat Analysis. Rev Ecuat Med Cienc Biológicas [Internet]. 2018 May 21 [cited 2022 Jul 4];39(1). Available from: https://remcb-puce.edu.ec/remcb/article/view/568.
34. Zurita J, Espinel N, Barba P, Ortega-Paredes D, Zurita-Salinas C, Rojas Y, et al. Genetic diversity and drug resistance of Mycobacterium tuberculosis in Ecuador. Int J Tuberc Lung Dis. 2019;23(2):166–73.
35. Garzon-Chavez D, Zurita J, Mora-Pinargote C, Franco-Sotomayor G, Leon-Benitez M, Granda-Pardo JC, et al. Prevalence, Drug Resistance, and Genotypic Diversity of the Mycobacterium tuberculosis Beijing Family in Ecuador. Microb Drug Resist. 2019;25(6):931–7.
36. Garzon-Chavez D, Garcia-Bereguiain MA, Mora-Pinargote C, Granda-Pardo JC, Leon-Benitez M, Franco-Sotomayor G, et al. Population structure and genetic diversity of Mycobacterium tuberculosis in Ecuador. Sci Rep. 2020;10(1):6237.
37. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691–3.
38. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics. 2008;9(1):75.
39. Smith I. Mycobacterium tuberculosis Pathogenesis and Molecular Determinants of Virulence. Clin Microbiol Rev [Internet]. 2003 [cited 2022 Jul 9]; Available from: https://journals.asm.org/doi/epub/https://doi.org/10.1128/CMR.16.3.463-496.2003.
40. Chatterjee A, Nilgiriwala K, Saranath D, Rodrigues C, Mistry N. Whole genome sequencing of clinical strains of Mycobacterium tuberculosis from Mumbai, India: A potential tool for determining drug-resistance and strain lineage. Tuberculosis. 2017;1(107):63–72.

41. Díaz Acosta CC, Russomando G, Candia N, Ritacco V, Vasconcellos SEG, de Berrêdo Pinho Moreira M, et al. Exploring the "Latin American Mediterranean" family and the RDRio lineage in Mycobacterium tuberculosis isolates from Paraguay, Argentina and Venezuela. BMC Microbiol. 2019 Dec;19(1):131.
42. Hadi SA, Kolte IV, Brenner EP, Cunha E a. T, Simonsen V, Ferrazoli L, et al. Identification of a predominant genotype of Mycobacterium tuberculosis in Brazilian indigenous population. Sci Rep. 2021 Jan 13;11(1):1224.
43. Salvato RS, Reis AJ, Schiefelbein SH, Gómez MAA, Salvato SS, da Silva LV, et al. Genomic-based surveillance reveals high ongoing transmission of multi-drug-resistant Mycobacterium tuberculosis in Southern Brazil. Int J Antimicrob Agents. 2021;58(4):106401.
44. Sánchez-Corrales L, Tovar-Aguirre OL, Galeano-Vanegas NF, Jiménez PAC, Martínez-Vega RA, Maldonado-Londoño CE, et al. Phylogenomic analysis and Mycobacterium tuberculosis antibiotic resistance prediction by whole-genome sequencing from clinical isolates of Caldas, Colombia. PLoS ONE. 2021;16(10):e0258402.
45. Cerezo I, Jiménez Y, Hernandez J, Zozio T, Murcia MI, Rastogi N. A first insight on the population structure of Mycobacterium tuberculosis complex as studied by spoligotyping and MIRU-VNTRs in Bogotá, Colombia. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. 2012;12(4):657–63.
46. Ferro BE, Nieto LM, Rozo JC, Forero L, van Soolingen D. Multidrug-Resistant Mycobacterium tuberculosis. Southwestern Colombia Emerg Infect Dis. 2011;17(7):1259–62.
47. Munro-Rojas D, Fernandez-Morales E, Zarrabal-Meza J, Martínez-Cazares MT, Parissi-Crivelli A, Fuentes-Domínguez J, et al. Genetic diversity of drug and multidrug-resistant Mycobacterium tuberculosis circulating in Veracruz, Mexico. PLoS ONE. 2018;13(3):e0193626.
48. Lefort B, Pérez-Martínez D, Viveros D, Zenteno-Cuevas R, Lefort B, Pérez-Martínez D, et al. Diversidad genética de Mycobacterium tuberculosis circulando en la jurisdicción V de Jalapa, Veracruz. México Rev Chil Infectol. 2021;38(5):639–46.
49. Sheen P, Couvin D, Grandjean L, Zimic M, Dominguez M, Luna G, et al. Genetic Diversity of Mycobacterium tuberculosis in Peru and Exploration of Phylogenetic Associations with Drug Resistance. PLoS ONE. 2013;8(6):e65873.
50. Taype CA, Agapito JC, Accinelli RA, Espinoza JR, Godreuil S, Goodman SJ, et al. Genetic diversity, population structure and drug resistance of Mycobacterium tuberculosis in Peru. Infect Genet Evol. 2012;12(3):577–85.
51. Realpe T, Correa N, Rozo JC, Ferro BE, Gomez V, Zapata E, et al. Population Structure among Mycobacterium tuberculosis Isolates from Pulmonary Tuberculosis Patients in Colombia. PLoS ONE. 2014;9(4):e93848.
52. Wiens KE, Woyczynski LP, Ledesma JR, Ross JM, Zenteno-Cuevas R, Goodridge A, et al. Global variation in bacterial strains that cause tuberculosis disease: a systematic review and meta-analysis. BMC Med. 2018;30(16):196.
53. Brynildsrud OB, Pepperell CS, Suffys P, Grandjean L, Monteserin J, Debech N, et al. Global expansion of Mycobacterium tuberculosis lineage 4 shaped by colonial migration and local adaptation. Sci Adv. 2018;4(10):eaat5869.
54. Jiménez P, Calvopiña K, Herrera D, Rojas C, Pérez L, Grijalva M, et al. Identification of the Mycobacterium tuberculosis Beijing lineage in Ecuador. Biomédica [Internet]. 2017 Jun 1 [cited 2022 Jun 18];37(2). Available from: http://www.revistabiomedica.org/index.php/biomedica/article/view/3450.
55. Brites D, Gagneux S. Co-evolution of Mycobacterium tuberculosis and Homo sapiens. Immunol Rev. 2015;264(1):6–24.
56. Chiner-Oms Á, López MG, Moreno-Molina M, Furió V, Comas I. Gene evolutionary trajectories in *Mycobacterium tuberculosis* reveal temporal signs of selection. Proc Natl Acad Sci. 2022;119(17):e2113600119.
57. Moreno-Molina M, Shubladze N, Khurtsilava I, Avaliani Z, Bablishvili N, Torres-Puente M, et al. Genomic analyses of Mycobacterium tuberculosis from human lung resections reveal a high frequency of polyclonal infections. Nat Commun. 2021;12(1):2716.
58. Yang T, Zhong J, Zhang J, Li C, Yu X, Xiao J, et al. Pan-Genomic Study of Mycobacterium tuberculosis Reflecting the Primary/Secondary Genes, Generality/Individuality, and the Interconversion Through Copy Number Variations. Front Microbiol [Internet]. 2018 [cited 2022 Jun 20];9. Available from: https://www.frontiersin.org/article/https://doi.org/10.3389/fmicb.2018.01886
59. Zakham F, Sironen T, Vapalahti O, Kant R. Pan and Core Genome Analysis of 183 Mycobacterium tuberculosis Strains Revealed a High Inter-Species Diversity among the Human Adapted Strains. Antibiotics. 2021;10(5):500.
60. Dar HA, Zaheer T, Ullah N, Bakhtiar SM, Zhang T, Yasir M, et al. Pangenome Analysis of Mycobacterium tuberculosis Reveals Core-Drug Targets and Screening of Promising Lead Compounds for Drug Discovery. Antibiotics. 2020;9(11):819.
61. Periwal V, Patowary A, Vellarikkal SK, Gupta A, Singh M, Mittal A, et al. Comparative Whole-Genome Analysis of Clinical Isolates Reveals Characteristic Architecture of Mycobacterium tuberculosis Pangenome. Hoshino Y, editor. PLOS ONE. 2015 Apr 8;10(4):e0122979.
62. Cassio Barreto de Oliveira M, Balan A. The ATP-Binding Cassette (ABC) Transport Systems in Mycobacterium tuberculosis: Structure, Function, and Possible Targets for Therapeutics. Biology. 2020;9(12):443.
63. Yang Z, Zeng X, Tsui SKW. Investigating function roles of hypothetical proteins encoded by the Mycobacterium tuberculosis H37Rv genome. BMC Genomics. 2019;20(1):394.
64. Diamanti E, Hamed MM, Lacour A, Bravo P, Illarionov B, Fischer M, et al. Targeting the IspD Enzyme in the MEP Pathway: Identification of a Novel Fragment Class. ChemMedChem. 2022;17(5):e202100679.
65. Wang X, Dowd CS. The MEP pathway: Promising drug targets in the fight against tuberculosis. ACS Infect Dis. 2018;4(3):278–90.
66. Mi J, Gong W, Wu X. Advances in Key Drug Target Identification and New Drug Development for Tuberculosis. BioMed Res Int. 2022;25(2022):5099312.
67. Quan TP, Bawa Z, Foster D, Walker T, del Ojo Elias C, Rathod P, et al. Evaluation of Whole-Genome Sequencing for Mycobacterial Species Identification and Drug Susceptibility Testing in a Clinical Setting: a Large-Scale Prospective Assessment of Performance against Line Probe Assays and Phenotyping. Diekema DJ, editor. J Clin Microbiol. 2018 Feb;56(2):e01480–17.
68. Katale BZ, Mbelele PM, Lema NA, Campino S, Mshana SE, Rweyemamu MM, et al. Whole genome sequencing of Mycobacterium tuberculosis isolates and clinical outcomes of patients treated for multidrug-resistant tuberculosis in Tanzania. BMC Genomics. 2020;21(1):174.
69. Guo Y, Cao X, Yang J, Wu X, Liu Y, Wan B, et al. Rifampin-resistance-associated mutations in the rifampin-resistance-determining region of the rpoB gene of Mycobacterium tuberculosis clinical isolates in Shanghai, PR China. J Med Microbiol. 2021 Mar;70(3).
70. Kardan-Yamchi J, Amini S, Hamzelou G, Rahimi Foroushani A, Ghodousi A, Cirillo DM, et al. Assessment of the GenoType MTBDRsl VER 2.0 compared to the phenotypic drug susceptibility testing and whole genome sequencing for the rapid detection of resistance to fluoroquinolone and second-line injectable drugs among rifampicin-resistant Mycobacterium tuberculosis isolates. Arch Microbiol. 2021 Sep 1;203(7):3989–96.
71. Cox H, Goig GA, Salaam-Dreyer Z, Dippenaar A, Reuter A, Mohr-Holland E, et al. Whole-Genome Sequencing Has the Potential To Improve Treatment for Rifampicin-Resistant Tuberculosis in High-Burden Settings: a Retrospective Cohort Study. J Clin Microbiol. 2022;60(3):e02362-e2421.
72. Pinhata JMW, Brandao AP, Mendes F de F, Rabello MC da S, Ferrazoli L, de Oliveira RS. Correlating genetic mutations with isoniazid phenotypic levels of resistance in Mycobacterium tuberculosis isolates from patients with drug-resistant tuberculosis in a high burden setting. Eur J Clin Microbiol Infect Dis. 2021 Dec 1;40(12):2551–61.
73. Bakhtiyariniya P, Khosravi AD, Hashemzadeh M, Savari M. Detection and characterization of mutations in genes related to isoniazid resistance in Mycobacterium tuberculosis clinical isolates from Iran. Mol Biol Rep [Internet]. 2022 Apr 2 [cited 2022 Jun 18]; Available from: https://link.springer.com/https://doi.org/10.1007/s11033-022-07404-2.
74. Xu Y, Jia H, Huang H, Sun Z, Zhang Z. Mutations Found in embCAB, embR, and ubiA Genes of Ethambutol-Sensitive and -Resistant Mycobacterium tuberculosis Clinical Isolates from China. BioMed Res Int. 2015;2015:951706.
75. Sun Q, Xiao T yang, Liu H can, Zhao X qin, Liu Z guang, Li Y nan, et al. Mutations within embCAB Are Associated with Variable Level of Ethambutol Resistance in Mycobacterium tuberculosis Isolates from China. Antimicrob Agents Chemother. 2017 Dec 21;62(1):e01279–17.

76. Rocha DMGC, Magalhães C, Cá B, Ramos A, Carvalho T, Comas I, et al. Heterogeneous Streptomycin Resistance Level Among Mycobacterium tuberculosis Strains From the Same Transmission Cluster. Front Microbiol [Internet]. 2021 [cited 2022 Jun 18];12. Available from: https://www.frontiersin.org/article/https://doi.org/10.3389/fmicb.2021.659545.
77. Cohen KA, Stott KE, Munsamy V, Manson AL, Earl AM, Pym AS. Evidence for Expanding the Role of Streptomycin in the Management of Drug-Resistant Mycobacterium tuberculosis. Antimicrob Agents Chemother. 2020;64(9):e00860-e920.
78. Chaiyachat P, Chaiprasert A, Nonghanphithak D, Smithtikarn S, Kamolwat P, Pungrassami P, et al. Whole-genome analysis of drug-resistant Mycobacterium tuberculosis reveals novel mutations associated with fluoroquinolone resistance. Int J Antimicrob Agents. 2021;58(3):106385.
79. Kabir S, Tahir Z, Mukhtar N, Sohail M, Saqalein M, Rehman A. Fluoroquinolone resistance and mutational profile of gyrA in pulmonary MDR tuberculosis patients. BMC Pulm Med. 2020;11(20):138.
80. Mirzayev F, Viney K, Linh NN, Gonzalez-Angulo L, Gegia M, Jaramillo E, et al. World Health Organization recommendations on the treatment of drug-resistant tuberculosis, 2020 update. Eur Respir J. 2021;57(6):2003300.
81. Aydın Kayalı R, Özkan SA, Biçmen C, Erer OF. The Relation Between the Emergence of Fluoroquinolone Resistance and Fluoroquinolone Exposure in New Cases of Active Pulmonary Tuberculosis. Turk Thorac J. 2021;22(1):45–9.
82. Dalhoff A. Global Fluoroquinolone Resistance Epidemiology and Implictions for Clinical Use. Interdiscip Perspect Infect Dis. 2012;2012:1–37.
83. Xia H, Zheng Y, Liu D, Wang S, He W, Zhao B, Increase S, in Moxifloxacin Resistance Rate among Multidrug-Resistant Mycobacterium tuberculosis Isolates in China, et al. to 2013. Microbiol Spectr. 2007;9(3):e00409-e421.
84. McNerney R, Zignol M, Clark TG. Use of whole genome sequencing in surveillance of drug resistant tuberculosis. Expert Rev Anti Infect Ther. 2018;16(5):433–42.
85. Davedow T, Carleton H, Kubota K, Palm D, Schroeder M, Gerner-Smidt P, et al. PulseNet International Survey on the Implementation of Whole Genome Sequencing in Low and Middle-Income Countries for Foodborne Disease Surveillance. Foodborne Pathog Dis. 2022;19(5):332–40.
86. Waddington C, Carey ME, Boinett CJ, Higginson E, Veeraraghavan B, Baker S. Exploiting genomics to mitigate the public health impact of antimicrobial resistance. Genome Med. 2022;14(1):15.
87. Sanchini A, Jandrasits C, Tembrockhaus J, Kohl TA, Utpatel C, Maurer FP, et al. Improving tuberculosis surveillance by detecting international transmission using publicly available whole genome sequencing data. Eurosurveillance. 2021;26(2):1900677.
88. van der Werf MJ, Ködmön C. Whole-Genome Sequencing as Tool for Investigating International Tuberculosis Outbreaks: A Systematic Review. Front Public Health [Internet]. 2019 [cited 2022 Aug 4];7. Available from: https://www.frontiersin.org/articles/https://doi.org/10.3389/fpubh.2019.00087
89. Canetti G, Rist N, Grosset J. [Measurement of sensitivity of the tuberculous bacillus to antibacillary drugs by the method of proportions. Methodology, resistance criteria, results and interpretation]. Rev Tuberc Pneumol (Paris). 1963 Mar;27:217–72.
90. Van Soolingen D. Molecular epidemiology of tuberculosis and other mycobacterial infections: main methodologies and achievements. J Intern Med. 2001;249(1):1–26.
91. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15(3):R46.
92. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. 2017.
93. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
94. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31(10):1674–6.
95. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun. 2016;7(1):11257.
96. Kaiju: Fast and sensitive taxonomic classification for metagenomics [Internet]. [cited 2022 Jun 20]. Available from: https://kaiju.binf.ku.dk/.
97. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072–5.
98. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–9.
99. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
100. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
101. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021 Feb 16;10(2):giab008.
102. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–8.
103. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w 1118; iso-2; iso-3. Fly (Austin). 2012;6(2):80–92.
104. Zhao Y, Jia X, Yang J, Ling Y, Zhang Z, Yu J, et al. PanGP: A tool for quickly analyzing bacterial pangenome profile. Bioinformatics. 2014;30(9):1297–9.
105. PanGP: pangenome Profile Tool Website [Internet]. [cited 2022 Jun 20]. Available from: https://pangp.zhaopage.com/.
106. Jünemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinowski J, et al. Updating benchtop sequencing performance comparison. Nat Biotechnol. 2013;31(4):294–6.
107. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016 Jul 8;44(Web Server issue):W242–5.
108. Lipworth S, Jajou R, de Neeling A, Bradley P, van der Hoek W, Maphalala G, et al. SNP-IT Tool for Identifying Subspecies and Associated Lineages of Mycobacterium tuberculosis Complex. Emerg Infect Dis. 2019;25(3):482–8.

Publisher's Note

Artículo 2: Transmission network and genomic diversity of *Mycobacterium tuberculosis* L4 lineage in Ecuadorian isolates provided clues to transoceanic epidemiology

En este artículo se analizan las posibles rutas de transmisión de los sublinajes de *M. tuberculosis* que circulan en Ecuador. Para ello, se comparan las relaciones filogenéticas y las variaciones genéticas identificadas en 88 aislados ecuatorianos (64 resistentes y 24 sensibles). También se investiga la similitud genética con 415 genomas previamente publicados, provenientes de 20 países. Es importante destacar que todos los genomas analizados pertenecen al linaje Euroamericano, que es el más prevalente en Ecuador. Nuestros datos revelan que nuestros aislados se dividen en quince grupos de transmisión.Sin embargo , cuatro aislados no se agrupan dentro de ninguno de los 51 grupos globales presentes en todos los genomas analizados. Además, se observó que 17 aislados muestran clonalidad con aislados identificados en migrantes latinoamericanos entre 2014 y 2015, incluyendo ocho ecuatorianos que migraron a España hace casi 20 años.También se encontraron relaciones genéticas entre nuestros aislados y aislados de otros países que son frecuentemente seleccionados para migración, conformando ocho redes de transmisión. La detección de clones en aislados ecuatorianos con los provenientes de migrantes sugiere la posible transmisión de tuberculosis latente. Esto podría estar relacionado con visitas de migrantes desde o hacia Ecuador. Estos hallazgos destacan la necesidad de mejorar los sistemas de vigilancia para identificar oportunamente brotes asociados con el movimiento de personas. Además, subraya la relevancia de utilizar la secuenciación de genoma completo para un seguimiento más preciso en estos casos.

Transmission network and genomic diversity of *Mycobacterium tuberculosis* L4 lineage in Ecuadorian isolates provided clues to transoceanic epidemiology.

1.1 Author names

Gabriel Morey-León[a,c,#], https://orcid.org/0000-0003-1824-8285

Paulina M. Mejía-Ponce[b], https://orcid.org/0000-0003-0673-020X

Juan Carlos Fernández-Cadena[a], https://orcid.org/0000-0002-0398-8371

Evelyn García-Moreira[c], https://orcid.org/0000-0002-2181-3364

Derly Andrade-Molina[a], https://orcid.org/0000-0002-2651-5884

Cuauhtémoc Licona-Cassani[b], https://orcid.org/0000-0002-0360-3945

Pablo Fresia[f], https://orcid.org/0000-0002-8465-476X

Luisa Berná[g,h,#] ,https://orcid.org/0000-0002-9358-3414

1.2 Affiliation(s)

[a.] Facultad de Ciencias de la Salud. Universidad Espíritu Santo, Samborondón, Ecuador.

[b.] Escuela de Ingeniería y Ciencias, Tecnológico de Monterrey, Nuevo León, México

[c.] Universidad de la República. Montevideo, Uruguay

[d.] Laboratorio de Ciencias Ómicas. Universidad Espíritu Santo, Samborondón, Ecuador.

[e.] Instituto Superior Tecnológico Argos, Guayaquil, Ecuador

[f.] Unidad Mixta Pasteur + INIA, Institut Pasteur de Montevideo, Montevideo, Uruguay.

[g.] Laboratorio de Interacciones Hospedero-Patógeno, Unidad de Biología Molecular, Institut Pasteur de Montevideo, Montevideo, Uruguay.

[h.] Unidad de Genómica Evolutiva, Facultad de Ciencias, Universidad de la República, Uruguay.

1.3 Corresponding author and email address

[#] Corresponding author. Luisa Berná email: lberna@pasteur.edu.uy, Gabriel Morey-León email: garielmorey@uees.edu.ec

1.4 Keywords

Genomic epidemiology, genomic clusters, tuberculosis, TMRCA, Ecuador, surveillance.

2. Abstract

Tuberculosis (TB) is a global public health concern and understanding *Mycobacterium tuberculosis* (*Mtb*) transmission routes and genetic diversity is crucial for outbreak control. This study aimed to investigate the genomic epidemiology and genetic diversity of *Mtb* in Ecuador by comparing 88 local isolates with 415 public genomes from 19 countries in the Euro-American lineage (L4). Our analysis identified 51 transmission groups among all isolates, with 19.3% of Ecuadorian isolates exhibiting clonality with isolates from Colombia and Latin American migrants, including eight Ecuadorians who migrated to Spain two decades ago. The isolates from Ecuador had variations in virulence genes crucial for intracellular replication and trafficking during acute infection and latency. Our findings suggest the possible transmission of reactivated latent TB associated with nucleotide variation, highlighting the need for strengthened surveillance systems to detect outbreaks related to human movement. Whole Genome Sequencing (WGS) is essential for follow-up investigations due to the complex epidemiology of TB, including variable latency periods and within-host heterogeneity.

3. Impact statement

Tuberculosis is a pressing global health concern, primarily attributed to increased multidrug (MDR) and extreme drug (XDR) resistance cases. Despite the availability of efficacious drug regimens, the efficacy of treatment has been compromised by patient non-adherence due to patient-related, medication-related, health service-related, and sociocultural factors. Among these prolonged treatment periods, effective antimicrobials, and delayed diagnosis testing, raise the emergence of drug-resistant strains. The escalating of resistant cases, exacerbated by the disruptive impact of the COVID-19 pandemic, and intensified migration patterns have highlighted the importance of implementing efficient and timely diagnostic methodologies worldwide, remarkably to follow the outbreak. This study identified Ecuadorian isolates with nucleotide changes in genes associated with drug resistance or virulence factors and defined transmission networks between individuals and across borders. Furthermore, our results show that personal relationships among individuals in shared work environments may contribute to transmission between them. These suggest possible transnational transmission events involving Ecuador and its border countries, frequently migration destinations. This reinforces the need to adopt a more extensive approach based on a complete genome sequence analysis to improve TB strain surveillance programs locally. The integration of these advanced genomic technologies in a local context is indispensable, for improving the outcome of TB strain monitoring efforts.

4. Data summary

Raw read sequences analyzed in this study can be found in the SRA database (http://www.ncbi.nlm.nih.gov/sra) under the following BioProject accession numbers: PRJNA827129, PRJEB23245, PRJEB23681, PRJEB27366, PRJEB29069, PRJEB44165, PRJEB48543, PRJEB50999,

PRJEB7669, PRJNA37301, PRJNA599957, PRJNA628024, PRJNA227755, PRJNA227756, PRJNA755956, PRJNA272873, PRJNA824124, PRJNA707145, PRJNA870648, PRJNA270004, PRJNA422870, and PRJNA438689. Refer to Table S1-2 for SRA accession numbers. Protocols have been provided within the article. Eight supplementary tables and five supplementary figures are available with the online version of this article.

5. Introduction

Tuberculosis (TB) caused by *Mycobacterium tuberculosis* (*Mtb*) is a significant global health concern, with an estimated of 10.6 million infections and 1.6 million casualties in 2021(1). The COVID-19 pandemic disrupted access to medical healthcare, including TB diagnosis and treatment programs, aggravating the TB burden and compromising the progress in TB control achieved in the last decades(2,3). Furthermore, the incidence of TB has been increasing in several countries, including Ecuador, where the rate reached 48 cases per 100,000 inhabitants in 2021(4) due, among other factors, to COVID-19 containment measures(5,6).

In recent years, the global increase in human migration has contributed to the spread of TB, mainly in high-income countries, where migrants are more likely to immigrate, whether in search of economic or educational opportunities or better living conditions. In the Ecuadorian context, the US, Spain, Italy, Canada, and Chile have been the countries most preferred(7–9). Identifying the transmission routes of TB cases is essential to reducing the potential transmission networks. Different public health organizations have implemented screening programs among household contacts to control the spread of the disease, including the TB-Directly Observed Treatment Short Course (TB-DOTS) (10). Genomic approaches, such as Whole Genome Sequencing (WGS) of *Mtb* strains, have improved the tracing of transmitted TB cases and provided invaluable information on drug resistance and sub-lineage patterns(11–16). In addition, studies combining WGS and Bayesian phylogenetic approaches have reconstructed the historical patterns of TB spread in Central and South America, dating the introduction of *Mtb* strains in those regions. The apparent ancestral emergence and marked diversification of the L2 (ancestral and modern Beijing sublineages) and L4 lineages have also been emphasized, reflecting their close correlation with restricted geographic distribution, leading to the independent emergence of multiple sub-lineages and local adaptation to distinct human populations(17–20). Studies conducted in Brazil, Paraguay, Mexico, and the United States have applied genomic approaches to identify frequent transmission clusters between prison inmates, drug users, migrants, and mixed groups (18-21), highlighting the need to prioritize contact tracing to groups with a higher likelihood of clustering retrospectively to improve TB control.

The *Mtb* complex (MTBC) comprises distinct phylogenetic lineages that have evolved over centuries (21–23). The Euro-American lineage (L4) exhibits a wide distribution of sub-lineages between and within countries. The Latin-American (LAM), Haarlem, X-type, and T families are the most prevalent L4 sub-lineages in Central and South America, the Caribbean, Europe, and Middle Africa (24–29). In Ecuador,

previous reports have revealed that the LAM and Haarlem families are the most prevalent, with a few cases of the Beijing family(30–33), where transmission clustering has been linked to isolates from neighboring countries(34,35). However, the genetic diversity of *Mtb* in local isolates from Ecuador has been understudied. A unique WGS study has identified the presence of 4.3.2/3 (LAM) and 4.1.2 (Haarlem) sub-lineages in a small dataset of *Mtb* strains from Ecuador(36). According to the migratory history from Ecuador to several countries, further evidence is needed to establish the extent to which migration contributes to the TB burden and ongoing transmission among the local population.

In this study, we aim to investigate the transmission routes of 88 *Mtb* strains classified as L4 originating from Ecuador by comparing them with 415 publicly available genomes from 19 Latin-American countries. To accomplish this, we used a combination of phylogenetics and comparative genomic analysis. Additionally, we explored the role of genetic variation in virulence genes in the recent transmission of TB in Ecuador. Our findings provide valuable insights into the genetic diversity of *Mtb* in Ecuador and shed light on the potential impact of migration on the TB burden and transmission within the local population. Ultimately, this research contributes to developing more effective and targeted TB control strategies tailored to the region's specific characteristics of the local *Mtb* population.

6. Methods

Genome database of the *Mtb* samples

This study included 88 clinical isolates of *Mtb* collected from different provinces in Ecuador between 2019 to 2021, which are available in the Sequence Read Archive (SRA) database (http://www.ncbi.nlm.nih.gov/sra) under the BioProject PRJNA827129. In addition, we included 415 public sequences, previously characterized as isolates corresponding to the L4 lineage of *Mtb*, covering 19 countries cataloged in the Ecuadorian context as continental and transoceanic migratory countries. Among the continental migratory countries, we considered isolates from Argentina (n=18), Brazil (n=84), Canada (n=41), Colombia (n=8), Guatemala (n=16), Mexico (n=35), Panama (n=4), Paraguay (n=67), Peru (n=44), and USA (n=6). In contrast, transoceanic migratory countries isolates were selected from Hungary (n=5), Netherlands (n=12), Portugal (n=8), Spain (n=59), and the United Kingdom (n=5). Particularly the 59 isolates assigned to Spain were considered isolates from Latin American migrants who had arrived in Spain years ago, distributed in Bolivia (n=20), Colombia, Ecuador (n=8 each), Chile, and Honduras (n=1, each) according to(37). The accession numbers and countries' distribution were detailed in (Supplementary Table 1-2).

Variant calling, lineage classification, and drug-resistance genes

All 503 *Mtb* genomes were processed with the MTBseq pipeline using the standard input parameters(38) to obtain the complete list of variants, including the single nucleotide polymorphisms (SNPs) and the insertions and deletions (InDels), by using the reference genome of *Mtb* H37Rv (NC_000962.3). Briefly,

the MTBseq pipeline uses BWA-mem and SAMtools algorithms for mapping the reads to the reference genome, GATK v3 for base call recalibration and realignment of reads around InDels, and SAMtools mpileup for variant calling. The criteria for selecting good-quality genomes from MTBseq analysis were coverage mean >20x, read depth DP<5, and reference genome coverage >95%. The sub-lineage classification and the genetic distance matrix among transmission groups were also obtained from MTBseq analysis.

The TB-Profiler v4 pipeline was used to predict the canonical mutations in genes associated with resistance to rifampicin (RIF), isoniazid (INH), pyrazinamide (PZA), and ethambutol (EMB) [first-line drugs], as well as fluoroquinolones (FQs, including ciprofloxacin, levofloxacin, moxifloxacin, and oxifloxacin), streptomycin (STR), ethionamide (ETH), and aminoglycosides (including amikacin, kanamycin, and capreomycin) [second-line drugs]. We used BCFtool as the caller algorithm within the TB-profiler. When the isolate has only isoniazid resistance was defined as HR-TB, for rifampicin RR-TB, pre-XDR was defined as isolate MDR/RR-TB plus resistance to any fluoroquinolone, all definitions are according to WHO guidelines.

Phylogenetic and population structure analysis

Phylogenetic reconstruction was achieved using the genomic alignment of concatenated SNPs derived from MTBseq, which included only those variable positions that fulfilled coverage criteria >20x and variant frequency >95% of all samples. Repetitive regions (PPE/PE-PGRS genes), consecutive InDels, and genes involved in antibiotic resistance were excluded. The substitution model of the alignment was calculated with ModelTest-NG v0.1.7(39). The phylogenetic tree was reconstructed with the maximum–likelihood (ML) method using RAxML-NG(40), the GTR+GAMMA (General Time Reversible with Gamma-distributed rates heterogeneity) nucleotide substitution model, and 1000 bootstrap replicates. To correct the likelihood of ascertainment bias introduced by only using polymorphic sites and rescaling the branch tree, we make an ascertainment bias correction by specifying the number of invariant sites (41). The invariant sites were obtained as indicated in: https://github.com/conmeehan/pathophy. Visualization of the phylogenetic tree was achieved with iTOL v6.6 (42). The *Mycobacterium microti* (Accession number: SRR3647357) genome was used as an outgroup to obtain the rooted tree.

The package rhierBAPS (43) was used to determine the population structure of these samples dataset, with a max depth of 2 and n.pops of 20, the algorithm clusters the samples hierarchically by a nested clustering approach, based on genetic data, such as SNPs distance matrix, to identify subpopulations or clusters.

Transmission cluster analysis and modeling of the genetic clustering network

The transmission genomic transmission clusters (TGCs) were determined from concatenated high-quality SNPs. A pairwise distance threshold of 12 SNPs between strains was evaluated to obtain an optimal population clustering, as previously reported in genomic TB research(37,44–46). Then, the TGCs were classified according to the number of members in each group: *small* (less than three isolates), *medium* (three to five isolates), and *large* (>five isolates). To represent the spatial connectivity and improve the local transmission inference, TransFlow(47) was used to reconstruct the transmission network in the clusters generated with at least three samples based on genomic and epidemiological data, like sampling dates and geographic coordinates. To address the bias inherent in lineage-specific reference genomes, TransFlow employs the PANPASCO pipeline(48), which utilizes a computational pan-genome comprising 146 MTBC complete genomes from the major lineages 1–4 to conduct pairwise SNP distance calculations to address the bias inherent in lineage-specific reference genomes. This approach helps to ensure a more comprehensive and representative analysis of genetic variation within the MTBC population.

Phylodynamic of TB transmission clusters

A time-calibrated phylogeny was inferred with BEAST v1.10.4(49) based on the collection dates of isolates and tip dates. A multiple sequence alignment of concatenated SNPs derived from MTBseq created an XML file in BEAUTi. The XML file was corrected to specify the number of invariant sites as indicated in:https://groups.google.com/forum/#!topic/beastusers/QfBHMOqImFE. The temporal signal in the sequence alignments was investigated using TempEst v1.5.3 (50). Dating was performed using the GTR+GAMMA substitution model, a strict molecular clock model, and a coalescent constant size demographic model. Markov chain Monte Carlo (MCMC) chains were run for 250 million iterations with 10% burn-in and sampled every 10,000 generations to evaluate the independent convergence of the chains. Results were summarized and convergence assessed using Tracer v1.6(51) to ensure that all relevant parameters reached an adequate population size of >200. The Maximum clade credibility tree was calculated with the software Treeannotator v2.5.0.

7. Results

Genomic variability, functional annotation, and resistance profile of Ecuadorian samples.

The sequencing of the 88 *Mtb* isolates from Ecuador resulted in an average depth coverage of 60.9X±23.6. The single nucleotide variant (SNV) calling analysis identified an average of 731 ± 100.4 SNPs and 48 ± 6.0 InDels, with 105 ± 17.8 being intergenic, 372 ± 47.4 non-synonymous, and 232 ± 35.8 synonymous. The annotation of genomes revealed 4,294 ± 14.5 genes, including 4,254 coding sequences (CDS), and 44 transfer RNA (tRNA) genes, with 3,378 ± 12.0 (77%) having functional assignments. A total of 1,061 ± 4.5 proteins with Enzyme Commission (EC) numbers, 918 ± 4.0 had Gene Ontology (GO) assignments, and 814 ± 3.3 proteins were mapped to KEGG pathways. Approximately 1,961 ± 19.1 genes were assigned

196 to the subsystems (a group of proteins that implement a specific biological process or structural complex)
197 (Supplementary Table 3).

198 The analysis of pathways showed that 19.2% of the annotated protein-coding genes were associated with
199 amino acid metabolism (444 ± 2.1), followed by genes related to carbohydrates (16.2%, 375 ± 2.7), lipids
200 metabolism (13.6%, 314 ± 2.0), and xenobiotics, biodegradation, and metabolism (12.5%, 288 ± 2.8
201 genes). The genes involved in amino acid biosynthesis were conserved and essential for pathogenicity in
202 bacteria, including *Mtb*(52). Furthermore, when analyzing the phylogenetic relationship and genetic
203 diversity associated with resistance in Ecuadorian isolates, our findings revealed that 4.3.2/3 (46.6%) and
204 4.1.1 (26.1%) sub-lineages exhibited higher prevalence. We identified a total of 42 single nucleotide
205 variants (SNVs) in resistance-related genes, with the *rpoB* gene showing the most significant mutational
206 occurrence. Furthermore, the SNV were associated according to the affectation in *Mtb* genomes as follow:
207 48.13% represented Non-synonymous SNP, followed of synonymous SNP with 30.08%, Intergenic SNV
208 with 13.67%, Deletions with 4.76% and Insertions with 3.37%, been the SNV more frequently presented
209 in genes related with Intermediary metabolism and respiration (21.71%) followed for Cell wall and
210 processes (21.34%) and Conserved hypotheticals (20.44%). An exhaustive detailed of SNV is showed in
211 Supplementary Table 4.

212 To gain a deeper insight into the genomic epidemiology of tuberculosis (TB) in Ecuador, we incorporated
213 an additional 415 Mtb samples from 19 countries categorized as both continental and transoceanic
214 migratory nations. This expanded our dataset from 88 isolates to a total of 503 samples. All these *Mtb*
215 genomes belonged to the L4 lineage, including mostly from families 4.3.2/3 (35.4%), 4.1.2 (22.7%), 4.4.1
216 (12.7%), and 4.1.1 (10.7%) (Supplementary Table 2). Moreover, 63.8% (321/503) *Mtb* samples were
217 classified as genotypically susceptible to all antituberculosis drugs. In contrast, 36.2% (180/503) were
218 found to be drug-resistant to at least one antibiotic. The clinical classification of these drug-resistant
219 samples showed that 16.7% were identified as having multi-drug resistance (MDR-TB), 8.5% display
220 resistance to isoniazid (HR-TB), 4.2% were classified as pre-extensively drug-resistant (pre-XDR-TB),
221 and 1.6% were resistant to rifampicin (RR-TB). Table 1 summarizes the drug-resistant classification of the
222 503 *Mtb* samples per region.

223 We determined the frequency of canonical mutations associated with conferring resistance in all analyzed
224 sequences to establish a correlation between drug resistance and transmissibility. Among these, the most
225 frequently found were *katG* Ser315Thr (n=92), *rpoB* Ser450Leu (n=68), *rpsL* Lys43Arg (n=21), *embB*
226 Met306Ile (n=17), *pncA* Gln10Pro (n=14), and *gyrA* Ala90Val (n=11) which confer resistance to INH, RIF,
227 STR, EMB, PZA, and FQs, respectively.

228 **Table 1.** Frequency distribution of the 503 *Mtb* samples, showing their sub-lineage and their drug resistance profile
229 per geographical region.

Geographic region	HR-TB n (%)	RR-TB n (%)	MDR-TB n (%)	Pre-XDR-TB n (%)	Other n (%)	Sensitive n (%)
South America	**33 (6.6)**	**4 (0.8)**	**71 (14.1)**	**19 (3.8)**	**12 (2.4)**	**207 (41.2)**
Euro-American	3 (0.6)	2 (0.4)	4 (0.8)	1 (0.2)	6 (1.2)	10 (2.0)
Haarlem	5 (1.0)	1 (0.2)	20 (4.0)	2 (0.4)	2 (0.4)	65 (12.9)
LAM	8 (1.6)	1 (0.2)	29 (5.8)	8 (1.6)	2 (0.4)	95 (18.9)
mainly-T			1 (0.2)	1 (0.2)		15 (3.0)
S-type	15 (3.0)		10 (2.0)		1 (0.2)	8 (1.6)
X-type	2 (0.4)		7 (1.4)	7 (1.4)	1 (0.2)	14 (2.8)
Central America	**7 (1.4)**		**7 (1.4)**		**4 (0.8)**	**38 (7.6)**
Euro-American					1 (0.2)	2 (0.4)
H37Rv-like						1 (0.2)
Haarlem			2 (0.4)			4 (0.8)
LAM	2 (0.4)		1 (0.2)		1 (0.2)	16 (3.2)
mainly-T			1 (0.2)		1 (0.2)	7 (1.4)
S-type	2 (0.4)					1 (0.2)
X-type	3 (0.6)		3 (0.6)		1 (0.2)	7 (1.4)
North America					**7 (1.4)**	**40 (8.0)**
Euro-American						3 (0.6)
H37Rv-like						10 (2.0)
LAM						1 (0.2)
mainly-T					1 (0.2)	
S-type					2 (0.4)	24 (4.8)
X-type					4 (0.8)	2 (0.4)
Europe	**3 (0.6)**	**4 (0.8)**	**6 (1.2)**	**2 (0.4)**	**1 (0.2)**	**38 (7.6)**
Cameroon	1 (0.2)					1 (0.2)
Euro-American			1 (0.2)		1 (0.2)	2 (0.4)
H37Rv-like						2 (0.4)
Haarlem	1 (0.2)	1 (0.2)	1 (0.2)	1 (0.2)		9 (1.8)
LAM	1 (0.2)	1 (0.2)	3 (0.6)			9 (1.8)
M. microti						1 (0.2)
mainly-T		1 (0.2)	1 (0.2)	1 (0.2)		9 (1.8)
S-type						1 (0.2)
TUR						2 (0.4)
X-type		1 (0.2)				2 (0.4)

Other frequent canonical mutations were fabG*1* -15C>T (n=16), *ahpC* -74G>A (n=15), and *gid* 329_330delTG (n=16) associated with INH and STR resistance, respectively. A comprehensive overview of drug-resistant canonical mutations identified in the 503 Mtb samples is presented in Table 2 and Supplementary Table 5.

TB population structure and transmission clusters

Isolate clustering, based on the concatenation of high-quality SNPs, revealed four primary groups at the initial hierarchical level, aligning with distinct lineages. These clusters are visually depicted in Figure 1, with varying branch colors in the phylogeny. Subsequently, the analysis revealed the presence of 15 secondary groups at the subsequent hierarchical level, corresponding to sub-lineages. Cluster 1 (in pink) encompasses 128 isolates, distributed across three subclusters, originating from lineages 4, 4.4.1.1, 4.2.2, 4.6.2, and 4.7/8. Cluster 2 (in orange) comprises 125 isolates divided into three subclusters (4-6), with the majority belonging to the lineages 4.1.2.1, and some isolates from 4.1.2 (subcluster 6). Cluster 3 (in green) is composed of 178 isolates, distributed across three subclusters (7-9), primarily associated with the lineage 4.3.2, 4.3.3, 4.3.4.1, 4.3.4.2 and surprisingly, it includes an isolate 4.1.2 within subcluster 9. Finally, cluster 4 (in blue) encompasses six subclusters (10-15), housing 71 isolates predominantly from the lineage 4.1.1,

4.1.1.1, and 4.1.1.3. Notably, isolates form subclusters 13, 15, and a small portion of subcluster 10 corresponding to lineage 4.1. Importantly, these 15 subclusters exhibit genomic distances exceeding 12 SNPs. Consequently, we conducted an analysis of cluster composition based on pairwise comparisons of isolates, utilizing a pairwise distance threshold of <12 SNPs, to determine their inclusion within the same transmission genomic cluster (TGC).

Table 2. Canonical variants associated with drug resistance frequently distributed within the 503 *Mtb* samples.

Drug	Gene	Canonical variants (n)
RIF	*rpoB*	Ser450Leu (68), Asp435Val (13), His445Asn (6)
INH	*ahpC*	-74G>A (15)
	fabG1	-15C>T (16)
	inhA	-154G>A (9)
	katG	Ser315Thr (92)
PZA	*pncA*	Gln10Pro (14)
EMB	*embB*	Met306Ile (17), Gly406Ala (15)
STR	*gid*	329_330delTG (16)
	rpsL	Lys43Arg (21)
FQ	*gyrA*	Ala90Val (11)
	gyrB	Asp461His (2)
AMG	*rrs*	1401A>G (18)
	eis	-12C>T (2)
	tlyA	Gly232Asp (1), Lys69Glu (1)
ETH	*ethA*	1222delT (3)
	ethR	Phe110Leu (8)
PAS	*folC*	Glu40Gly (1)

#*inhA* promoter mutations include mutations in *fabG1* open reading frame (ORF) because they create alternative promoters for *inhA* and mutations upstream of *fabG1* because they act as promoters of the entire operon, which includes *inhA*. (*) Stop codon. (†) variants associated with INH and ETH resistance.

According to this strategy, 92.8% (n=467) of the isolates were classified into 51 TGCs ranging from 2 to 63 members. Considering the number of members per TGC, we classified seven as *small*, 20 as *medium*, and 24 as *large* (Supplementary Table 6). Notably, the Mtb samples identified as susceptible were more likely to be in larger clusters than the resistant samples. In most pre-XDR isolates, clonality existed between them or with members from other countries. The TGCs distribution exhibits a remarkable grouping primarily within Cluster 3 (comprising 37.3% TGCs) and Cluster 1 (constituting 29.4% TGCs). Additionally, there exists a diversity of ungrouped isolates composed of 35 isolates. Among these, 28.6% (10 out of 35 isolates) were linked to the lineages 4.3.2/3/4, while 25.7% (9 out of 35 isolates) show association with being associated principally with lineage 4.8 (Figure 2A and Table 3).

Upon closer examination of the five largest Transmission Genomic Clusters (TGCs), namely TGC_1, TGC_2, TGC_9, TGC_18, and TGC_19, distinct demographic and epidemiological profiles emerged (see

Table 3). The predominant composition of TGC_1 was primarily attributed to isolates hailing from Paraguay (46.3%), Peru (19.0%), and Spain (15.9%). TGC_18 encompassed isolates originating from Ecuador (41.0%) and Paraguay (26.8%). In contrast, TGC_19 primarily consisted of isolates from Spain (38.7%). TGC_2 predominantly comprised isolates from Argentina (53.3%), whereas TGC_9 was predominantly characterized by isolates from Canada (54.1%).

Table 3. Distribution of 503 isolates of *Mtb* according to Clustering (rhierBAPS) and TGCs approach.

Clusters classification	*n* of isolates	Distribution of Transmission Genomic Clusters (TGCs)	Largest TGCs *n* isolate, Lineage
Cluster 1	128	15 TGCs + 14 ungrouped isolates	TGC_9: 24 isolates. S-type (4.4.1.1)
Cluster 2	125	7 TGCs+ 5 ungrouped isolates + 1 isolate TGC18 (Cluster3)	TGC_1: 63 isolates. Haarlem (4.1.2.1) TGC_2: 30 isolates. Haarlem (4.1.2.1, *n*=14; 4.1.2.1.1, *n*=16)
Cluster 3	178	19 TGCs+ 10 ungrouped isolates + 1 isolate TGC14 (Cluster2)	TGC_18: 56 isolates. LAM (4.3.3) TGC_19: 31 isolates. LAM (4.3.2)
Cluster 4	71	10 TGCs+ 6 ungrouped isolates	

Analyzing the Ecuadorian *Mtb* isolates, 84 out of 88 samples cluster into 15 TGCs (5 from *medium* TGCs, and ten from *large* TGCs), and the remaining four isolates (S0137, S0516, S2192, and S2193) were ungrouped. Members of ten of these TCGs (TGC_2, _3, _5, _14, _15, _18, _20, _21, _37, and _47) were found to be related to isolates from different countries, revealing unseen links and possible routes of large-scale transmission between them. Interestingly, one of the clusters (TGC_37) presented a close relationship (genetic distance of 8-9 SNPs) between samples derived from Ecuador (*n*=5) and an Ecuadorian migrant identified in Spain (Figure 2B). The TGC_14 contained close *Mtb* samples (genetic distance of 6 SNPs) from Ecuador (*n*=2) and Brazil (*n*=5). In addition, a high clonality (60.2%, 53/88) among members of the same Ecuadorian cluster was observed. The three largest Ecuadorian TGCs, composed of 49 isolates, showed clonality in 68.8% (11/16 isolates in TGC_15), 69.6% (16/23 isolates in TGC_18), and 90.0% (9/10 isolates in TGC_11) of the cases, respectively. (Supplementary Tables 6-7 and Supplementary Figure 1) Similarly, TGC_20, TGC_30, and TGC_48 had a higher proportion of singletons (100%, 3 out of 3 isolates), followed by TGC_18 exhibited a singleton proportion of 30.4%, encompassing 7 out of 23 isolates, while TGC_15 displayed a slightly higher rate of 31.25%, involving 5 out of 16 isolates. (Supplementary Table 8)

When focusing on lineage composition, the largest TGCs were predominantly composed of 4.3.3 (46.9%, 23/49; TGC_18), followed by the 4.1.1 (32.7%, (16/49); TGC_15), and the 4.4.1.1 (20.4% (10/49); TGC_11). Regarding the distribution of drug-resistant isolates among these TGCs, we found clonal pre-XDR-TB isolates present in TGC_15 (n=4, genetic distance 0 SNPs), TGC_18 (n=2, genetic distance 0 SNPs) and TGC_37 (n=1, genetic distance 1 SNP); furthermore, 40.9% of samples from these TGCs were

derived from patients previously treated for TB, followed by treatment-naive (34.1%), and in-treatment (25%).

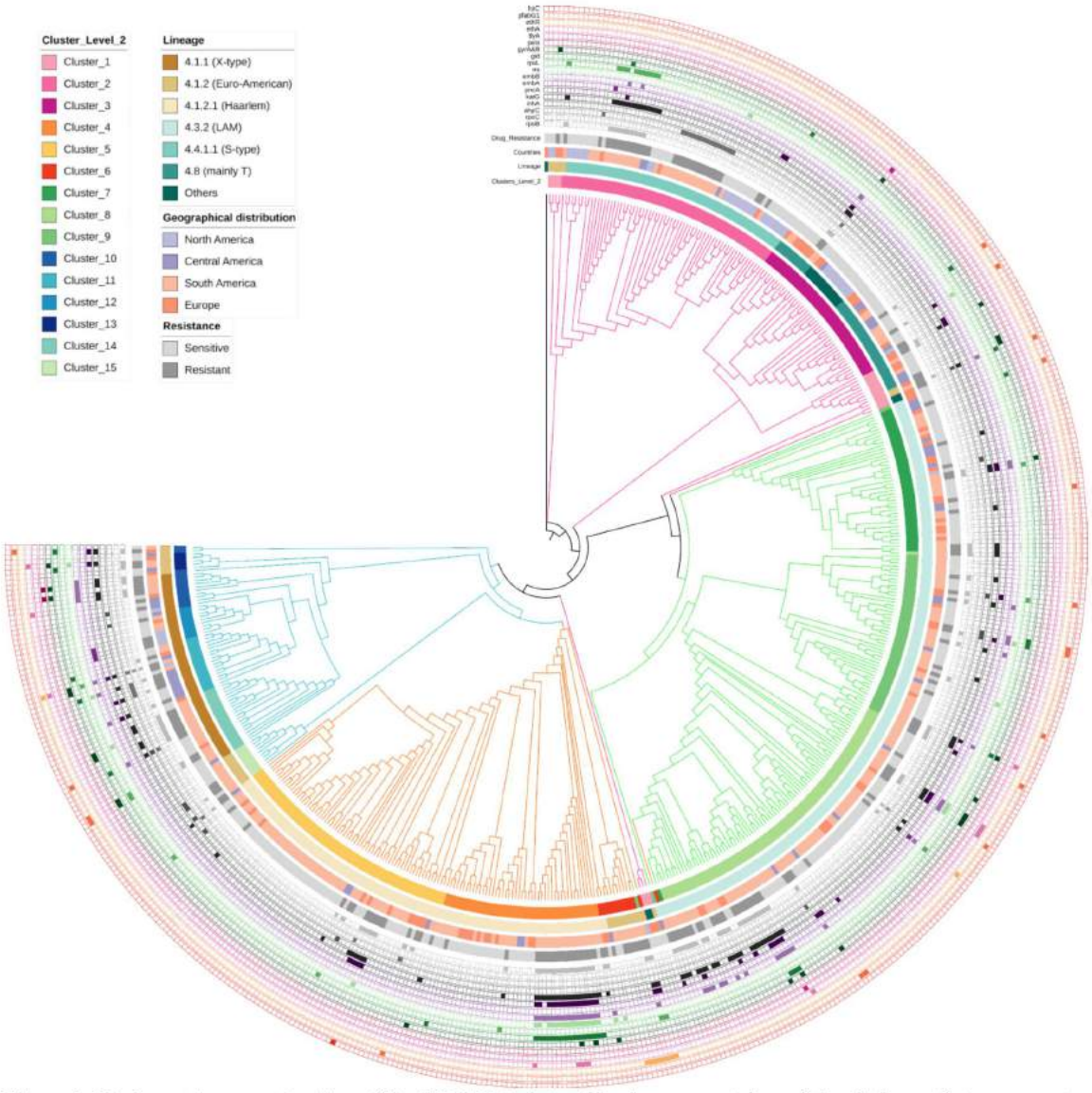

Figure 1. Phylogenetic reconstruction of the 503 *Mtb* isolates. Circular representation of the phylogenetic tree reconstructed with 1,437 SNPs using the ML method, the GTR+GAMMA substitution model, bootstrap support of 1000 replicates, and rooted with *M. microti*. Circular strips indicated the following metadata (inside out): i) fifteen subclusters classified from the second clustering level, ii) the geographic distribution of samples, iii) sub-lineage classification, iv) drug resistance clinical classifications, and v) canonical variants associated with drug resistance. Color in branches on trees represents the first level of clustering (Lineages) of *Mtb* isolates.

Genomic clustering networks

Genomic networks were inferred for the 503 *Mtb* isolates, suggesting possible transmission routes. Three strategies were applied for this identification: the rhierBAPS algorithm to cluster genetic sequence data; the application of a genomic distance threshold of 12 SNPs to obtain an optimal population clustering; and

Transflow algorithm to reconstruct the transmission network within clusters comprising a minimum of three samples, based on genomic and epidemiological data. Different distributions were revealed in both genomic cluster and network analysis (Figure 2A). Among Ecuadorian isolates (collected between 2019 to 2021), 17 out of 88 samples showed clonality with isolates from Colombia (n=2) or Latin-American migrants in Spain (n=4), which were detected in 2014 and 2015, respectively (Figure 2B).

In particular, within TGC_03, the Ecuadorian isolates S0110 and S0128 (identified in 2021) were found to exhibit complete similarity to the isolate SRR15508558 from Colombia (2020). Moreover, the Ecuadorian isolate S2186 (2020) was determined to be identical to ERR4195626 originating from Colombia (2015). Furthermore, noteworthy findings emerged in TGC_15, where Ecuadorian isolates S0106, S1063, and S1135 (2021), as well as S2203 and S2206 (2020), demonstrated clonality with the isolate ERR4195320 derived from an Ecuadorian migrant identified in Spain (2014). Furthermore, within TGC_18, a cluster of Ecuadorian isolates encompassing S0149, S0205, S0555, S0585, S0605, and S2191 (2020), alongside S0498, S1133, and S1212 (2021), displayed marked clonality with three *Mtb* isolates recovered from Latin-American migrants in Spain (2015). This cluster notably included an Ecuadorian isolate (ERR2884613) and two Bolivian isolates (ERR2884716 and ERR2884718) (See Figure 2C). Among these 17 isolates, the most predominant family affiliations were 4.3.2/3 (52.9%, n=9), followed by 4.1.1 (29.4%, n=5), and 4.1.2.1 (17.7%, n=3).

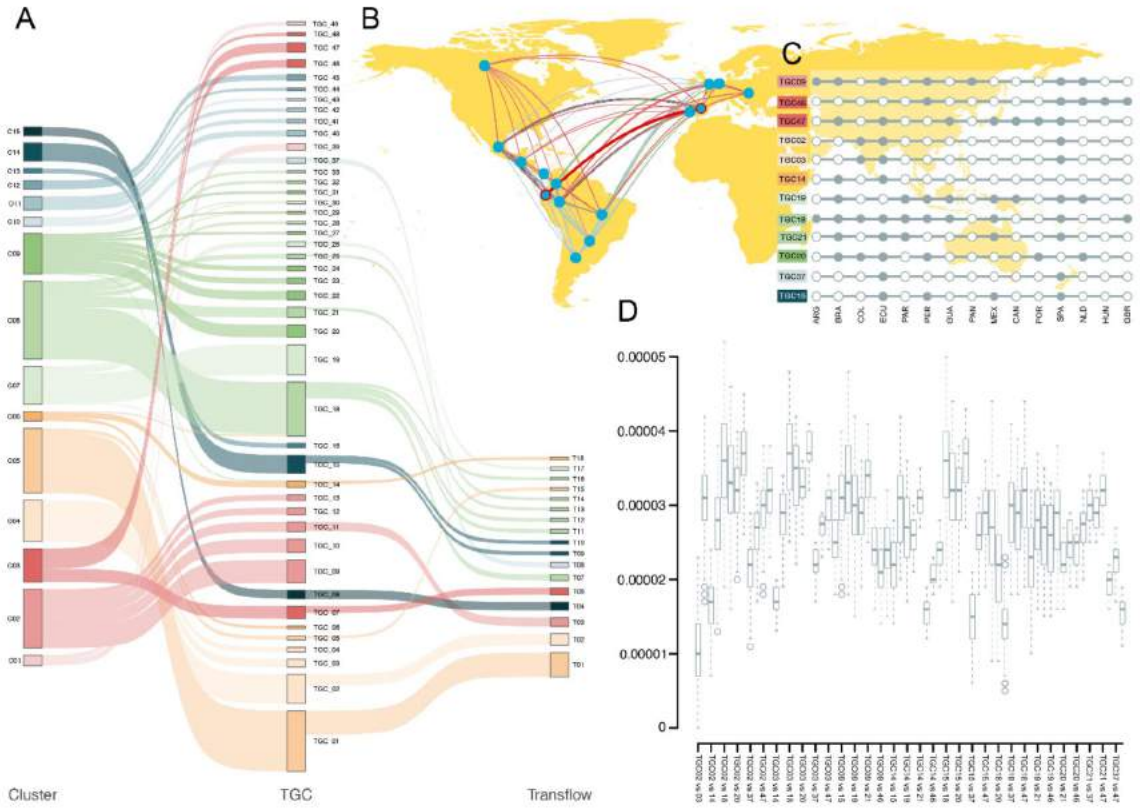

Figure 2. Genomics clustering network of the 503 *Mtb* samples. A: Plot showed the relation from Genetic cluster analyzed by rhierBAPs, TGC (<12 SNPs) by MTBseq, and Networks identified by Transflow analysis. **B**: TGCs Networks considered the Ecuadorian isolates analyzed in this study, in red highlight the connection confirmed by Transflow. The width of the links is due to the genetic distance (wider, less distance). The color of the links is according to the TGCs. More details at https://microreact.org/project/phylo-tb. **C**: The geographical distribution of TGCs from B panel revealed 12 networks. Nonetheless, two of these TGCs did not incorporate isolates from Ecuador. Interesting, isolates from Ecuadorian migrants in Spain were, however, identified within these TGCs. **D:** Boxplot displaying genetic distances both within and between the TGCs that involve Ecuadorian isolates.

However, through TransFlow analysis, we have identified 17 transmission networks. Notably, within the most extensive transmission network (TGC_01), consisting of 25 Paraguayan isolates, the isolates SRR21240246, SRR21240038, and SRR21240046 emerged as potential index cases of transmission among network members, with isolate SRR21240246 responsible for 11 cases within this transmission network. Similarly, a comparable pattern of transmission organization is evident in the second network associated with TGC_02, exclusively composed of Argentinian isolates. Additionally, within the fifth transmission network exclusive to Peruvian isolates (TGC_10), the isolate SRR10851678 from Brazil stands out as the potential index case (Supplementary Figure 2).

Ecuadorian isolates were involved in a total of eight transmission networks. Notably, isolates S0158, S0134, and S2205 were identified as potential index cases within their respective transmission networks among the Ecuadorian isolates (specifically, network three within TGC_11, network eight associated with TGC_37, and network nine considered in TGC_15). Conversely, within network seven (associated with TGC_18), it was observed that Ecuadorian isolate S2191 likely originated from the potential index case ERR2884711 found in a Bolivian isolate, with a transmission sequence that involved ERR2884716 (Bolivian), all occurring in Spain. The genomic distance difference between the index case and isolate S2191 amounted to 15 SNPs. These findings are concisely outlined in Supplementary Figure 3.

Genetic diversity of the virulence-associated genes

To characterize mutations in genes affecting adaptability to the host, we describe in Ecuadorian isolates the SNPs in genes commonly associated with MTBC virulence(52). A total of 303 SNPs were identified in these virulence genes, and all isolates were determined to have at least one SNP in the genes *mce1F, mmpL4, phoR, ctpV, pepD, mce3F, fadD13,* and *nuoG*. Additionally, genes *pks12, fadD5, mce3C, pks12, nuoG,* and *katG* were mutated with at least one SNP in more than 50% of the isolates. The top six genes containing the most significant SNPs include *plcA, plcB, pks7, pks12, phoR,* and *PPE46* (Supplementary Tables 9-10).

46.6% (41/88) of isolates had more than 40 SNPs, with a maximum of 79 SNPs identified within the virulence genes. The isolates corresponding to 4.1.1 (TGC_15 and TGC_37), 4.4.1.1 (TGC_11), and 4.3.2/3/4(TGC_18, _20, _21, _26, _29, and _30) sub-lineages presented higher numbers of polymorphisms (39 to 51, 39 to 43, and 28 to 79 SNPs, respectively). Additionally, two members of TGC_14 (S0017 and

S0039) and one ungrouped sample (S2193) showed high numbers of mutations (70, 71, and 74 SNPs, respectively) in virulence genes. Interestingly, the five isolates associated with the sub-lineage 4.7/8 showed the lowest number of SNPs in the virulence genes (15 to 25 SNPs).

Phylodynamics of TB in Ecuador

To obtain a more comprehensive understanding of the population dynamics of TB in Ecuador we estimated the time to the most recent common ancestor (TMRCA) for each TGC. Our findings showed consistent estimates of the ages of TMRCAs within different TGCs of TB in Ecuadorian isolates. The estimated TMRCA for the Ecuadorian isolate was around 1054 years before the present (YBP), placing it at approximately 967, although with a relatively broad posterior density interval of 697 to 1475 (95% highest posterior density (HPD) interval). Our analysis of *Mtb* isolates from different transmission groups in Ecuador revealed a range of TMRCA estimates, TGC_ 15 was the oldest (218 YBP) while TGC_03 was the youngest (105 to 155 YBP) followed by TGC_18 (113 to 173 YBP). Particularly, the isolates corresponding to TGC_02, _21, and _47, where only one Ecuadorian isolate was included, had more than 100 YBP (191 YBP for TGC_02, 208 YBP (TGC_21), and 229 YBP (TGC_47)), but surprisingly the isolate S0108 (TGC_47) form a clade with S0137, which is classified as ungrouped. Moreover, we found that ungrouped isolates S2192, S0137, S2193, and S0516 had TMRCA estimates of 157, 229, 373, and 404 YBP, respectively (Supplementary Figure 4).

Based on the findings, most Ecuadorian *Mtb* isolates were closely related to the LAM and X-type sub-lineages, We focused further analysis on all isolates belonging to these sub-lineages. The corresponding TMRCA estimates for LAM and X-type sub-lineages were 470 YBP (95% HPD, 358 to 720) and 450 YBP (95% HPD, 349 to 647), respectively. Within the X-type sub-lineage, TGC_15 members have a TMRCA of 291 YBP, while those from TGC_37 have a TMRCA of 254 YBP. Isolates from TGC_14 have a TMRCA of 238 YBP, while ungrouped isolates S0516 and S2193 have TMRCA estimates of 482 and 354 YBP, respectively; all of them were from the Euroamerican sub-lineage. While, the isolates corresponding to the LAM sub-lineage have a TMRCA estimate ranging from 242 to 295 YBP, including six TGCs (289 YBP (TGC_26), 242 to 265 YBP for TGC_18, 249 YBP (TGC_30), 271 to 293 YBP (TGC_20), 295 YBP (TGC_21), 292 YBP (TGC_29) and one ungrouped isolate (289 YBP for S2192), surprisingly the isolate S2192 was including into the clade that contained the TGC_26, and the TGC_18 was divided in three clades.

On the other hand, in the Haarlem sub-lineage, isolates have a TMRCA estimated to be between 237 and 292 YBP, including 263 YBP for isolates corresponding to TGC_05, 237 to 273 YBP for TGC_03 isolates, and 292 YBP for TGC_02. The S-type sub-lineage has a TMRCA of 257 YBP (TGC_11), while isolates corresponding to the mainly-T sub-lineage have a TMRCA between 278 (TGC_48) and 269 YBP

(TGC_47), while the S0137 isolate has a TMRCA of 340 YBP, particularly the isolate S0137 forms clade with TGC_48 (Supplementary Figure 5).

8. Discussion

Recurrent TB can arise from either reactivation of latent infection or reinfection with drug-resistant strains. Differentiating between these two can be challenging, especially when there are genomic variations in genes involved in persistence and pathogenicity. Human-adapted MTB strains show a high degree of genomic similarity but vary in terms of geographic distribution, virulence, transmissibility, and drug resistance pattern(53). To investigate transmission pathways within our study population, we analyzed 503 Mtb strains of the L4 lineage, mostly from neighboring regions in Ecuador and other areas. Our analysis of high-quality genomes revealed that 63.8% of isolates were sensitive to all drugs, while the remaining 36.2% were resistant to at least one drug. All genomes in this study belonged to the Euro-American lineage, with the most common sub-lineages being 4.3.2/3 (35.4%), 4.1.2.1 (22.7%), 4.4.1 (12.7%), and 4.1.1. (10.7%). In South and Central America and the Caribbean, 4.3.2/3, 4.1.2, and 4.8 sub-lineages primarily represent the L4 lineage(24,54,55), indicating their European origin, the 4.4.1, and 4.1.1 could be mostly included due to migratory processes, as other studies have highlighted (18,56,57). We identified a total of nineteen genes that harbor mutations associated with resistance to first- and second-line drugs. Among these, 63.4% were specifically linked to resistance against first-line drugs. The presence of these canonical drug-resistant genes serves as a valuable resource for further identification and characterization studies in the local context, enhancing genetic data sources for TB research(13,58,59).

Spanning network analysis using WGS enables researchers to connect TB transmission events (60,61), providing insights into the spatial and temporal dynamics of TB transmission and identifying individuals and locations that play a critical role(62,63). In this study, we used WGS to investigate the possible transmission network between *Mtb* isolates from Ecuador and other 19 countries. We found evidence for the potential spread of Ecuadorian *Mtb* isolates to individuals in different countries based on the most significant clustering proportion. In addition, in some instances, it was possible to observe the formation of transmission networks restricted to isolates from a single geographical location. Combining genetic data with epidemiological data could facilitate the management of TB transmission, particularly in immigrant communities where changing socio-epidemiological features due to migration may increase transmission complexity(64).

The influx of immigrants to and from high-burden countries is a significant driver of TB incidence in areas where transmission is limited and can also contribute to disease reactivation(37,65). TB case clusters may involve autochthonous cases, mixed multinational cases, or cases among foreign-born individuals concentrated in a specific country(66,67). Our study observed clonality between Mtb isolates from 17 Ecuadorian and from Colombia or Latin-American migrants in Spain detected in 2014 and 2015. By

reconstructing networks, it was possible to provide evidence that the potential transmission route involving migrants may have occurred through direct contact with Bolivians and Ecuadorians in Spain (TGC_18), likely in workspaces such as plantations, factories, or restaurants. Particularly, isolate S2191, part of this transmission network, exhibits clonality with other Ecuadorian isolates detected in 2020 and 2021. This finding proposes the hypothesis that S2191 could represent the initial transmission point in the local context. In the case of clonality with Colombian isolates, the formation of transmission networks is not evident. However, a similar situation could have occurred in the case of network seven (TGC_18), with shared work environments. Surprisingly, three isolates corresponding to female patients were found in both scenarios, indicating a possible personal relationship among the individuals that may contribute to transmission between them. These findings suggest possible transnational transmission events involving Ecuador and its border countries and frequent migration destinations, highlighting the need to strengthen disease surveillance to reduce the possibility of more dangerous strains of tuberculosis entering the country(68,69).

Genome comparisons offer valuable insights into the molecular mechanisms that bacteria employ to establish a pathogenic lifestyle in specific hosts. This includes the ability to survive and multiply within intracellular or extracellular host environments and induce lesions and disease(70). However, our understanding of the virulence factors expressed by *Mtb* is limited, and genetic variations resulting from selection pressure may affect their expression. Despite these challenges, performing this type of analysis may contribute to our understanding of how these factors function under local conditions(71). Our study identified 303 SNPs located within 103 genomic regions associated with virulence, with *plcA, plcB, pks7, pks12, phoR*, and *PPE46* genes displaying the highest number of SNPs. These are associated with phospholipid metabolism, specific biosynthetic pathways responsible for unique lipids, or glycolipid conjugates that are crucial for the virulence of *Mtb* and are integral components of the complex mycobacterial cell envelope(72,73). The observed modification or inactivation of these genes in our study isolates can affect the virulence of *Mtb,* leading to a less severe disease stage for the host, resulting in reduced lung damage and prolonged disease progression.

While our study has provided valuable insights, it is important to acknowledge a few limitations. Primarily, the sample size was relatively modest, consisting of 88 Mtb isolates from cultured samples, which is relatively small compared to the total number of reported TB cases in 2021(5,595). Additionally, due to limited comprehensive epidemiological information, including contact details, our understanding of transmission dynamics was constrained. To enhance our understanding, future research should aim to combine genomic data with additional epidemiological details to uncover potential transmission pathways for Mtb. Moreover, our sequencing focused on cultured sputum isolates, a standard approaches in *Mtb* genomic epidemiology which may not fully capture the complete spectrum of mycobacteria in the lungs, thus limiting our ability to capture the intricacies of within-host variation in an individual's infection (74).

Longitudinal studies have the potential to enhance sequencing analyses, uncovering variations in resistance genes and virulence factors, thereby assisting in refining treatment strategies. These insights hold the promise of contributing to the alleviation of the global TB burden.

To the best of our knowledge, this is the first study in the country to establish local transmission networks of *Mtb* using complete genome analysis. However, our findings also contribute to the knowledge of the transmission networks characterized in Ecuador based on the MIRU-VNTR approach(32,33,75). Our study provides valuable insights into the genomic and epidemiological characteristics of *Mtb* isolates in Ecuador. In addition, our analysis identified drug-resistant isolates and transmission events between individuals and across borders, underscoring the need for more extensive whole-genome sequencing and network analysis to guide public health interventions.

9. Author statements

9.1 Author contributions

GML and PMMP conceived and designed the study performed data analysis (bioinformatic processing of the raw sequencing data), and drafted the original manuscript; DAM, EGM, and JCFC performed data analysis and wrote the draft manuscript; CLC provided advice on data analysis; PF and LB conceived and designed the study wrote the main manuscript and reviewed it; and GM and DAM performed the funding acquisition. All authors read and approved the final manuscript.

9.2 Conflicts of interest

The author(s) declare that there are no conflicts of interest.

9.3 Funding information

GM is a doctoral student in the PEDECIBA program. LB and PF are members of PEDECIBA and the Sistema Nacional de Investigadores (SNI) of ANII. The funders had no role in the study design, data collection, analysis, the decision to publish, or the preparation of the manuscript. This work received no specific grant from any funding agency.

9.4 Ethical approval

Our study was reviewed and approved by the ethics committee of University Espíritu Santo under code 2022-001A. The research was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. Informed consent of the patients was not required since our study worked with isolates from a collection. The ethics committee determined the exception of informed consent because data of the isolates were anonymized, and no data on the patients were disclosed.

9.5 Acknowledgements

We thank all the researchers from the different BioProject that were used in this study for the availability of sequences on the SRA website. To the Omics Sciences laboratory of the Faculty of Health Sciences, for the availability of its servers.

10. References

1. World Health Organization. Global Tuberculosis Report, 2022 [Internet]. 2022 [cited 2022 Nov 6]. Available from: https://www.who.int/teams/global-tuberculosis-programme/tb-reports
2. Souza M do R, Paz WS da, Sales VB dos S, Jesus GFH de, Tavares D dos S, Lima SVMA, et al. Impact of the COVID-19 Pandemic on the Diagnosis of Tuberculosis in Brazil: Is the WHO End TB Strategy at Risk? Front Pharmacol [Internet]. 2022 [cited 2023 Mar 13];13. Available from: https://www.frontiersin.org/articles/10.3389/fphar.2022.891711
3. Jones AJ, Jones-López EC, Butler-Wu SM, Wilson ML, Rodman J, Flors L, et al. Impact of COVID-19 on diagnosis and testing for TB in a high-resource, low-burden setting. Int J Tuberc Lung Dis. 2022 Sep;26(9):888–90.
4. World Bank. Incidence of tuberculosis (per 100,000 people) - Ecuador | Data [Internet]. [cited 2023 Feb 18]. Available from: https://data.worldbank.org/indicator/SH.TBS.INCD?end=2021&locations=EC&name_desc=false&start=2002&view=chart
5. Mushomi JA, Palattiyil G, Bukuluki P, Sidhva D, Myburgh ND, Nair H, et al. Impact of Coronavirus Disease (COVID-19) Crisis on Migrants on the Move in Southern Africa: Implications for Policy and Practice. Health Syst Reform. 2022 Jan 1;8(1):e2019571.
6. Mellado-Sola I, Rodríguez-Molino P, Armas EA, Nogueira López J, Falces-Romero I, Rey CC, et al. Impact of Coronavirus Pandemic on Tuberculosis and Other Imported Diseases Screening among Migrant Minors in Spain. Trop Med Infect Dis. 2023 Jan;8(1):28.
7. Jokisch B. migrationpolicy.org. 2007 [cited 2023 Oct 31]. Ecuador: Diversity in Migration. Available from: https://www.migrationpolicy.org/article/ecuador-diversity-migration
8. Herrera G. Migration and Migration Policy in Ecuador in 2000–2021. 2022 Oct 18;
9. Jokisch BD. migrationpolicy.org. 2023 [cited 2023 Oct 31]. Ecuador Juggles Rising Emigration and Challenges Accommodating Venezuelan Arrivals. Available from: https://www.migrationpolicy.org/article/ecuador-migration-trends-emigration-venezuelans
10. Cavalcante SC, Durovni B, Barnes GL, Souza FBA, Silva RF, Barroso PF, et al. Community-randomized trial of enhanced DOTS for tuberculosis control in Rio de Janeiro, Brazil. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2010 Feb;14(2):203–9.

11. Yang T, Gan M, Liu Q, Liang W, Tang Q, Luo G, et al. SAM-TB: a whole genome sequencing data analysis website for detection of Mycobacterium tuberculosis drug resistance and transmission. Brief Bioinform. 2022 Feb 24;23(2):bbac030.
12. Lam C, Martinez E, Crighton T, Furlong C, Donnan E, Marais BJ, et al. Value of routine whole genome sequencing for Mycobacterium tuberculosis drug resistance detection. Int J Infect Dis. 2021 Dec;113:S48–54.
13. Wang L, Yang J, Chen L, Wang W, Yu F, Xiong H. Whole-genome sequencing of *Mycobacterium tuberculosis* for prediction of drug resistance. Epidemiol Infect. 2022;150:e22.
14. Brown AC. Whole-Genome Sequencing of Mycobacterium tuberculosis Directly from Sputum Samples. In: Parish T, Kumar A, editors. Mycobacteria Protocols [Internet]. New York, NY: Springer US; 2021 [cited 2022 Jun 18]. p. 459–80. (Methods in Molecular Biology). Available from: https://doi.org/10.1007/978-1-0716-1460-0_20
15. Torres Ortiz A, Coronel J, Vidal JR, Bonilla C, Moore DAJ, Gilman RH, et al. Genomic signatures of pre-resistance in Mycobacterium tuberculosis. Nat Commun. 2021 Dec;12(1):7312.
16. Freschi L, Vargas R, Husain A, Kamal SMM, Skrahina A, Tahseen S, et al. Population structure, biogeography and transmissibility of Mycobacterium tuberculosis. Nat Commun. 2021 Dec;12(1):6099.
17. Stucki D, Brites D, Jeljeli L, Coscolla M, Liu Q, Trauner A, et al. Mycobacterium tuberculosis lineage 4 comprises globally distributed and geographically restricted sublineages. Nat Genet. 2016 Dec;48(12):1535–43.
18. Brynildsrud OB, Pepperell CS, Suffys P, Grandjean L, Monteserin J, Debech N, et al. Global expansion of Mycobacterium tuberculosis lineage 4 shaped by colonial migration and local adaptation. Sci Adv. 2018 Oct 17;4(10):eaat5869.
19. Ajawatanawong P, Yanai H, Smittipat N, Disratthakit A, Yamada N, Miyahara R, et al. A novel Ancestral Beijing sublineage of Mycobacterium tuberculosis suggests the transition site to Modern Beijing sublineages. Sci Rep. 2019 Sep 23;9:13718.
20. Ektefaie Y, Dixit A, Freschi L, Farhat MR. Globally diverse Mycobacterium tuberculosis resistance acquisition: a retrospective geographical and temporal analysis of whole genome sequences. Lancet Microbe. 2021 Mar 1;2(3):e96–104.
21. Napier G, Campino S, Merid Y, Abebe M, Woldeamanuel Y, Aseffa A, et al. Robust barcoding and identification of Mycobacterium tuberculosis lineages for epidemiological and clinical studies. Genome Med. 2020 Dec;12(1):114.

559 22. Coll F, McNerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, et al. A robust SNP
560 barcode for typing Mycobacterium tuberculosis complex strains. Nat Commun. 2014
561 Dec;5(1):4812.
562 23. Senghore M, Diarra B, Gehre F, Otu J, Worwui A, Muhammad AK, et al. Evolution of Mycobacterium
563 tuberculosis complex lineages and their role in an emerging threat of multidrug resistant
564 tuberculosis in Bamako, Mali. Sci Rep. 2020 Dec;10(1):327.
565 24. Demay C, Liens B, Burguière T, Hill V, Couvin D, Millet J, et al. SITVITWEB – A publicly available
566 international multimarker database for studying Mycobacterium tuberculosis genetic diversity and
567 molecular epidemiology. Infect Genet Evol. 2012 Jun 1;12(4):755–66.
568 25. Santos-Lazaro D, Gavilan RG, Solari L, Vigo AN, Puyen ZM. Whole genome analysis of extensively
569 drug resistant Mycobacterium tuberculosis strains in Peru. Sci Rep. 2021 Dec;11(1):9493.
570 26. Tatara MB, Perdigão J, Viveiros M, Kritski A, Silva KE da, Sacchi FPC, et al. Genetic Diversity and
571 Molecular Epidemiology of Mycobacterium tuberculosis in Roraima State, Brazil. Am J Trop Med
572 Hyg. 2019 Aug 6;101(4):774–9.
573 27. Lagos J, Couvin D, Arata L, Tognarelli J, Aguayo C, Leiva T, et al. Analysis of Mycobacterium
574 tuberculosis Genotypic Lineage Distribution in Chile and Neighboring Countries. Manganelli R,
575 editor. PLOS ONE. 2016 Aug 12;11(8):e0160434.
576 28. Jiménez-Ruano AC, Madrazo-Moya CF, Cancino-Muñoz I, Mejía-Ponce PM, Licona-Cassani C, Comas
577 I, et al. Whole genomic sequencing based genotyping reveals a specific X3 sublineage restricted to
578 Mexico and related with multidrug resistance. Sci Rep. 2021 Jan 21;11(1):1870.
579 29. Mejía-Ponce PM, Ramos-González EJ, Ramos-García AA, Lara-Ramírez EE, Soriano-Herrera AR,
580 Medellín-Luna MF, et al. Genomics epidemiology analysis reveals hidden signatures of drug
581 resistance in Mycobacterium tuberculosis [Internet]. bioRxiv; 2022 [cited 2023 May 3]. p.
582 2022.03.15.484552. Available from:
583 https://www.biorxiv.org/content/10.1101/2022.03.15.484552v1
584 30. Arias APJ, Lahiguera MJ, Borrás R, Cardona CG, Silva MG, López MJV, et al. Comparative study of the
585 genetic diversity of Mycobacterium tuberculosis Complex by Simplified Amplified Fragment Length
586 Polymorphism and Mycobacterial Interspersed Repetitive Unit Variable Number Tandem Repeat
587 Analysis. Rev Ecuat Med Cienc Biológicas [Internet]. 2018 May 21 [cited 2022 Jul 4];39(1). Available
588 from: https://remcb-puce.edu.ec/remcb/article/view/568
589 31. Zurita J, Espinel N, Barba P, Ortega-Paredes D, Zurita-Salinas C, Rojas Y, et al. Genetic diversity and
590 drug resistance of Mycobacterium tuberculosis in Ecuador. Int J Tuberc Lung Dis. 2019 Feb
591 1;23(2):166–73.

592 32. Garzon-Chavez D, Zurita J, Mora-Pinargote C, Franco-Sotomayor G, Leon-Benitez M, Granda-Pardo
593 JC, et al. Prevalence, Drug Resistance, and Genotypic Diversity of the Mycobacterium tuberculosis
594 Beijing Family in Ecuador. Microb Drug Resist. 2019 Jul;25(6):931–7.
595 33. Garzon-Chavez D, Garcia-Bereguiain MA, Mora-Pinargote C, Granda-Pardo JC, Leon-Benitez M,
596 Franco-Sotomayor G, et al. Population structure and genetic diversity of Mycobacterium
597 tuberculosis in Ecuador. Sci Rep. 2020 Dec;10(1):6237.
598 34. Xu G, Mao X, Wang J, Pan H. Clustering and recent transmission of Mycobacterium tuberculosis in
599 a Chinese population. Infect Drug Resist. 2018 Mar 6;11:323–30.
600 35. Ng IC, Wen TH, Yang ST, Fang CT, Hsueh PR. Detecting tuberculosis clusters in urban neighborhoods,
601 Taipei, Taiwan: Linking geographic and genotyping evidence. Appl Geogr. 2019 Mar 1;104:56–64.
602 36. Morey-León G, Andrade-Molina D, Fernández-Cadena JC, Berná L. Comparative genomics of drug-
603 resistant strains of Mycobacterium tuberculosis in Ecuador. BMC Genomics. 2022 Dec 21;23(1):844.
604 37. Cancino-Muñoz I, López MG, Torres-Puente M, Villamayor LM, Borrás R, Borrás-Máñez M, et al.
605 Population-based sequencing of Mycobacterium tuberculosis reveals how current population
606 dynamics are shaped by past epidemics. Lewnard J, Franco E, Meehan CJ, Walker TM, editors. eLife.
607 2022 Jul 26;11:e76605.
608 38. Kohl TA, Utpatel C, Schleusener V, De Filippo MR, Beckert P, Cirillo DM, et al. MTBseq: a
609 comprehensive pipeline for whole genome sequence analysis of Mycobacterium tuberculosis
610 complex isolates. PeerJ. 2018 Nov 13;6:e5895.
611 39. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. ModelTest-NG: A New and Scalable
612 Tool for the Selection of DNA and Protein Evolutionary Models. Mol Biol Evol. 2020 Jan 1;37(1):291–
613 4.
614 40. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly
615 tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019 Nov 1;35(21):4453–5.
616 41. Coscolla M, Gagneux S, Menardo F, Loiseau C, Ruiz-Rodriguez P, Borrell S, et al. Phylogenomics of
617 Mycobacterium africanum reveals a new lineage and a complex evolutionary history. Microb
618 Genomics. 2021;7(2):000477.
619 42. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and
620 annotation. Nucleic Acids Res. 2021 Apr 22;49(W1):W293–6.
621 43. Cheng L, Connor TR, Sirén J, Aanensen DM, Corander J. Hierarchical and Spatially Explicit Clustering
622 of DNA Sequences with BAPS Software. Mol Biol Evol. 2013 May 1;30(5):1224–8.
623 44. Jajou R, Kohl TA, Walker T, Norman A, Cirillo DM, Tagliani E, et al. Towards standardisation:
624 comparison of five whole genome sequencing (WGS) analysis pipelines for detection of
625 epidemiologically linked tuberculosis cases. Eurosurveillance. 2019 Dec 12;24(50):1900130.

626 45. Hall MB, Rabodoarivelo MS, Koch A, Dippenaar A, George S, Grobbelaar M, et al. Evaluation of
627 Nanopore sequencing for Mycobacterium tuberculosis drug susceptibility testing and outbreak
628 investigation: a genomic analysis. Lancet Microbe. 2023 Feb 1;4(2):e84–92.
629 46. Séraphin MN, Norman A, Rasmussen EM, Gerace AM, Chiribau CB, Rowlinson MC, et al. Direct
630 transmission of within-host Mycobacterium tuberculosis diversity to secondary cases can lead to
631 variable between-host heterogeneity without de novo mutation: A genomic investigation.
632 EBioMedicine. 2019 Sep 1;47:293–300.
633 47. Pan J, Li X, Zhang M, Lu Y, Zhu Y, Wu K, et al. TransFlow: a Snakemake workflow for transmission
634 analysis of Mycobacterium tuberculosis whole-genome sequencing data. Bioinformatics. 2023 Jan
635 1;39(1):btac785.
636 48. Jandrasits C, Kröger S, Haas W, Renard BY. Computational pan-genome mapping and pairwise SNP-
637 distance improve detection of Mycobacterium tuberculosis transmission clusters. PLOS Comput
638 Biol. 2019 Dec 9;15(12):e1007527.
639 49. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and
640 phylodynamic data integration using BEAST 1.10. Virus Evol. 2018 Jan 1;4(1):vey016.
641 50. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous
642 sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016 Apr 9;2(1):vew007.
643 51. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian
644 Phylogenetics Using Tracer 1.7. Syst Biol. 2018 Sep;67(5):901–4.
645 52. Smith I. *Mycobacterium tuberculosis* Pathogenesis and Molecular Determinants of Virulence. Clin
646 Microbiol Rev. 2003;16(3):463–96.
647 53. Zakham F, Sironen T, Vapalahti O, Kant R. Pan and Core Genome Analysis of 183 Mycobacterium
648 tuberculosis Strains Revealed a High Inter-Species Diversity among the Human Adapted Strains.
649 Antibiotics. 2021 Apr 28;10(5):500.
650 54. Realpe T, Correa N, Rozo JC, Ferro BE, Gomez V, Zapata E, et al. Population Structure among
651 Mycobacterium tuberculosis Isolates from Pulmonary Tuberculosis Patients in Colombia. PLoS ONE.
652 2014 Apr 18;9(4):e93848.
653 55. Díaz Acosta CC, Russomando G, Candia N, Ritacco V, Vasconcellos SEG, de Berrêdo Pinho Moreira
654 M, et al. Exploring the "Latin American Mediterranean" family and the RDRio lineage in
655 Mycobacterium tuberculosis isolates from Paraguay, Argentina and Venezuela. BMC Microbiol.
656 2019 Dec;19(1):131.
657 56. Woodman M, Haeusler IL, Grandjean L. Tuberculosis Genetic Epidemiology: A Latin American
658 Perspective. Genes. 2019 Jan;10(1):53.

659 57. Wiens KE, Woyczynski LP, Ledesma JR, Ross JM, Zenteno-Cuevas R, Goodridge A, et al. Global
660 variation in bacterial strains that cause tuberculosis disease: a systematic review and meta-analysis.
661 BMC Med. 2018 Oct 30;16:196.
662 58. Chernyaeva E, Rotkevich M, Krasheninnikova K, Lapidus A, Polev DE, Solovieva N, et al. Genomic
663 Variations in Drug Resistant Mycobacterium tuberculosis Strains Collected from Patients with
664 Different Localization of Infection. Antibiotics. 2020 Dec 31;10(1):27.
665 59. do Carmo Guimarães Diniz J de L, von Groll A, Unis G, Dalla-Costa ER, Rosa Rossetti ML, Vianna JS,
666 et al. Whole-genome sequencing as a tool for studying the microevolution of drug-resistant serial
667 Mycobacterium tuberculosis isolates. Tuberculosis. 2021 Dec;131:102137.
668 60. Yassine E, Galiwango R, Ssengooba W, Ashaba F, Joloba ML, Zalwango S, et al. Assessing a
669 transmission network of Mycobacterium tuberculosis in an African city using single nucleotide
670 polymorphism threshold analysis. MicrobiologyOpen. 2021 Jun 16;10(3):e1211.
671 61. López-Cortés A, Zambrano AK, Guevara-Ramírez P, Echeverría BA, Guerrero S, Cabascango E, et al.
672 Clinical, genomics and networking analyses of a high-altitude native American Ecuadorian patient
673 with congenital insensitivity to pain with anhidrosis: a case report. BMC Med Genomics. 2020 Aug
674 17;13:113.
675 62. Cudahy PG, Andrews JR, Bilinski A, Dowdy DW, Mathema B, Menzies NA, et al. Spatially targeted
676 screening to reduce tuberculosis transmission in high incidence settings: a systematic review and
677 synthesis. Lancet Infect Dis. 2019 Mar;19(3):e89–95.
678 63. Alaridah N, Hallbäck ET, Tångrot J, Winqvist N, Sturegård E, Florén-Johansson K, et al. Transmission
679 dynamics study of tuberculosis isolates with whole genome sequencing in southern Sweden. Sci
680 Rep. 2019 Mar 20;9(1):4931.
681 64. Abascal E, Pérez-Lago L, Martínez-Lirola M, Chiner-Oms Á, Herranz M, Chaoui I, et al. Whole
682 genome sequencing–based analysis of tuberculosis (TB) in migrants: rapid tools for cross-border
683 surveillance and to distinguish between recent transmission in the host country and new
684 importations. Eurosurveillance. 2019 Jan 24;24(4):1800005.
685 65. Meumann EM, Horan K, Ralph AP, Farmer B, Globan M, Stephenson E, et al. Tuberculosis in
686 Australia's tropical north: a population-based genomic epidemiological study. Lancet Reg Health –
687 West Pac [Internet]. 2021 Oct 1 [cited 2023 Apr 17];15. Available from:
688 https://www.thelancet.com/journals/lanwpc/article/PIIS2666-6065(21)00138-3/fulltext
689 66. Yang C, Lu L, Warren JL, Wu J, Jiang Q, Zuo T, et al. Internal migration and the transmission dynamics
690 of tuberculosis in Shanghai, China: an epidemiological, spatial, and genomic analysis. Lancet Infect
691 Dis. 2018 Jul;18(7):788–95.

692 67. Ayabina D, Ronning JO, Alfsnes K, Debech N, Brynildsrud OB, Arnesen T, et al. Genome-based
693 transmission modelling separates imported tuberculosis from recent transmission within an
694 immigrant population. Microb Genomics. 2018;4(10):e000219.
695 68. Pescarini JM, Simonsen V, Ferrazoli L, Rodrigues LC, Oliveira RS, Waldman EA, et al. Migration and
696 tuberculosis transmission in a middle-income country: a cross-sectional study in a central area of
697 São Paulo, Brazil. BMC Med. 2018 Apr 30;16(1):62.
698 69. Sanabria GE, Sequera G, Aguirre S, Méndez J, dos Santos PCP, Gustafson NW, et al. Phylogeography
699 and transmission of Mycobacterium tuberculosis spanning prisons and surrounding communities
700 in Paraguay. Nat Commun. 2023 Jan 19;14(1):303.
701 70. Sapriel G, Brosch R. Shared Pathogenomic Patterns Characterize a New Phylotype, Revealing
702 Transition toward Host-Adaptation Long before Speciation of Mycobacterium tuberculosis.
703 Genome Biol Evol. 2019 Aug 1;11(8):2420–38.
704 71. Forrellad MA, Klepp LI, Gioffré A, Sabio y García J, Morbidoni HR, Santangelo M de la P, et al.
705 Virulence factors of the Mycobacterium tuberculosis complex. Virulence. 2013 Jan 1;4(1):3–66.
706 72. Quadri LEN. Biosynthesis of mycobacterial lipids by polyketide synthases and beyond. Crit Rev
707 Biochem Mol Biol. 2014;49(3):179–211.
708 73. Le Chevalier F, Cascioferro A, Frigui W, Pawlik A, Boritsch EC, Bottai D, et al. Revisiting the role of
709 phospholipases C in virulence and the lifecycle of Mycobacterium tuberculosis. Sci Rep. 2015 Nov
710 25;5(1):16918.
711 74. Kaplan G, Post FA, Moreira AL, Wainwright H, Kreiswirth BN, Tanverdi M, et al. Mycobacterium
712 tuberculosis Growth at the Cavity Surface: a Microenvironment with Failed Immunity. Infect
713 Immun. 2003 Dec;71(12):7099–108.
714 75. Garcés E, Cifuentes L, Franco G, Romero-Sandoval N, Arias PJ. Study of the distribution of lineages
715 of Mycobacterium tuberculosis in a prison in Guayaquil, Ecuador. 2023;

Artículo 3: A precision overview of genomic resistance screening in Ecuadorian isolates of *Mycobacterium tuberculosis* **using web-based bioinformatics tools.** PlosOne (https://doi.org/10.1371/journal.pone.0294670).

En este artículo se evaluó la información obtenida mediante la secuenciación de 88 muestras de *M. tuberculosis* (64 resistentes y 24 sensibles) en Ecuador. Se utilizaron varios programas web Mykrobe, TB-profiler, PhyReSse, KvarQ, y SAM-TB para detectar las variantes presentes en aislados. El objetivo fue determinar la concordancia entre estos métodos y el enfoque tradicional, con el fin de evaluar su aplicación en programas de vigilancia. Nuestros datos muestran que el sublinaje principal identificado en los aislamientos fue LAM, representando el 44,3 %de los casos, seguido por el sublinaje X-type, que representó el 23,0 %. En cuanto a la detección de resistencia global, el programa Mykrobe mostró una alta correlación con los resultados del DST (98 % de acuerdo y un índice Kappa de Cohen de 0,941). Sin embargo, para el análisis de fármacos de primera línea de forma individual, se observó que SAM-TB, TB-profiler, PhyReSse y Mykrobe presentaron una mejor correlación con los resultados del DST Hemos observado que el 50 % de las mutaciones identificadas por todas las herramientas basadas en la web corresponden a mutaciones canónicas en las regiones *rpoB, katG, embB, pncA, gyrA* y *rrs*. Nuestros resultados sugieren que SAM-TB, TB-profiler, PhyReSse y Mykrobe fueron las herramientas web más efectivas para detectar estas mutaciones canónicas relacionadas con la resistencia. Sin embargo, es necesario realizar más análisis para mejorar la detección de mutaciones relacionadas con drogas de segunda línea. Mejorar los programas de vigilancia de los aislamientos de TB mediante uso de herramientas de secuenciación de genoma completo para medicamentos de primera línea, TB-MDR y TB-XDR es una prioridad para comprender mejor la epidemiología molecular de esta enfermedad en nuestro país.

PLOS ONE

RESEARCH ARTICLE

A precision overview of genomic resistance screening in Ecuadorian isolates of *Mycobacterium tuberculosis* using web-based bioinformatics tools

Gabriel Morey-León[1,2,3], Paulina M. Mejía-Ponce[4], Juan Carlos Granda Pardo[5], Karen Muñoz-Mawyin[6], Juan Carlos Fernández-Cadena[7], Evelyn García-Moreira[8], Derly Andrade-Molina[1,6], Cuauhtémoc Licona-Cassani[4], Luisa Berná[9,10]*

1 Facultad de Ciencias de la Salud, Universidad Espíritu Santo, Samborondón, Ecuador, **2** Universidad de la República, Montevideo, Uruguay, **3** University of Guayaquil, Guayaquil, Ecuador, **4** Escuela de Ingeniería y Ciencias, Tecnológico de Monterrey, Monterrey, Nuevo León, México, **5** Centro de Referencia Nacional de Micobacterias, Instituto Nacional de Investigación en Salud Pública Dr Leopoldo Izquieta Perez, INSPI-LIP, Guayaquil, Ecuador, **6** Laboratorio de Ciencias Ómicas, Universidad Espíritu Santo, Samborondón, Ecuador, **7** African Genome Center, University Mohammed VI Polytechnic (UM6P), Ben Guerir, Morocco, **8** Instituto Superior Tecnológico Argos, Guayaquil, Ecuador, **9** Laboratorio de Interacciones Hospedero-Patógeno, Unidad de Biología Molecular, Institut Pasteur de Montevideo, Montevideo, Uruguay, **10** Unidad de Genómica Evolutiva, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay

* lberna@pasteur.edu.uy

Citation: Morey-León G, Mejía-Ponce PM, Granda Pardo JC, Muñoz-Mawyin K, Fernández-Cadena JC, García-Moreira E, et al. (2023) A precision overview of genomic resistance screening in Ecuadorian isolates of *Mycobacterium tuberculosis* using web-based bioinformatics tools. PLoS ONE 18(12): e0294670. https://doi.org/10.1371/journal.pone.0294670

Editor: Christophe Sola, Universite Paris XI: Universite Paris-Saclay, FRANCE

Received: March 18, 2023

Accepted: November 7, 2023

Published: December 5, 2023

Data Availability Statement: Raw sequence reads of all Mycobacterium tuberculosis isolates subjected these WGS analysis were deposited in the Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra) accessible under the BioProject number PRJNA827129.

Funding: This work was funded by grant FCI-016-2017 from the University of Guayaquil. The funders had no role in study design, data collection and

Abstract

Introduction

Tuberculosis (TB) is among the deadliest diseases worldwide, and its impact is mainly due to the continuous emergence of resistant isolates during treatment due to the laborious process of resistance diagnosis, nonadherence to treatment and circulation of previously resistant isolates of *Mycobacterium tuberculosis*. In this study, we evaluated the performance and functionalities of web-based tools, including Mykrobe, TB-profiler, PhyResSE, KvarQ, and SAM-TB, for detecting resistance in 88 Ecuadorian isolates of *Mycobacterium tuberculosis* drug susceptibility tested previously. Statistical analysis was used to determine the correlation between genomic and phenotypic analysis. Our results showed that with the exception of KvarQ, all tools had the highest correlation with the conventional drug susceptibility test (DST) for global resistance detection (98% agreement and 0.941 Cohen's kappa), while SAM-TB, PhyResSE, TB-profiler and Mykrobe had better correlations with DST for first-line drug analysis individually. We also identified that in our study, only 50% of mutations characterized by the web-based tools in the *rpoB*, *katG*, *embB*, *pncA*, *gyrA* and *rrs* regions were canonical and included in the World Health Organization (WHO) catalogue. Our findings suggest that SAM-TB, PhyResSE, TB-profiler and Mykrobe were efficient in determining canonical resistance-related mutations, but more analysis is needed to improve second-line detection. Improving surveillance programs using whole-genome sequencing tools for first-line drugs, MDR-TB and XDR-TB is essential to understand the molecular epidemiology of TB in Ecuador.

analysis, decision to publish, or preparation of the manuscript. GM is a doctoral student in the PEDECIBA program. LB is a member of PEDECIBA and the Sistema Nacional de Investigadores (SNI) of ANII. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Importance

Tuberculosis, an infectious disease caused by *Mycobacterium tuberculosis*, most commonly affects the lungs and is often spread through the air when infected people cough, sneeze, or spit. However, despite the existence of effective drug treatment, patient adherence, long duration of treatment, and late diagnosis have reduced the effectiveness of therapy and increased drug resistance. The increase in resistant cases, added to the impact of the COVID-19 pandemic, has highlighted the importance of implementing efficient and timely diagnostic methodologies worldwide. The significance of our research is in evaluating and identifying a more efficient and user-friendly web-based tool to characterize resistance in *Mycobacterium tuberculosis* by whole-genome sequencing, which will allow more routine application to improve TB strain surveillance programs locally.

Introduction

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* (Mtb), is one of the top 10 causes of death worldwide. In 2021, the World Health Organization (WHO) estimated that 10.6 million people were infected and 1.6 million people died [1]. Despite many innovations in tuberculosis diagnosis, drug resistance detection, drug therapy, prevention and control programs, several challenges, such as patient adherence, long duration of treatment, and late diagnosis, continue to limit the effectiveness of TB therapy [2]. The increase in resistant TB patients is not only due to exposure to multidrug-resistant (MDR) and extensively resistant strains (XDR) but also due to late or inadequate diagnosis (associated with slow growth of Mtb), ineffective treatment or poor adherence to treatment [2]. On the other hand, the accumulation of point mutations in coding regions for drug targets and/or drug-converting enzymes is a major mechanism for acquiring resistance in Mtb [3], which has further complicated the situation, making the timely detection of resistance-conferring mutations crucial for effective treatment and prevention of onwards transmission [4], for which the use of procedures that reduce the time of diagnosis is recommended.

The emergence of whole-genome sequencing (WGS) as a tool for detecting drug resistance in Mtb has revolutionized tuberculosis diagnosis and treatment [5]. Compared to traditional drug susceptibility testing, WGS provides results in a fraction of the time [6], making it a faster and more effective method for detecting resistance to all drugs simultaneously, identifying lineages, tracing transmission, and defining outbreaks [7–11]. With large datasets generated in tuberculosis research by WGS, the emergence of bioinformatics web-based tools such as TBProfiler [12], KvarQ [13], TGS-TB [14], Mykrobe Predictor TB [15], CASTB [16], PhyTB [17], ReSeqTB-UVP [18], GenTB [19], PhyResSE [20], SAM-TB [21] and others has significantly improved the efficiency of genotyping and drug resistance identification in *Mycobacterium tuberculosis*. These tools are popular due to their ease of use, ability to classify SNPs, feasibility of batch analysis, and user-friendly interfaces. However, they do have some technical limitations, such as the lack of standardized criteria for discerning noncanonical mutations, and selecting mutations with diagnostic criteria for resistance in new and repurposed antibiotics remains a challenge for utilizing WGS technology effectively [22]. Although web-based tools for TB lineage and resistance characterization have limitations, their combination with the WHO updated catalogue of resistance-associated mutations in TB [23] has enabled efficient implementation of WGS in surveillance programmes [24–26]. Despite these challenges,

WGS technology offers reliable and efficient results that can guide TB diagnosis and treatment. The potential benefits of WGS in TB diagnosis and treatment make it a promising tool for improving public health policies and global TB control efforts.

The incidence of tuberculosis has been on the rise in many countries, including Ecuador, since 2015 [27]. The COVID-19 pandemic has added further complexity to this issue. The emergence of drug-resistant strains is a serious global threat and poses significant challenges to public health, particularly in low- and middle-income countries. This study aims to evaluate the effectiveness and capabilities of Mykrobe, TB-profiler, PhyResSE, KvarQ, and SAM-TB web-based tools for detecting drug resistance in *Mycobacterium tuberculosis* isolates with conventional DST. This evaluation is particularly important in the local context, where rapid diagnosis is critical for controlling multidrug-resistant tuberculosis (MDR-TB).

Material and methods

A total of 88 clinical isolates were included in this study, obtained by convenience sampling from private laboratories and the National Reference of Mycobacteria at the National Institute of Public Health Research "Leopoldo Izquieta Pérez" (INSPI-LIP) in Guayaquil between 2019 and 2021. These isolates were identified as multidrug-resistant through DST agar [28] or the Bactec MGIT 960 System protocol [29] prior to their inclusion in the study. The resistance pattern of first-line and second-line anti-tuberculosis drugs was determined according to the proportional method by the Bactec MGIT 960 System protocol in 67 samples with critical concentrations of drugs as follows: rifampicin, 1.0 μg/mL; isoniazid, 0.1 μg/mL; ethambutol, 5.0 μg/mL; kanamycin, 2.5 μg/mL, capreomycin, 1.0 μg/mL, amikacin, 1.0 μg/mL, 1.0 μg/mL for levofloxacin and 1.0 μg/mL for moxifloxacin. However, in 21 samples, the resistance was determined by DST agar, and the critical concentrations of anti-TB drugs for the DST assays were as follows: rifampicin, 40.0 μg/mL; isoniazid, 0.2 μg/mL; ethambutol, 0.4 μg/mL; streptomycin, 4.0 μg/mL, and 200 μg/mL for pyrazinamide (pyrazinamidase assay). The resistance profile was defined based on the results from all isolates according to WHO recommendations.

DNA extraction and sequencing

Genomic DNA was extracted from 88 isolates of *Mycobacterium tuberculosis* grown in Lowenstein-Jensen medium by two different protocols. In 21 isolates, the DNA was extracted by the CTAB method [30, 31], while for others, 67 were obtained by the PureLink DNA Mini Kit (Fisher Scientific, Pennsylvania, USA) according to the manufacturer's instructions. Each extracted DNA sample was quantified by a Qubit 4.0 fluorometer (Invitrogen, Carlsbad, CA, USA.). DNA samples that fulfilled the quality standards in terms of integrity, purity and quantity were sequenced. Genomic DNA libraries were prepared for whole-genome sequencing using the Tagmentation-based library prep kit according to the manufacturer's instructions (Illumina Inc., San Diego, CA, USA). Next, libraries with different indices were multiplexed and loaded on an Illumina MiniSeq platform with a High Output Reagent Kit, according to the manufacturer's instructions (Illumina, San Diego, CA, USA). Sequencing was carried out using a 2 x 150 paired-end (PE) configuration.

Bioinformatic analyses

Using the Galaxy platform (https://usegalaxy.org/), reads were classified by Kraken version 2 [32] to detect possible contamination or the presence of other mycobacteria. In addition, FastQC version 0.11.9 [33] and Trimmomatic version 0.38 [34] were used to control quality and trim the low-quality ends of the reads, respectively. In particular, a sliding window was

used to trim sequences with an average quality value lower than 20. The high-quality reads were mapped to *M. tuberculosis* strain HR37v (NC_000962.3) using BWA-MEM [35]. To ensure high reliability, only isolates with sequences meeting strict criteria were included in the analysis. Specifically, any isolate with a sequencing depth of less than 20X or a reference coverage of less than 90% was excluded from subsequent analyses.

For phylogenetic analysis, clean reads were used as input in the MTBseq pipeline [36] to obtain i) sublineage classification, ii) transmission group identification and iii) SNP-based alignment for phylogenomic analysis. *Mycobacterium microti* Maus III (accession number: ERR4618952) was used as an outgroup to obtain the rooted tree. The substitution model was calculated with ModelTest-NG v0.1.7 [37]. Phylogenetic reconstruction was performed with RAxML-NG [38] using the maximum likelihood method and a bootstrap cut-off = 0.01. Visualization of the phylogenetic tree was achieved with iTOL v6.6 [39]. The transmission groups were determined by evaluating a distance of 12 SNPs between strains due showing the highest agreement between epidemiological research and genomic data [40, 41] The transmission cluster sizes were classified according to the number of shared isolates as follows: small if it had fewer than three isolates, medium if it was composed of three to five isolates and large if it had more than five isolates.

Predicting susceptibility and drug resistance

The web-based tools TB-Profiler v5.0 (https://github.com/jodyphelan/TBProfiler), PhyReSse v1.0 (The Phylo-Resistance-Search-Engine) (https://bioinf.fz-borstel.de/mchips/phyresse/), Mykrobe v0.12.2 (https://www.mykrobe.com/), KvarQ v0.12.2 (https://kvarq.readthedocs.io/en/latest/index.html) and SAM-TB (https://samtb.uni-medica.com/index) were used to predict canonical mutations in genes related to resistance, such as rifampicin (RIF), isoniazid (INH), pyrazinamide (PZA), and ethambutol (EMB) (first-line drug) and fluoroquinolone (ciprofloxacin, levofloxacin, moxifloxacin, oxifloxacin, FQ), streptomycin (STR), ethionamide (ETH), and aminoglycosides (amikacin, AMK, kanamycin, KM, capreomicin, CAP) (second-line drugs). All programs were run under the default parameters using the high-quality reads from the 88 samples.

Statistical analysis

Global and drug-specific sensitivities, specificities, positive predictive values (PPV) and negative predictive values (NPV) were computed using MedCalc® Statistical Software version 22.013 (MedCalc Software Ltd, Ostend, Belgium; https://www.medcalc.org; 2023) with a 95% confidence interval. For global analysis, we considered the tools capacity for detecting any resistance-related mutations that permit inclusion of the isolate/patient in a surveillance and drug scheme adjustment. The diagnostic results based on DST were used as a reference for the analysis.

Ethics statements

Our study was reviewed and approved by the ethics committee of University Espíritu Santo under code 2022-001A. The research was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. Informed consent of the patients was not required since our study worked with isolates from a collection. The ethics committee determined the exception of informed consent because the isolates were anonymized, and no data on the patients were disclosed. The present work is based on INSPI-LIP permission to use the positive samples.

Results

The drug susceptibility of *M. tuberculosis* isolates obtained from patients between 2019 and 2021 was assessed via drug susceptibility testing (DST) for four first-line drugs (rifampicin, isoniazid, pyrazinamide, and ethambutol), three second-line drugs (streptomycin, levofloxacin, and moxifloxacin), and three injectable drugs (kanamycin, amikacin, and capreomycin). Of the 88 samples, 22 were identified as susceptible, while 66 were found to be drug resistant (DR-TB) and further categorized as MDR (n = 46), polyresistant (n = 13), rifampicin and fluoroquinolone resistant (n = 5), and rifampicin monoresistant (n = 2). Among the isolates, 26.1% were from female patients, with a median age of 39.2 ± 14.5 years (range 21–73 years), and 73.9% were from male patients, with a median age of 40.3 ± 15.0 years (range 11–74 years). With regard to treatment history, 37.5% of the isolates were from treatment-naive cases (45.5% were drug-resistant), 23.9% were from patients receiving treatment (90.5% were drug-resistant), and 38.6% were from previously treated cases (94.1% were drug-resistant). The majority of the drug-resistant isolates (59.1%) were obtained from patients in Guayaquil, one of the most important economic centers in Ecuador, followed by Babahoyo, El Empalme, and Quito (4.5%, 3.4%, and 2.3%, respectively). A small number of cases were detected in Chone, Duran, Guaranda, Machala, and Nueva Loja (1.1% each) (S1 Table).

The MTBseq pipeline was used to assess the phylogenetic relationships among the 88 isolates by analysing 5586 SNPs. The results indicated a high diversity of circulating isolates in Ecuador, with LAM sublineages being the most predominant (44.3%), followed by X-type (23.9%), S-type and Haarlem (11.4%, each), as shown in Fig 1.

Although no association was found between sublineages and drug sensitivity, it is worth noting that out of the 10 Haarlem strains, 7 (70.0%) displayed MDR resistance, while three were sensitive. Cluster analysis identified 54 isolates (61.4%) grouped into 16 potentially molecular-related transmission clusters (with at least two isolates and $\leq$12 SNP differences), while 34 isolates were singletons (S1 Table, Fig 1). Seven clusters were considered small (< 3 isolates), eight medium (3–5 isolates), with the most prevalent being clusters 13 and 16 (five isolates each), and one large cluster with nine isolates (group 01). Pre-XDR isolates were distributed in four clusters, with three isolates in cluster 10 and two isolates each in clusters 11 and 16. The remaining pre-XDR isolates did not form clusters, and one isolate was unique. Notably, the only large cluster consisting of nine isolates was entirely composed of the S-type sublineage (S1 Table, Fig 1).

The performance and functionalities of five web-based tools (TB-Profiler v5.0, PhyResSE v1.0, Mykrobe v0.12.2, KvarQ v0.12.2, and SAM-TB) were evaluated by comparing the phenotypic resistance profile of the 88 strains (S2 Table). Descriptive statistics such as sensitivity, specificity, accuracy, percentage of agreement and kappa coefficient were calculated and are presented in Table 1.

All the evaluated programs exhibited good sensitivity and specificity globally, but their performance varied for preconditions of resistance to specific drugs. The highest level of agreement was observed for isoniazid on all programs (agreement and kappa coefficient above 93% and k = 0,840), while KvarQ had the lowest agreement for rifampicin (agreement 90.91% and k = 0,784).

In terms of detecting specific resistances, all tools showed good detection values for isoniazid and rifampicin. However, the detection of ethambutol and pyrazinamide varied across the different tools. TB-profiler and SAM-TB demonstrated adequate detection values, while KvarQ and PhyResSE exhibited lower quality values. In summary, TB-profiler and SAM-TB were found to be the most effective in detecting resistance to first-line antibiotics, whereas KvarQ was the least effective. For streptomycin, Mykrobe and KvarQ showed fair agreement

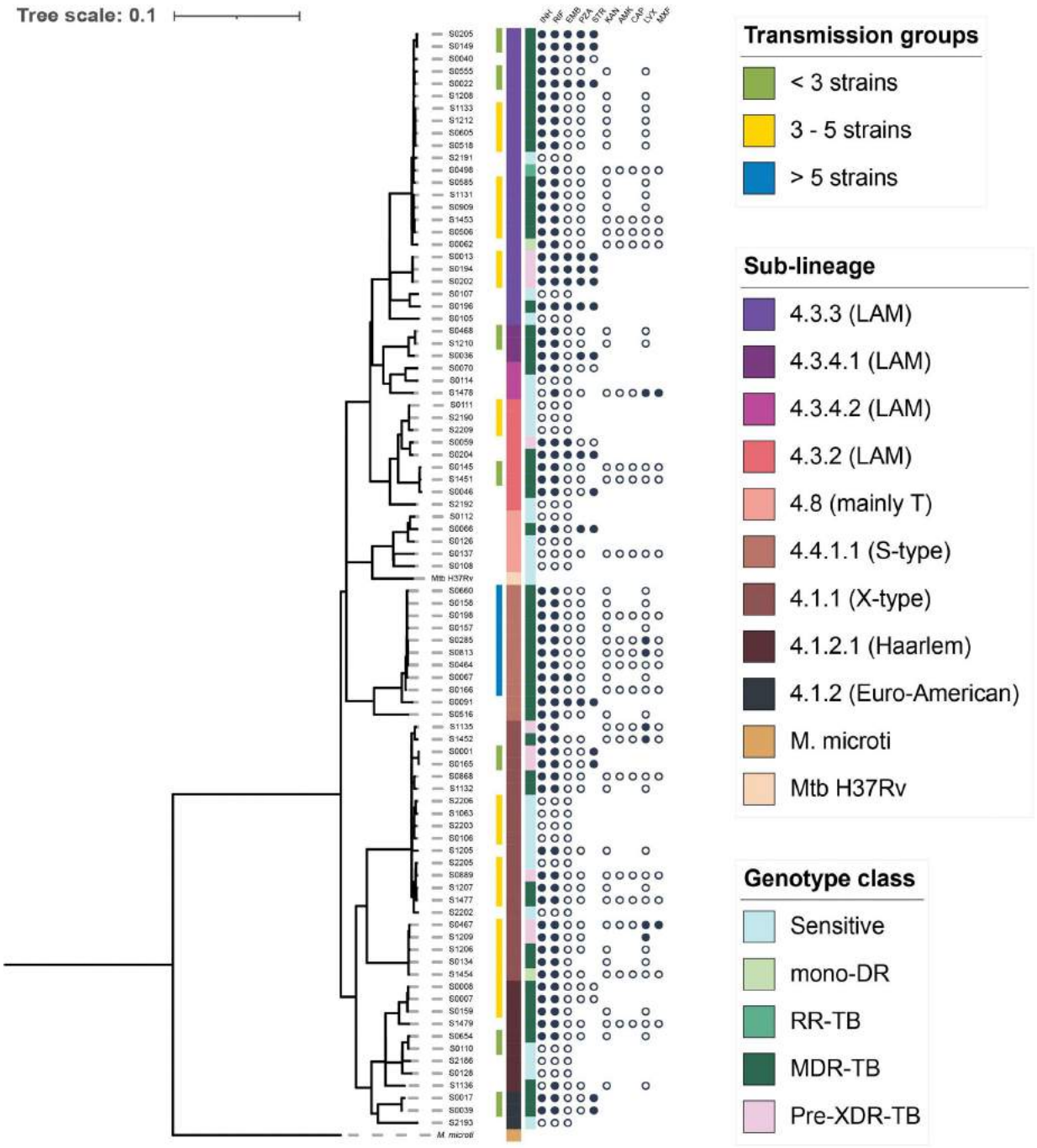

Fig 1. Phylogenetic reconstruction of the 88 Mtb isolates from Ecuador. A total of 5,586 SNPs were used to reconstruct the phylogenetic tree using the maximum likelihood method, the GTR substitution model and a bootstrap cut-off of 0.01. Metadata include i) transmission groups (lines green, yellow, and blue), ii) sublineage classification, iii) global genotypic class of drug resistance and iv) DST results of first- and second-line antibiotics (full circle = resistant; empty circle = sensitive; missing circle = not tested).

https://doi.org/10.1371/journal.pone.0294670.g001

Table 1. Comparative statistical analysis of five software programs for predicting anti-TB drug resistance.

Statistic	Mykrobe	KvarQ	TB-Profiler	PhyResSE	SAM-TB
Sensitivity	0.97 (0.89 to 1.00)	0.92 (0.83 to 0.97)	0.97 (0.89 to 1.00)	0.97 (0.89 to 1.00)	0.97 (0.89 to 1.00)
Specificity	1.00 (0.85 to 1.00)	1.00 (0.85 to 1.00)	1.00 (0.85 to 1.00)	1.00 (0.85 to 1.00)	1.00 (0.85 to 1.00)
Positive Likelihood Ratio					
Negative Likelihood Ratio	0.03 (0.01 to 0.12)	0.08 (0.03 to 0.18)	0.03 (0.01 to 0.12)	0.03 (0.01 to 0.12)	0.03 (0.01 to 0.12)
Positive Predictive Value	1.00 (0.94 to 1.00)	1.00 (0.94 to 1.00)	1.00 (0.94 to 1.00)	1.00 (0.94 to 1.00)	1.00 (0.94 to 1.00)
Negative Predictive Value	0.92 (0.74 to 0.98)	0.82 (0.65 to 0.91)	0.92 (0.74 to 0.98)	0.92 (0.74 to 0.98)	0.92 (0.74 to 0.98)
Accuracy (*)	0.98 (0.92 to 1.00)	0.94 (0.87 to 0.98)	0.98 (0.92 to 1.00)	0.98 (0.92 to 1.00)	0.98 (0.92 to 1.00)
% of agreement	0.98	0.94	0.98	0.98	0.98
Cohen's kappa	0.941	0.859	0.941	0.941	0.941
	Almost perfect agreement	Substantial agreement	Almost perfect agreement	Almost perfect agreement	Almost perfect agreement

Numbers in brackets represent the corresponding 95% confidence intervals. Cohen's kappa: 0.01–0.20: slight agreement; 0.21–0.40: fair agreement; 0.41–0.60: moderate agreement; 0.61–0.80: substantial agreement; 0.81–1.00: almost perfect or perfect agreement

https://doi.org/10.1371/journal.pone.0294670.t001

in detecting resistance. However, all tools had difficulties detecting resistance to fluoroquinolones, particularly levofloxacin, and only KvarQ and SAM-TB presented the best agreement, while for moxifloxacin, only KvarQ showed moderate agreement. The statistical parameters for kanamycin, amikacin, and capreomycin could not be determined because only sensitive isolates were identified phenotypically (Table 2).

By genomic analysis of 88 phenotypic isolates, we identified genotypically 59.1% of strains MDR (52/88), 11.36% pre-XDR, 1.13% Monoresistant (1/88 each for HR and RR), and 27.3% Sensitive (24/88). Diversity in the genotypic resistance profile from the isolates was observed among the drugs assessed. Mykrobe, PhyResSE and SAM-TB detected the highest proportion of resistance to isoniazid, with 61 out of 88 isolates (69.3%). Similarly, PhyResSE and SAM-TB detected 64 out 88 isolates (72.7%) for rifampicin, while for ethambutol with 22 out 87 isolates (25.3%) and pyrazinamide with 20 out 64 isolates (31.3%), TB-profiler and SAM-TB were the best; however, the lowest resistance proportions were mainly described by KvarQ.

It is noteworthy that all programs detected resistance to second-line drugs that were not detected by DST. This may be associated with the slower growth of some of the isolates at the limit of the DST detection range. In particular, all programs detected a range of 2 to 5 isolates with mutations associated with resistance to kanamycin, amikacin and capreomycin, none of which were previously detected by DST. In terms of fluoroquinolones, between 10 and 12 isolates were characterized as resistant to moxifloxacin and levofloxacin by the programs, while DST detected only 2 and 7 resistant isolates, respectively. Overall, KvarQ detected less resistance, while SAM-TB was the program that detected more resistance, followed by PhyResSE and TB-Profiler (Table 3).

Polymorphisms associated with drug resistance identified by whole-genome sequencing and web-based tools

Whole-genome sequencing and web-based tool analysis of the 88 isolates revealed a total of 60 mutations distributed among 4 promoter regions, 5 intergenic regions, and 51 coding regions leading to changes in the reading frame. The majority of the mutations were single nucleotide changes, although 8 insertions and 4 deletions were also detected. Notably, three of the deletions were of considerable length, with one at 193 nt in the gid gene and two affecting the *pncA* and Rv2041c-Rv2042c genes at 891 and 833 nt, respectively. The *pncA* and *rpoB* genes associated with resistance to pyrazinamide and rifampicin, respectively, had the highest

Table 2. Sensitivity, specificity and accuracy of ten drugs of anti-TB drug resistance analysed by WGS plus web-based tools compared with DST.

Mykrobe	Isoniazid (n = 87)	Rifampicin (n = 88)	Ethambutol (n = 87)	Pyrazinamide (n = 64)	Streptomycin (n = 21)	Kanamycin (n = 45)	Amikacin (n = 21)	Capreomycin (n = 21)	Levofloxacin (n = 46)	Moxifloxacin (n = 21)
Sensitivity	95.2 (86.7 to 99.0)	95.5 (87.3 to 99.1)	100.0 (71.5 to 100.0)	83.3 (51.6 to 97.9)	56.3 (29.9 to 82.3)				42.9 (9.9 to 81.6)	50.0 (1.3 to 98.7)
Specificity	96.0 (79.7 to 99.9)	100.0 (84.6 to 100)	86.8 (77.1 to 93.5)	94.2 (84.1 to 98.8)	100.0 (47.8 to 100)	95.6 (84.9 to 99.5)	95.2 (76.2 to 99.9)	95.2 (76.2 to 99.9)	97.4 (86.5 to 99.9)	89.5 (66.9 to 98.7)
Accuracy	95.5 (88.8 to 98.8)	96.6 (90.4 to 99.3)	88.5 (79.8 to 94.4)	92.2 (82.7 to 97.4)	66.7 (43.0 to 85.4)				89.1 (76.4 to 96.4)	85.7 (63.7 to 97.0)
Cohen's kappa	0.891	0.913	0.625	0.751	0.380				0.489	0.323
TB-Profiler										
Sensitivity	93.7 (84.5 to 98.2)	100.0 (94.2 to 100)	100.0 (71.5 to 100)	83.3 (51.6 to 97.9)	68.8 (41.3 to 89.0)				42.9 (9.9 to 81.6)	50.0 (1.3 to 98.7)
Specificity	96.0 (79.7 to 99.9)	84.6 (65.1 to 95.6)	85.5 (72.5 to 90.6)	86.5 (74.2 to 94.4)	100.0 (47.8 to 100)	95.6 (84.9 to 99.5)	95.2 (76.2 to 99.9)	95.2 (76.2 to 99.9)	97.4 (86.5 to 99.9)	89.5 (66.9 to 98.7)
Accuracy	94.3 (87.2 to 98.1)	95.5 (88.8 to 98.8)	87.4 (75.8 to 91.8)	85.9 (75.0 to 93.4)	76.2 (52.8 to 91.8)				89.1 (76.4 to 96.4)	85.7 (63.7 to 97.0)
Cohen's kappa	0.865	0.886	0.599	0.602	0.511				0.489	0.323
PhyResSE										
Sensitivity	95.2 (86.7 to 99.0)	97.0 (89.5 to 99.6)	90.9 (58.7 to 99.8)	75.0 (42.8 to 94.5)	68.8 (41.3 to 89.0)				42.9 (9.9 to 81.6)	50.0 (1.3 to 98.7)
Specificity	96.0 (79.7 to 99.9)	100.0 (84.6 to 100)	85.5 (75.6 to 92.6)	86.5 (74.2 to 94.4)	100.0 (47.8 to 100)	95.6 (84.9 to 99.5)	95.2 (76.2 to 99.9)	95.2 (76.2 to 99.9)	97.4 (86.5 to 99.9)	89.5 (66.9 to 98.7)
Accuracy	95.5 (88.8 to 98.8)	97.7 (92.0 to 99.7)	86.2 (77.2 to 92.7)	84.4 (73.1 to 92.2)	76.2 (52.8 to 91.8)				89.1 (76.4 to 96.4)	85.7 (63.7 to 97.0)
Cohen's kappa	0.891	0.941	0.523	0.545	0.511				0.489	0.323
KvarQ										
Sensitivity	92.1 (82.4 to 97.4)	87.9 (77.5 to 94.6)	72.7 (39.0 to 94.0)	66.7 (34.9 to 90.1)	56.3 (29.9 to 82.3)				42.9 (9.9 to 81.6)	50.0 (1.3 to 98.7)
Specificity	96.0 (79.7 to 99.9)	100.0 (84.6 to 100)	93.4 (85.3 to 97.8)	88.5 (76.6 to 95.7)	100.0 (47.8 to 100)	100.0 (92.1 to 100)	100.0 (83.9 to 100)	100.0 (83.9 to 100)	100.0 (91.0 to 100)	94.7 (74.0 to 99.9)
Accuracy	93.2 (85.8 to 97.5)	90.9 (82.9 to 96.0)	90.8 (82.7 to 96.0)	84.4 (73.1 to 92.2)	66.7 (43.0 to 85.4)				91.3 (79.2 to 97.6)	90.5 (69.6 to 98.8)
Cohen's kappa	0.840	0.784	0.614	0.518	0.380				0.56	0.447
SAM-TB										
Sensitivity	95.2 (86.7 to 99.0)	97.0 (89.5 to 99.6)	90.9 (58.7 to 99.8)	100.0 (73.5 to 100)	75.0 (47.6 to 92.7)				71.4 (29.0 to 96.3)	50.0 (1.3 to 98.7)
Specificity	96.0 (79.7 to 99.9)	100.0 (84.6 to 100)	85.5 (75.6 to 92.6)	84.6 (71.9 to 93.1)	100.0 (47.8 to 100)	100.0 (92.1 to 100)	95.2 (76.2 to 99.9)	95.2 (76.2 to 99.9)	97.4 (86.5 to 99.9)	79.0 (54.4 to 94.0)
Accuracy	95.5 (88.8 to 98.8)	97.7 (92.0 to 99.7)	86.2 (77.2 to 92.7)	87.5 (76.9 to 94.5)	81.0 (58.1 to 94.6)				93.5 (82.1 to 98.6)	76.2 (52.8 to 91.8)
Cohen's kappa	0.891	0.941	0.523	0.674	0.588				0.732	0.173

Cohen's kappa: 0.01–0.20: slight agreement; 0.21–0.40: fair agreement; 0.41–0.60: moderate agreement; 0.61–0.80: substantial agreement; 0.81–1.00: almost perfect or perfect agreement.

https://doi.org/10.1371/journal.pone.0294670.t002

number of mutations, with *pncA* showing 13 mutations, two insertions of 4 nt each, and one deletion of 833 nt, while *rpoB* had 14 nonsynonymous mutations, some of which occurred in the same codon. Position 445 of the *rpoB* gene was the most variable codon, with five different configurations, followed by position 406 of the *embB* gene with three different mutations.

Table 3. Comparison of the number of cases of resistance detection by phenotypic (DST) and genotypic web-based tools for ten TB anti-drugs.

Treatment drugs	DST n (%)	Mykrobe n (%)	TB-profiler n (%)	PhyReSse n (%)	KvarQ n (%)	SAM-TB n (%)
Isoniazid	63 (71.6)	61 (69.3)	60 (68.2)	61 (69.3)	59 (67.1)	61 (69.3)
Rifampicin	66 (75.0)	63 (71.6)	62 (70.5)	64 (72.7)	58 (65.9)	64 (72.7)
Ethambutol	11 (12.5)	21 (23.9)	22 (25.0)	21 (23.9)	13 (14.7)	21 (23.9)
Pyrazinamide	12 (13.6)	14 (15.9)	18 (20.5)	18 (20.5)	15 (17.0)	21 (23.9)
Streptomycin	16 (18.2)	21 (23.9)	23 (26.1)	23 (26.1)	19 (21.6)	24 (27.3)
Kanamycin	0 (0.0)	5 (5.7)	5 (5.7)	5 (5.7)	2 (2.3)	5 (5.7)
Amikacin	0 (0.0)	3 (3.4)	3 (3.4)	3 (3.4)	2 (2.3)	2 (2.3)
Capreomycin	0 (0.0)	3 (3.4)	3 (3.4)	3 (3.4)	2 (2.3)	3 (3.4)
Levofloxacin	7 (8.0)	11 (12.5)	10 (11.4)	11 (12.5)	10 (11.4)	12 (13.6)
Moxifloxacin	2 (2.3)	11 (12.5)	10 (11.4)	11 (12.5)	10 (11.4)	12 (13.6)

https://doi.org/10.1371/journal.pone.0294670.t003

Additionally, the *embB*, *gyrA*, and *rrs* genes, which confer resistance to ethambutol, fluoroquinolones, and streptomycin, respectively, had at least 4 mutations (Table 4, S3 Table).

Polymorphisms conferring rifampicin resistance

Mutations in the *rpoB*, *rpoC*, and *rpoA* genes were frequently identified in the 66 isolates defined as phenotypically resistant to rifampicin through whole-genome sequencing. The S450L *rpoB* mutation was the most prevalent (44/64), followed by D435V (6/64), V170F (3/64), D435Y, H445C/R/Y, and L430P (1/64 each). One isolate showing the V170F and L430P mutations and three isolates with the L430P and H445Q mutations were also identified. TB-profiler, PhyResSE and SAM-TB were able to identify mutations in one isolate (S0555_Mtb_Ec) with H445P and K446Q, while SAM-TB was able to identify in one isolate (S0506_Mtb_Ec) the mutations S450L plus K274N and S450L plus P45S (S0196_Mtb_Ec), while the others only detected S450L. KvarQ was able to identify isolates with the compensatory mutations W484G (S0001_Mtb_Ec and S0165_Mtb_Ec) and I491V (S1131_Mtb_Ec) in *rpoC* and the T187A (S0091_Mtb_Ec) mutation in the *rpoA* gene, while PhyResSE detected the compensatory mutation I491V (S0204_Mtb_Ec) in the *rpoB* gene. None of the sensitive isolates had any characterized mutations.

Polymorphisms conferring isoniazid and ethionamide resistance

Out of 63 clinical isolates where isoniazid resistance was reported by phenotypic susceptibility testing, at least one known isoniazid-resistance mutation was found in 61. For isolates S1135_Mtb_Ec and S1205_Mtb_Ec with DST-detected resistance, the 426C>T mutation was found, but it was not previously reported for isoniazid. One isolate reported as sensitive by DST, S1136_Mtb_Ec, had one resistance-related mutation. The most prevalent mutation related to isoniazid resistance was S315T in the *katG* gene (55/63), particularly S315N, which was detected in only one isolate (S0070_Mtb_Ec). Five isolates presented the mutation -15C>T in the *fabG1*/*inhA* promoter gene related to isoniazid/ethionamide resistance. Other ethionamide resistance mutations were reported by TB-profiler and SAM-TB in the *ethR* gene (A95T and F110L mutations) and the *ethA* gene (the 1222delT deletion). For three isolates with DST-detected isoniazid resistance, no known mutations related to this resistance were found.

Polymorphisms conferring ethambutol resistance

The analysis of the studied isolates revealed the presence of multiple mutations in the *embB* and *embA* genes, including M306I/V, D354A, G406A/D/S, and a -16C>T substitution. M306I

Table 4. Mutational changes in resistant isolates of *Mycobacterium tuberculosis* identified by WGS.

RIF		INH		EMB		PZA		STR		KAM		FQ		ETH	
rpoA	T187A	***katG***	S315T	***embA***	-16C>T	***pncA***	-10A>G	***rpsL***	K43R	***eis***	-12C>T	***gyrA***	D89G	***ethA***	1222delT
rpoB	P45S		S315N	***embB***	M306I		V7A	***rrs***	514A>C	***rrs***	1401A>G		A90V	***ethR***	A95T
	V170F				M306V		D8H		517C>T				D94A		F110L
	K274N				D354A		H43P		888G>A				D94N	***inhA/ fabG1****	-15C>T
	T400A						A46V		906A>G			***gyrB***	N499D		
	L430P				G406A		D49A	***gid***	102delG						
	D435V D435Y	***inhA/ fabG1****	-15C>T		G406D		H51R								
	H445C				G406S		P69L		148_340del193						
	H445P						H82R								
	H445Q						S104G								
	H445R						G108E								
	H445Y						V125F H137Y								
	K446Q						M175I								
	S450L I491V						254insGGGC								
rpoC	W484G						292insCCCG								
	I491V						503-*833del								
						***pncA*, Rv2041c-Rv2042c**	2287848_2288738del891								

**inhA* promoter mutations include mutations in the *fabG1* open reading frame (ORF) because they create alternative promoters for *inhA* and mutations upstream of *fabG1* because they act as promoters of the entire operon, which includes *inhA*. Abbreviations: rifampicin (RIF), isoniazid (INH), pyrazinamide (PZA), ethambutol (EMB) (first-line drug) and fluoroquinolone (ciprofloxacin, levofloxacin, moxifloxacin, oxifloxacin, FQ), streptomycin (STR), ethionamide (ETH), and aminoglycosides (amikacin, AMK, kanamycin, KM, capreomicin, CAP) (second-line drugs).

https://doi.org/10.1371/journal.pone.0294670.t004

was the most frequently observed mutation, found in 11 out of 22 isolates, often in combination with G406D and D354A. Additionally, M306V and G406S were each observed in four isolates. It is noteworthy that mutations in the *embB* and *embA* genes were identified in 11 initially ethambutol-sensitive isolates as determined by DST.

Polymorphisms conferring pyrazinamide resistance

A diverse range of mutations in the *pncA* gene associated with pyrazinamide resistance was observed, including 8 insertions, 1 deletion, and 14 single nucleotide polymorphisms that spanned the entire *pncA* gene, including its promoter region. The most frequently occurring mutations were H82R and V125F (3/21), followed by H51R and G108E (2/21), along with some unique mutations, including V7A, D8H, H43P, A46V, D49A, P69L, S104G, H137Y, and M175I, and a substitution of -10A>G in the promoter region. Two isolates had larger deletions, one spanning 891 nucleotides encompassing the *pncA* and Rv2041c-Rv2042c genes and the other having an 833-nucleotide deletion within the *pncA* gene (503-*del833). Furthermore, specific insertions were identified in two isolates: 292GC, 293CC, 294CC, and 295GG in S0036_Mtb_Ec and 254TG, 255GG, 256GG, and 257CC in S0040_Mtb_Ec. Remarkably, mutations in the *pncA* gene were identified in 8 isolates initially characterized as pyrazinamide-sensitive by DST. Additionally, mutations in the rpsA gene were detected in 2 isolates, but these were not associated with pyrazinamide resistance. In isolate S2193_Mtb_Ec, only PhyResSE detected the H137Y mutation in pncA, however this is not thought to be associated with resistance.

Polymorphisms conferring fluoroquinolone resistance

Thirteen isolates were found to have mutations associated with resistance to FQ. The most common mutation, present in six out 13 isolates, was A90V, followed by D94N (three isolates) and D89G (one isolate), all located in the *gyrA* gene. One DST-sensitive isolate had a combination of two mutations, D94A in the *gyrA* gene and N499D in *gyrB*. However, one isolate that was identified as resistant to levofloxacin and moxifloxacin by DST did not have any resistance-related mutations. Two isolates resistant to levofloxacin but sensitive to moxifloxacin were detected to have resistance-related mutations only by SAM-TB. No mutations related to FQ resistance were found in the majority of the isolates identified as sensitive by DST.

Polymorphisms conferring aminoglycoside resistance

Among the 28 isolates identified as resistant to streptomycin, the K43R mutation in the *rpsL* gene was the most common, found in 16 isolates, while the K88R mutation was the least frequent, present in only one isolate. Mutations in the *rrs* gene were also observed, including 514A>C (3/28), 517C>T, 888G>A, and 906A>G. In addition, deletions in the gid gene were identified, including 102delG and 148_340del193. Interestingly, even isolates classified as sensitive to kanamycin, amikacin, and capreomycin by DST had the known resistance-associated substitution 1401A>G in the *rrs* gene in three isolates and two isolates exhibiting the -12C>T mutation in the *eis* gene, which is associated with kanamycin resistance.

Discussion

Although sequencing-based diagnostic information for *M. tuberculosis* has been available for several years and recent methodological advancements have enabled the use of direct whole-genome sequencing for accurate prediction of drug resistance, this is the first study to assess and compare the effectiveness of whole-genome sequencing and user-friendly web-based tools

in inferring drug resistance in tuberculosis drug susceptibility surveillance in Ecuador. The study utilized isolates obtained from nine cities, with Guayaquil being the most representative city, accounting for 59.1% of the isolates. This high number of tuberculosis cases detected in Guayaquil may be attributed to the high mobility of people from other provinces for trade or work, in addition to the presence of leading health centers for monitoring this pathogen.

Accelerating diagnosis and administering appropriate treatment based on the rapid identification of resistant strains and transmission clusters are crucial in reducing the incidence of tuberculosis [4]. In our study, whole-genome sequencing detected 97% of phenotypically resistant isolates accurately and identified mutations associated with resistance in 18.2% of phenotypically sensitive isolates within a week. However, discordant results (2.3%) were observed, likely due to factors such as critical concentration in some DST systems, mutations outside the drug resistance regions, silent mutations, and heteroresistance, which are missed by conventional methods [42–44].

Whole-genome sequencing (WGS) has been evaluated as a useful tool to quantify transmission clusters, with a threshold of 12 SNPs. In our study, using 12 SNPs, we identified sixteen mostly medium-sized transmission clusters (3–5 isolates), indicating local aggregations, and showed a high transmission rate (54/88, 61.4%). This is in contrast to previous studies where a low transmission rate was reported using MIRU-VNTR [45, 46] or using WGS in other localities [47, 48]. This could be due to a large percentage of cluster of two isolates (43.8%) in our results or the highest SNPs umbral used in other studies (15 SNPs).

WGS is a powerful tool that can simultaneously differentiate *M. tuberculosis* species, detect drug resistance, identify genetic diversity, and track transmission dynamics [49–51]. However, its use in clinical practice is still limited due to challenges in analysing genomic sequencing data, including the lack of standardized criteria for identifying noncanonical mutations and selecting diagnostic criteria for resistance [22]. The WHO has updated its TB treatment guidelines, including related mutations for drug-susceptible, single drug-resistant, MDR-TB, and XDR-TB, to improve web-based tools [23]. Despite several TB-specific genome browsers and WGS analysis tools being available, their use in surveillance programs is limited in countries such as Ecuador due to inadequate access to computational infrastructure, funding, and technical expertise. Clinicians and public health officials may also lack specialized knowledge to interpret genomic data, further impeding their adoption. Thus, promoting the use of these tools and improving their accessibility and usability for routine clinical and public health purposes is essential.

To evaluate their usefulness, we compared the accuracy, sensitivity, and specificity of five tools for inferring the resistance of *M. tuberculosis* isolates against phenotypic tests. Our results showed that globally, TB-profiler, PhyResSE, Mykrobe, and SAM-TB have a better ability to identify any resistance pattern in *M. tuberculosis* isolates, with an overall sensitivity, specificity, and accuracy of 97.0%, 100.0% and 97.7%, respectively. Notably, SAM-TB, the most recently developed tool, combines features of other pipelines and improves the detection and interpretation of resistance, similar to a panel that shows variations in isolates analysed, including statistical data, quality control, drug resistance, phylogenetic tree reconstruction and pairwise SNP distance into isolates and references. Although these web-based tools demonstrated good overall performance in detecting drug resistance in *Mycobacterium tuberculosis* isolates with conventional DST, they also revealed limitations in the predictive power of some mutations in their catalogue. In addition, the tools showed varying levels of sensitivity and specificity due to differences in the sets of mutations used and their underlying genotyping pipelines. Notably, this would indicate that the predictive power of these tools should be considered with varying degrees of confidence depending on the predictive performance of the detected markers [15,

52]; whence, finding a solution to maintain and standardized update the automated pipelines to WHO guidelines is crucial to ensure their continued effectiveness.

Drug resistance in *M. tuberculosis* is primarily mediated by mutations in specific gene targets. Therefore, identifying the responsible or strongly associated SNPs is crucial for effective diagnosis and treatment [8, 25, 53, 54]. In our study, we identified 17 resistance-associated genes out of 42 candidate genes, which comprised a total of 69 mutations. Canonical mutations associated with resistance have been previously identified in various studies, along with deletions that can affect gene functionality or promote bacterial transmission [51, 55–63].

In first-line drugs, numerous studies have demonstrated that the rifampicin resistance-determining region (RRDR) in the *rpoB* gene accurately predicts rifampicin resistance [56, 64, 65]. In our study, we discovered that the S450L mutation was the most significant in predicting resistance and linked to high levels of resistance [66, 67]. Additionally, compensatory mutations in the *rpoC* gene (W848G and I491V) and the *rpoA* gene (T187A) were associated with multidrug-resistant (MDR) transmission [3, 68–71]. These mutations may contribute to the faster emergence of MDR strains and highlight the importance of monitoring these genetic changes to prevent the spread of drug resistance. Likewise, mutations in various genes have been associated with conferring resistance to isoniazid [55, 72, 73] and ethionamide [59, 74, 75]. Our study found a higher prevalence of mutations in the *katG* gene (88.8%), which is consistent with prior research [55, 73, 76]. On the other hand, mutations in the *embCAB* gene operon are frequently involved in ethambutol resistance [57, 77–80]. We found that a higher number of mutations occurred in the *embB* gene, which is consistent with previous studies. Additionally, all the mutations we identified as related to pyrazinamide resistance coincided with studies pointing to the importance of the promoter and the *pncA* gene [4, 57, 62, 81, 82]. Our study confirms the importance of monitoring mutations in the RRDR of the *rpoB* gene and the prevalence of mutations in the *katG*, *embB* and *pncA* genes for predicting isoniazid, ethambutol and pyrazinamide resistance, respectively. We also highlight the significance of compensatory mutations and the need to monitor them to prevent the emergence and spread of MDR strains.

For second-line drugs, our analysis showed that the majority (92.3%) of mutations associated with fluoroquinolone resistance were located specifically in the quinolone resistance-determining region (QRDR), consistent with previous reports linking mutations in these regions to cross-resistance to all FQs [74, 83–85]. Interestingly, one resistant isolate did not show any mutations in the *gyrA* or *gyrB* genes, suggesting the involvement of other mechanisms, such as efflux pumps, in generating fluoroquinolone resistance [86]. Similarly, 60.7% of mutations observed in the *rpsL* gene were the k43R mutation, which has been reported as the most recurrent streptomycin resistance-associated mutation [51, 85, 87]. In our study, we analysed the *rrs* gene and identified several substitutions associated with streptomycin (positions 514, 517, 888, and 906) or linked to cross-resistance to kanamycin, capreomycin, or amikacin (1401A>G). It is worth noting that this mutation has been found to confer high levels of resistance in previous studies [8, 72, 88–92]. Moreover, our study identified two isolates with the -12C>T mutation in the *eis* promoter that have been associated with low-level kanamycin resistance, supporting previous research findings [93, 94]. Therefore, our findings support the importance of monitoring this mutation and other known substitutions in the *rrs* gene to detect and prevent the spread of second-line drug-resistant TB strains.

Our study also identified genetic alterations that may impact drug resistance. Among these alterations were large deletions in the *pncA* and gid genes, which could have implications for drug susceptibility. Furthermore, we detected nonresistance-conferring mutations in the *katG* (R463L, L427P), *kasA* (G269S), *ahpC* (51G>A), *pncA* (H137Y) and *gyrA* (E21Q, S95T,

G668D) genes. While these mutations do not confer drug resistance, they may still be useful for characterizing the evolutionary history of *M. tuberculosis* [8, 95].

Our study faced limitations due to the COVID-19 pandemic, limited access to samples and epidemiological information, lack of access to phenotypic tests for all drugs and a small sample size. Due to the absence of complete susceptibility results for all drugs tested by DST, our study could lead to underestimation of the true magnitude of the resistance, especially for second-line drugs, as well as determine the real performance of tools used for detecting second-line drug resistance and loss of ability to determine rare resistance variants with low frequencies associated with intrinsic and extrinsic resistance. Future studies with larger sample sizes and broader coverage should be conducted to identify mutations in tuberculosis patients in other regions of Ecuador. The study suggests to using TB-specific genome browsers and WGS analysis tools such as SAM-TB to aid in genotyping, drug resistance, and transmission cluster identification. These tools should be implemented more routinely in TB surveillance programs, particularly in local conditions. The study highlights the potential of WGS technology and bioinformatic tools to improve TB diagnosis and treatment and ultimately reduce the global burden of TB.

Supporting information

S1 Table. Sociodemographic, clinical, and genotypic results in 88 isolates of *Mycobacterium tuberculosis*.
(XLSX)

S2 Table. Individual phenotypic and genotypic resistance pattern in 88 isolate of *Mycobacterium tuberculosis*. S: Sensitive, R: Resistance, ND: Not determined.
(XLSX)

S3 Table. Mutation profile from 88 isolates of *Mycobacterium tuberculosis* by five web-based tools.
(XLSX)

Acknowledgments

We acknowledge the private laboratories and National Reference of Micobacteria of INSPI-LIP for supporting the isolate analysed in this study. We also thank the Omics Lab of Universidad Espiritu Santo for permitting the use of the MiniSeq sequencing machine employed to run the samples. We also thank PEDECIBA and the Bioinformatics Unit of the Institute Pasteur de Montevideo for their assistance and support.

Author Contributions

Conceptualization: Gabriel Morey-León, Cuauhtémoc Licona-Cassani, Luisa Berná.

Data curation: Paulina M. Mejía-Ponce, Juan Carlos Granda Pardo.

Formal analysis: Gabriel Morey-León, Paulina M. Mejía-Ponce, Juan Carlos Granda Pardo, Karen Muñoz-Mawyin, Juan Carlos Fernández-Cadena, Derly Andrade-Molina.

Funding acquisition: Gabriel Morey-León.

Investigation: Gabriel Morey-León.

Methodology: Karen Muñoz-Mawyin, Juan Carlos Fernández-Cadena, Derly Andrade-Molina.

Resources: Juan Carlos Granda Pardo.

Validation: Evelyn García-Moreira.

Writing – original draft: Gabriel Morey-León, Luisa Berná.

Writing – review & editing: Gabriel Morey-León, Paulina M. Mejía-Ponce, Juan Carlos Granda Pardo, Karen Muñoz-Mawyin, Juan Carlos Fernández-Cadena, Derly Andrade-Molina, Luisa Berná.

References

1. World Health Organization. Global Tuberculosis Report, 2022 [Internet]. 2022 [cited 2022 Nov 6]. Available from: https://www.who.int/teams/global-tuberculosis-programme/tb-reports
2. Singh R, Dwivedi SP, Gaharwar US, Meena R, Rajamani P, Prasad T. Recent updates on drug resistance in Mycobacterium tuberculosis. J Appl Microbiol. 2020 Jun; 128(6):1547–67. https://doi.org/10.1111/jam.14478 PMID: 31595643
3. Alame Emane AK, Guo X, Takiff HE, Liu S. Drug resistance, fitness and compensatory mutations in Mycobacterium tuberculosis. Tuberculosis. 2021 Jul 1; 129:102091. https://doi.org/10.1016/j.tube.2021.102091 PMID: 34090078
4. Soetaert K, Ceyssens PJ, Boarbi S, Bogaerts B, Delcourt T, Vanneste K, et al. Retrospective evaluation of routine whole genome sequencing of Mycobacterium tuberculosis at the Belgian National Reference Center, 2019. Acta Clin Belg. 2022 Sep 3; 77(5):853–60. https://doi.org/10.1080/17843286.2021.1999588 PMID: 34751641
5. Bogaerts B, Delcourt T, Soetaert K, Boarbi S, Ceyssens PJ, Winand R, et al. A Bioinformatics Whole-Genome Sequencing Workflow for Clinical Mycobacterium tuberculosis Complex Isolate Analysis, Validated Using a Reference Collection Extensively Characterized with Conventional Methods and In Silico Approaches. J Clin Microbiol. 2021 Mar 31; 59(6):e00202–21. https://doi.org/10.1128/JCM.00202-21 PMID: 33789960
6. Maladan Y, Krismawati H, Wahyuni T, Tanjung R, Awaludin K, Audah KA, et al. The whole-genome sequencing in predicting Mycobacterium tuberculosis drug susceptibility and resistance in Papua, Indonesia. BMC Genomics. 2021 Nov 22; 22:844.
7. Lam C, Martinez E, Crighton T, Furlong C, Donnan E, Marais BJ, et al. Value of routine whole genome sequencing for Mycobacterium tuberculosis drug resistance detection. Int J Infect Dis. 2021 Dec; 113: S48–54. https://doi.org/10.1016/j.ijid.2021.03.033 PMID: 33753222
8. Wang L, Yang J, Chen L, Wang W, Yu F, Xiong H. Whole-genome sequencing of *Mycobacterium tuberculosis* for prediction of drug resistance. Epidemiol Infect. 2022; 150:e22.
9. Brown AC. Whole-Genome Sequencing of Mycobacterium tuberculosis Directly from Sputum Samples. In: Parish T, Kumar A, editors. Mycobacteria Protocols [Internet]. New York, NY: Springer US; 2021 [cited 2022 Jun 18]. p. 459–80. (Methods in Molecular Biology). Available from: https://doi.org/10.1007/978-1-0716-1460-0_20 PMID: 34235666
10. Torres Ortiz A, Coronel J, Vidal JR, Bonilla C, Moore DAJ, Gilman RH, et al. Genomic signatures of pre-resistance in Mycobacterium tuberculosis. Nat Commun. 2021 Dec; 12(1):7312. https://doi.org/10.1038/s41467-021-27616-7 PMID: 34911948
11. Freschi L, Vargas R, Husain A, Kamal SMM, Skrahina A, Tahseen S, et al. Population structure, biogeography and transmissibility of Mycobacterium tuberculosis. Nat Commun. 2021 Dec; 12(1):6099. https://doi.org/10.1038/s41467-021-26248-1 PMID: 34671035
12. Napier G, Campino S, Merid Y, Abebe M, Woldeamanuel Y, Aseffa A, et al. Robust barcoding and identification of Mycobacterium tuberculosis lineages for epidemiological and clinical studies. Genome Med. 2020 Dec; 12(1):114. https://doi.org/10.1186/s13073-020-00817-3 PMID: 33317631
13. Steiner A, Stucki D, Coscolla M, Borrell S, Gagneux S. KvarQ: targeted and direct variant calling from fastq reads of bacterial genomes. BMC Genomics. 2014 Oct 9; 15(1):881. https://doi.org/10.1186/1471-2164-15-881 PMID: 25297886
14. Sekizuka T, Yamashita A, Murase Y, Iwamoto T, Mitarai S, Kato S, et al. TGS-TB: Total Genotyping Solution for Mycobacterium tuberculosis Using Short-Read Whole-Genome Sequencing. PLOS ONE. 2015 Nov 13; 10(11):e0142951. https://doi.org/10.1371/journal.pone.0142951 PMID: 26565975
15. Mahé P, El Azami M, Barlas P, Tournoud M. A large scale evaluation of TBProfiler and Mykrobe for antibiotic resistance prediction in Mycobacterium tuberculosis. PeerJ. 2019 May 1; 7:e6857. https://doi.org/10.7717/peerj.6857 PMID: 31106066

16. Iwai H, Kato-Miyazawa M, Kirikae T, Miyoshi-Akiyama T. CASTB (the comprehensive analysis server for the Mycobacterium tuberculosis complex): A publicly accessible web server for epidemiological analyses, drug-resistance prediction and phylogenetic comparison of clinical isolates. Tuberc Edinb Scotl. 2015 Dec; 95(6):843–4.

17. Benavente ED, Coll F, Furnham N, McNerney R, Glynn JR, Campino S, et al. PhyTB: Phylogenetic tree visualisation and sample positioning for M. tuberculosis. BMC Bioinformatics. 2015 May 13; 16:155. https://doi.org/10.1186/s12859-015-0603-3 PMID: 25968323

18. Ezewudo M, Borens A, Chiner-Oms Á, Miotto P, Chindelevitch L, Starks AM, et al. Integrating standardized whole genome sequence analysis with a global Mycobacterium tuberculosis antibiotic resistance knowledgebase. Sci Rep. 2018 Oct 18; 8(1):15382. https://doi.org/10.1038/s41598-018-33731-1 PMID: 30337678

19. Gröschel MI, Owens M, Freschi L, Vargas R, Marin MG, Phelan J, et al. GenTB: A user-friendly genome-based predictor for tuberculosis resistance powered by machine learning. Genome Med. 2021 Aug 30; 13(1):138. https://doi.org/10.1186/s13073-021-00953-4 PMID: 34461978

20. Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, Cirillo DM, et al. PhyResSE: a Web Tool Delineating Mycobacterium tuberculosis Antibiotic Resistance and Lineage from Whole-Genome Sequencing Data. J Clin Microbiol. 2015 Jun; 53(6):1908–14. https://doi.org/10.1128/JCM.00025-15 PMID: 25854485

21. Yang T, Gan M, Liu Q, Liang W, Tang Q, Luo G, et al. SAM-TB: a whole genome sequencing data analysis website for detection of Mycobacterium tuberculosis drug resistance and transmission. Brief Bioinform. 2022 Feb 24; 23(2):bbac030. https://doi.org/10.1093/bib/bbac030 PMID: 35211720

22. Goldfeder RL, Priest JR, Zook JM, Grove ME, Waggott D, Wheeler MT, et al. Medical implications of technical accuracy in genome sequencing. Genome Med. 2016 Mar 2; 8(1):24. https://doi.org/10.1186/s13073-016-0269-0 PMID: 26932475

23. World Health Organization. Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance [Internet]. Geneva, Switzerland: World Health Organization; 2021 [cited 2022 Aug 19]. 86 p. Available from: https://www.who.int/publications-detail-redirect/9789240028173

24. McNerney R, Zignol M, Clark TG. Use of whole genome sequencing in surveillance of drug resistant tuberculosis. Expert Rev Anti Infect Ther. 2018 May 4; 16(5):433–42. https://doi.org/10.1080/14787210.2018.1472577 PMID: 29718745

25. Sanchini A, Jandrasits C, Tembrockhaus J, Kohl TA, Utpatel C, Maurer FP, et al. Improving tuberculosis surveillance by detecting international transmission using publicly available whole genome sequencing data. Eurosurveillance. 2021 Jan 14; 26(2):1900677. https://doi.org/10.2807/1560-7917.ES.2021.26.2.1900677 PMID: 33446303

26. van der Werf MJ, Ködmön C. Whole-Genome Sequencing as Tool for Investigating International Tuberculosis Outbreaks: A Systematic Review. Front Public Health [Internet]. 2019 [cited 2022 Aug 4];7. Available from: https://www.frontiersin.org/articles/10.3389/fpubh.2019.00087 PMID: 31058125

27. World Bank. Incidencia de tuberculosis (por cada 100.000 personas)—Ecuador | Data [Internet]. [cited 2022 Nov 29]. Available from: https://datos.bancomundial.org/indicador/SH.TBS.INCD?locations=EC

28. Canetti G, Rist N, Grosset J. [Measurement of sensitivity of the tuberculous bacillus to antibacillary drugs by the method of proportions. Methodology, resistance criteria, results and interpretation]. Rev Tuberc Pneumol (Paris). 1963 Mar; 27:217–72. PMID: 14018284

29. Direct Susceptibility Testing of Mycobacterium tuberculosis for Pyrazinamide by Use of the Bactec MGIT 960 System. [cited 2022 Nov 4]; Available from: https://journals.asm.org/doi/epdf/10.1128/JCM.03162-15?src=getftr

30. Van Soolingen D. Molecular epidemiology of tuberculosis and other mycobacterial infections: main methodologies and achievements. J Intern Med. 2001; 249(1):1–26. https://doi.org/10.1046/j.1365-2796.2001.00772.x PMID: 11168781

31. Larsen MH, Biermann K, Tandberg S, Hsu T, Jacobs WR Jr. Genetic Manipulation of Mycobacterium tuberculosis. Curr Protoc Microbiol. 2007; 6(1):10A.2.1-10A.2.21. https://doi.org/10.1002/9780471729259.mc10a02s6 PMID: 18770603

32. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014; 15(3):R46. https://doi.org/10.1186/gb-2014-15-3-r46 PMID: 24580807

33. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. 2017.

34. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1; 30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404

35. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009 Jul 15; 25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168

36. Kohl TA, Utpatel C, Schleusener V, De Filippo MR, Beckert P, Cirillo DM, et al. MTBseq: a comprehensive pipeline for whole genome sequence analysis of Mycobacterium tuberculosis complex isolates. PeerJ. 2018 Nov 13; 6:e5895. https://doi.org/10.7717/peerj.5895 PMID: 30479891

37. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Mol Biol Evol. 2020 Jan 1; 37(1):291–4. https://doi.org/10.1093/molbev/msz189 PMID: 31432070

38. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019 Nov 1; 35(21):4453–5. https://doi.org/10.1093/bioinformatics/btz305 PMID: 31070718

39. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021 Apr 22; 49(W1):W293–6. https://doi.org/10.1093/nar/gkab301 PMID: 33885785

40. Diel R, Kohl TA, Maurer FP, Merker M, Meywald Walter K, Hannemann J, et al. Accuracy of whole-genome sequencing to determine recent tuberculosis transmission: an 11-year population-based study in Hamburg, Germany. Eur Respir J. 2019 Nov; 54(5):1901154. https://doi.org/10.1183/13993003.01154-2019 PMID: 31467121

41. Cancino-Muñoz I, López MG, Torres-Puente M, Villamayor LM, Borrás R, Borrás-Máñez M, et al. Population-based sequencing of Mycobacterium tuberculosis reveals how current population dynamics are shaped by past epidemics. Lewnard J, Franco E, Meehan CJ, Walker TM, editors. eLife. 2022 Jul 26; 11:e76605.

42. Hofmann-Thiel S, Hoffmann H, Hillemann D, Rigouts L, Van Deun A, Kranzer K. How should discordance between molecular and growth-based assays for rifampicin resistance be investigated? Int J Tuberc Lung Dis. 2017 Jul 1; 21(7):721–6. https://doi.org/10.5588/ijtld.17.0140 PMID: 28633695

43. Mokaddas E, Ahmad S, Eldeen HS, Al-Mutairi N. Discordance between Xpert MTB/RIF Assay and Bactec MGIT 960 Culture System for Detection of Rifampin-Resistant Mycobacterium tuberculosis Isolates in a Country with a Low Tuberculosis (TB) Incidence. J Clin Microbiol. 2015 Apr; 53(4):1351–4. https://doi.org/10.1128/JCM.03412-14 PMID: 25609730

44. Danchuk SN, Solomon OE, Kohl TA, Niemann S, Soolingen D van, Ingen J van, et al. Challenging the gold standard: critical limitations in clinical detection of drug-resistant tuberculosis [Internet]. medRxiv; 2023 [cited 2023 Mar 9]. p. 2023.02.27.23286518. Available from: https://www.medrxiv.org/content/10.1101/2023.02.27.23286518v1

45. Garzon-Chavez D, Garcia-Bereguiain MA, Mora-Pinargote C, Granda-Pardo JC, Leon-Benitez M, Franco-Sotomayor G, et al. Population structure and genetic diversity of Mycobacterium tuberculosis in Ecuador. Sci Rep. 2020 Dec; 10(1):6237. https://doi.org/10.1038/s41598-020-62824-z PMID: 32277077

46. Zurita J, Espinel N, Barba P, Ortega-Paredes D, Zurita-Salinas C, Rojas Y, et al. Genetic diversity and drug resistance of Mycobacterium tuberculosis in Ecuador. Int J Tuberc Lung Dis. 2019 Feb 1; 23(2):166–73. https://doi.org/10.5588/ijtld.18.0095 PMID: 30808448

47. Xu Y, Cancino-Muñoz I, Torres-Puente M, Villamayor LM, Borrás R, Borrás-Máñez M, et al. High-resolution mapping of tuberculosis transmission: Whole genome sequencing and phylogenetic modelling of a cohort from Valencia Region, Spain. PLoS Med. 2019 Oct 31; 16(10):e1002961. https://doi.org/10.1371/journal.pmed.1002961 PMID: 31671150

48. Nelson KN, Talarico S, Poonja S, McDaniel CJ, Cilnis M, Chang AH, et al. Mutation of Mycobacterium tuberculosis and Implications for Using Whole-Genome Sequencing for Investigating Recent Tuberculosis Transmission. Front Public Health. 2022 Jan 13; 9:790544. https://doi.org/10.3389/fpubh.2021.790544 PMID: 35096744

49. Colijn C, Cohen T. Whole-genome sequencing of Mycobacterium tuberculosis for rapid diagnostics and beyond. Lancet Respir Med. 2016 Jan 1; 4(1):6–8. https://doi.org/10.1016/S2213-2600(15)00510-X PMID: 26669892

50. Sánchez-Corrales L, Tovar-Aguirre OL, Galeano-Vanegas NF, Castaño Jiménez PA, Martínez-Vega RA, Maldonado-Londoño CE, et al. Phylogenomic analysis and Mycobacterium tuberculosis antibiotic resistance prediction by whole-genome sequencing from clinical isolates of Caldas, Colombia. PLoS ONE. 2021 Oct 7; 16(10):e0258402. https://doi.org/10.1371/journal.pone.0258402 PMID: 34618869

51. Rocha DMGC, Magalhães C, Cá B, Ramos A, Carvalho T, Comas I, et al. Heterogeneous Streptomycin Resistance Level Among Mycobacterium tuberculosis Strains From the Same Transmission Cluster. Front Microbiol [Internet]. 2021 [cited 2022 Jun 18];12. Available from: https://www.frontiersin.org/article/10.3389/fmicb.2021.659545 PMID: 34177837

52. van Beek J, Haanperä M, Smit PW, Mentula S, Soini H. Evaluation of whole genome sequencing and software tools for drug susceptibility testing of Mycobacterium tuberculosis. Clin Microbiol Infect. 2019 Jan; 25(1):82–6. https://doi.org/10.1016/j.cmi.2018.03.041 PMID: 29653190

53. Senghore M, Diarra B, Gehre F, Otu J, Worwui A, Muhammad AK, et al. Evolution of Mycobacterium tuberculosis complex lineages and their role in an emerging threat of multidrug resistant tuberculosis in Bamako, Mali. Sci Rep. 2020 Dec; 10(1):327. https://doi.org/10.1038/s41598-019-56001-0 PMID: 31941887

54. Coll F, McNerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. Genome Med. 2015 Dec; 7(1):51. https://doi.org/10.1186/s13073-015-0164-0 PMID: 26019726

55. Pinhata JMW, Brandao AP, Mendes F de F, Rabello MC da S, Ferrazoli L, de Oliveira RS. Correlating genetic mutations with isoniazid phenotypic levels of resistance in Mycobacterium tuberculosis isolates from patients with drug-resistant tuberculosis in a high burden setting. Eur J Clin Microbiol Infect Dis. 2021 Dec 1; 40(12):2551–61. https://doi.org/10.1007/s10096-021-04316-0 PMID: 34297229

56. Guo Y, Cao X, Yang J, Wu X, Liu Y, Wan B, et al. Rifampin-resistance-associated mutations in the rifampin-resistance-determining region of the rpoB gene of Mycobacterium tuberculosis clinical isolates in Shanghai, PR China. J Med Microbiol. 2021 Mar; 70(3). https://doi.org/10.1099/jmm.0.001317 PMID: 33507146

57. Li M chao, Chen R, Lin S qiang, Lu Y, Liu H can, Li G lian, et al. Detecting Ethambutol Resistance in Mycobacterium tuberculosis Isolates in China: A Comparison Between Phenotypic Drug Susceptibility Testing Methods and DNA Sequencing of embAB. Front Microbiol [Internet]. 2020 [cited 2022 Nov 3];11. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2020.00781

58. Hameed HA, Tan Y, Islam MM, Lu Z, Chhotaray C, Wang S, et al. Detection of Novel Gene Mutations Associated with Pyrazinamide Resistance in Multidrug-Resistant Mycobacterium tuberculosis Clinical Isolates in Southern China. Infect Drug Resist. 2020 Jan 22; 13:217–27. https://doi.org/10.2147/IDR.S230774 PMID: 32158237

59. Ang MLT, Zainul Rahim SZ, de Sessions PF, Lin W, Koh V, Pethe K, et al. EthA/R-Independent Killing of Mycobacterium tuberculosis by Ethionamide. Front Microbiol [Internet]. 2017 [cited 2022 Nov 3];8. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2017.00710 PMID: 28487681

60. Sowajassatakul A, Prammananan T, Chaiprasert A, Phunpruch S. Overexpression of eis without a mutation in promoter region of amikacin- and kanamycin-resistant Mycobacterium tuberculosis clinical strain. Ann Clin Microbiol Antimicrob. 2018 Jul 16; 17(1):33. https://doi.org/10.1186/s12941-018-0285-6 PMID: 30008266

61. Chaiyachat P, Chaiprasert A, Nonghanphithak D, Smithtikarn S, Kamolwat P, Pungrassami P, et al. Whole-genome analysis of drug-resistant Mycobacterium tuberculosis reveals novel mutations associated with fluoroquinolone resistance. Int J Antimicrob Agents. 2021 Sep 1; 58(3):106385. https://doi.org/10.1016/j.ijantimicag.2021.106385 PMID: 34161790

62. Gomes LC, Campino S, Marinho CRF, Clark TG, Phelan JE. Whole genome sequencing reveals large deletions and other loss of function mutations in Mycobacterium tuberculosis drug resistance genes. Microb Genomics. 2021 Dec 10; 7(12):000724. https://doi.org/10.1099/mgen.0.000724 PMID: 34889724

63. Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, Hannan M, et al. Functional and evolutionary genomics of Mycobacterium tuberculosis: Insights from genomic deletions in 100 strains. Proc Natl Acad Sci. 2004 Apr 6; 101(14):4865–70. https://doi.org/10.1073/pnas.0305634101 PMID: 15024109

64. Kardan-Yamchi J, Amini S, Hamzelou G, Rahimi Foroushani A, Ghodousi A, Cirillo DM, et al. Assessment of the GenoType MTBDRsl VER 2.0 compared to the phenotypic drug susceptibility testing and whole genome sequencing for the rapid detection of resistance to fluoroquinolone and second-line injectable drugs among rifampicin-resistant Mycobacterium tuberculosis isolates. Arch Microbiol. 2021 Sep 1; 203(7):3989–96. https://doi.org/10.1007/s00203-021-02387-3 PMID: 34032874

65. Cox H, Goig GA, Salaam-Dreyer Z, Dippenaar A, Reuter A, Mohr-Holland E, et al. Whole-Genome Sequencing Has the Potential To Improve Treatment for Rifampicin-Resistant Tuberculosis in High-Burden Settings: a Retrospective Cohort Study. J Clin Microbiol. 2022 Mar 16; 60(3):e02362–21. https://doi.org/10.1128/jcm.02362-21 PMID: 35170980

66. Mvelase NR, Pillay M, Sibanda W, Ngozo JN, Brust JCM, Mlisana KP. rpoB Mutations Causing Discordant Rifampicin Susceptibility in Mycobacterium tuberculosis: Retrospective Analysis of Prevalence, Phenotypic, Genotypic, and Treatment Outcomes. Open Forum Infect Dis. 2019 Apr 1; 6(4):ofz065. https://doi.org/10.1093/ofid/ofz065 PMID: 31024968

67. Rando-Segura A, Aznar ML, Moreno MM, Espasa Soley M, Sulleiro Igual E, Bocanegra Garcia C, et al. Molecular characterization of rpoB gene mutations in isolates from tuberculosis patients in Cubal, Republic of Angola. BMC Infect Dis. 2021 Oct 12; 21(1):1056. https://doi.org/10.1186/s12879-021-06763-8 PMID: 34641802

68. Wang S, Zhou Y, Zhao B, Ou X, Xia H, Zheng Y, et al. Characteristics of compensatory mutations in the rpoC gene and their association with compensated transmission of Mycobacterium tuberculosis. Front Med. 2020 Feb 1; 14(1):51–9. https://doi.org/10.1007/s11684-019-0720-x PMID: 31938981

69. Yun YJ, Lee JS, Yoo JC, Cho E, Park D, Kook YH, et al. Patterns of rpoC Mutations in Drug-Resistant Mycobacterium tuberculosis Isolated from Patients in South Korea. Tuberc Respir Dis. 2018 Jul; 81 (3):222–7. https://doi.org/10.4046/trd.2017.0042 PMID: 29527837

70. Conkle-Gutierrez D, Ramirez-Busby S, Gorman B, Elghraoui A, Hoffner S, Elmaraachli W, et al. Novel and Reported Compensatory Mutations in rpoABC Associate Specifically with Predominant Mycobacterium tuberculosis Rifampicin Resistance Marker rpoB:S450L [Internet]. Microbiology; 2022 Feb [cited 2022 Nov 5]. Available from: http://biorxiv.org/lookup/doi/10.1101/2022.02.22.481565

71. Vargas AP, Rios AA, Grandjean L, Kirwan DE, Gilman RH, Sheen P, et al. Determination of Potentially Novel Compensatory Mutations in rpoC Associated with Rifampin Resistance and rpoB Mutations in Mycobacterium tuberculosis Clinical Isolates from Peru. Int J Mycobacteriology. 2020; 9(2):26. https://doi.org/10.4103/ijmy.ijmy_27_20 PMID: 32474533

72. Salvato RS, Reis AJ, Schiefelbein SH, Gómez MAA, Salvato SS, da Silva LV, et al. Genomic-based surveillance reveals high ongoing transmission of multi-drug-resistant Mycobacterium tuberculosis in Southern Brazil. Int J Antimicrob Agents. 2021 Oct 1; 58(4):106401. https://doi.org/10.1016/j.ijantimicag.2021.106401 PMID: 34289403

73. Bakhtiyariniya P, Khosravi AD, Hashemzadeh M, Savari M. Detection and characterization of mutations in genes related to isoniazid resistance in Mycobacterium tuberculosis clinical isolates from Iran. Mol Biol Rep [Internet]. 2022 Apr 2 [cited 2022 Jun 18]; Available from: https://link.springer.com/10.1007/s11033-022-07404-2 PMID: 35366177

74. Penn-Nicholson A, Georghiou SB, Ciobanu N, Kazi M, Bhalla M, David A, et al. Detection of isoniazid, fluoroquinolone, ethionamide, amikacin, kanamycin, and capreomycin resistance by the Xpert MTB/XDR assay: a cross-sectional multicentre diagnostic accuracy study. Lancet Infect Dis. 2022 Feb 1; 22 (2):242–9. https://doi.org/10.1016/S1473-3099(21)00452-7 PMID: 34627496

75. Ushtanit A, Kulagina E, Mikhailova Y, Makarova M, Safonova S, Zimenkov D. Molecular Determinants of Ethionamide Resistance in Clinical Isolates of Mycobacterium tuberculosis. Antibiotics. 2022 Feb; 11 (2):133. https://doi.org/10.3390/antibiotics11020133 PMID: 35203736

76. Rahman SMM, Ather MdF, Nasrin R, Hoque MA, Khatun R, Rahman T, et al. Performance of WHO-Endorsed Rapid Tests for Detection of Susceptibility to First-Line Drugs in Patients with Pulmonary Tuberculosis in Bangladesh. Diagnostics. 2022 Feb 5; 12(2):410. https://doi.org/10.3390/diagnostics12020410 PMID: 35204501

77. Sun Q, Xiao T yang, Liu H can, Zhao X qin, Liu Z guang, Li Y nan, et al. Mutations within embCAB Are Associated with Variable Level of Ethambutol Resistance in Mycobacterium tuberculosis Isolates from China. Antimicrob Agents Chemother. 2017 Dec 21; 62(1):e01279–17. https://doi.org/10.1128/AAC.01279-17 PMID: 29084750

78. Xu Y, Jia H, Huang H, Sun Z, Zhang Z. Mutations Found in embCAB, embR, and ubiA Genes of Ethambutol-Sensitive and -Resistant Mycobacterium tuberculosis Clinical Isolates from China. BioMed Res Int. 2015; 2015:951706. https://doi.org/10.1155/2015/951706 PMID: 26417605

79. Wang J, Zhao W, Liu R, Huo F, Dong L, Xue Y, et al. Rapid Detection of Ethambutol-Resistant Mycobacterium tuberculosis from Sputum by High-Resolution Melting Analysis in Beijing, China. Infect Drug Resist. 2020 Oct 20; 13:3707–13. https://doi.org/10.2147/IDR.S270542 PMID: 33116691

80. Madrazo-Moya CF, Cancino-Muñoz I, Cuevas-Córdoba B, González-Covarrubias V, Barbosa-Amezcua M, Soberón X, et al. Whole genomic sequencing as a tool for diagnosis of drug and multidrug-resistance tuberculosis in an endemic region in Mexico. PLOS ONE. 2019 Jun 5; 14(6):e0213046. https://doi.org/10.1371/journal.pone.0213046 PMID: 31166945

81. Whitfield MG, Engelthaler DM, Allender C, Folkerts M, Heupink TH, Limberis J, et al. Comparative Performance of Genomic Methods for the Detection of Pyrazinamide Resistance and Heteroresistance in Mycobacterium tuberculosis. J Clin Microbiol. 2022; 60(1):e01907–21. https://doi.org/10.1128/JCM.01907-21 PMID: 34757831

82. Shrestha D, Maharjan B, Thapa J, Akapelwa ML, Bwalya P, Chizimu JY, et al. Detection of Mutations in pncA in Mycobacterium tuberculosis Clinical Isolates from Nepal in Association with Pyrazinamide Resistance. Curr Issues Mol Biol. 2022 Sep; 44(9):4132–41. https://doi.org/10.3390/cimb44090283 PMID: 36135195

83. He W, Tan Y, Liu C, Wang Y, He P, Song Z, et al. Drug-Resistant Characteristics, Genetic Diversity, and Transmission Dynamics of Rifampicin-Resistant Mycobacterium tuberculosis in Hunan, China, Revealed by Whole-Genome Sequencing. Microbiol Spectr. 2022 Feb 16; 10(1):e01543–21. https://doi.org/10.1128/spectrum.01543-21 PMID: 35171016

84. Kabir S, Tahir Z, Mukhtar N, Sohail M, Saqalein M, Rehman A. Fluoroquinolone resistance and mutational profile of gyrA in pulmonary MDR tuberculosis patients. BMC Pulm Med. 2020 May 11; 20:138. https://doi.org/10.1186/s12890-020-1172-4 PMID: 32393213

85. Sun H, Zeng J, Li S, Liang P, Zheng C, Liu Y, et al. Interaction between rpsL and gyrA mutations affects the fitness and dual resistance of Mycobacterium tuberculosis clinical isolates against streptomycin and fluoroquinolones. Infect Drug Resist. 2018 Mar 27; 11:431–40. https://doi.org/10.2147/IDR.S152335 PMID: 29628767

86. El Garch F, Lismond A, Piddock LJV, Courvalin P, Tulkens PM, Van Bambeke F. Fluoroquinolones induce the expression of patA and patB, which encode ABC efflux pumps in Streptococcus pneumoniae. J Antimicrob Chemother. 2010 Oct; 65(10):2076–82. https://doi.org/10.1093/jac/dkq287 PMID: 20709735

87. Cohen KA, Stott KE, Munsamy V, Manson AL, Earl AM, Pym AS. Evidence for Expanding the Role of Streptomycin in the Management of Drug-Resistant Mycobacterium tuberculosis. Antimicrob Agents Chemother. 2020 Aug 20; 64(9):e00860–20. https://doi.org/10.1128/AAC.00860-20 PMID: 32540971

88. Conkle-Gutierrez D, Kim C, Ramirez-Busby SM, Modlin SJ, Mansjö M, Werngren J, et al. Rare alternative second line injectable drug resistance markers identified by gene-wise genome wide association in M. tuberculosis with unexplained resistance [Internet]. Genomics; 2021 Nov [cited 2022 Nov 6]. Available from: http://biorxiv.org/lookup/doi/10.1101/2021.11.23.469801

89. Matsui T, Pinhata JMW, Rabello MC da S, Brandão AP, Ferrazoli L, Leão SC, et al. Frequency of first and second-line drug resistance-associated mutations among resistant Mycobacterium tuberculosis clinical isolates from São Paulo, Brazil. Mem Inst Oswaldo Cruz. 2020; 115:e200055.

90. Li Q, Gao H, Zhang Z, Tian Y, Liu T, Wang Y, et al. Mutation and Transmission Profiles of Second-Line Drug Resistance in Clinical Isolates of Drug-Resistant Mycobacterium tuberculosis From Hebei Province, China. Front Microbiol [Internet]. 2019 [cited 2022 Nov 6];10. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2019.01838 PMID: 31447823

91. Coll F, Phelan J, Hill-Cawthorne GA, Nair MB, Mallard K, Ali S, et al. Genome-wide analysis of multi- and extensively drug-resistant Mycobacterium tuberculosis. Nat Genet. 2018 Feb; 50(2):307–16. https://doi.org/10.1038/s41588-017-0029-0 PMID: 29358649

92. Ali A, Hasan Z, McNerney R, Mallard K, Hill-Cawthorne G, Coll F, et al. Whole Genome Sequencing Based Characterization of Extensively Drug-Resistant Mycobacterium tuberculosis Isolates from Pakistan. PLOS ONE. 2015 Feb 26; 10(2):e0117771. https://doi.org/10.1371/journal.pone.0117771 PMID: 25719196

93. Georghiou SB, Magana M, Garfein RS, Catanzaro DG, Catanzaro A, Rodwell TC. Evaluation of Genetic Mutations Associated with Mycobacterium tuberculosis Resistance to Amikacin, Kanamycin and Capreomycin: A Systematic Review. PLOS ONE. 2012 Mar 29; 7(3):e33275. https://doi.org/10.1371/journal.pone.0033275 PMID: 22479378

94. Kambli P, Ajbani K, Nikam C, Sadani M, Shetty A, Udwadia Z, et al. Correlating rrs and eis promoter mutations in clinical isolates of Mycobacterium tuberculosis with phenotypic susceptibility levels to the second line injectables. Int J Mycobacteriology. 2016 Mar; 5(1):1–6.

95. Lau RWT, Ho PL, Kao RYT, Yew WW, Lau TCK, Cheng VCC, et al. Molecular Characterization of Fluoroquinolone Resistance in Mycobacterium tuberculosis: Functional Analysis of gyrA Mutation at Position 74. Antimicrob Agents Chemother. 2011 Feb; 55(2):608–14. https://doi.org/10.1128/AAC.00920-10 PMID: 20956608

Artículo 4: Decoding Ecuadorian *Mycobacterium tuberculosis* Isolates: Unveiling Lineage-Associated Signatures in Beta-Lactamase Resistance via Pangenome Analysis.

En este artículo se evaluó la composición genética de 88 aislados ecuatorianos (66 de pacientes masculinos y 22 de pacientes femeninos) de *M. tuberculosis* mediante el análisis de su pangenoma. Se identificó que el core genoma de los 88 aislados está compuesto por aproximadamente 3200 genes, mientras que los genes accesorios y/o únicos pueden variar entre 100 y 700, dependiendo del algoritmo utilizado. Se observó que los aislados pertenecientes a los sublinajes LAM y X-type presentaron una mayor diversidad genética. La categoría predominante dentro de las particiones del pangenoma fue la de genes relacionados con el metabolismo, según la clasificación de la base COG. Además, se identificaron las vías metabólicas más destacadas, como el procesamiento de información medioambiental según la base de datos KEGG.

Se encontraron genes asociados con el transporte y metabolismo de lípidos, aminoácidos y metabolitos secundarios, especialmente en los sublinajes LAM y S-type, lo que sugiere que estos aislados pueden tener una mayor capacidad de adaptación a diferentes condiciones del entorno, en concordancia con la hipervariabilidad del huésped. También se identificaron variaciones genéticas en genes relacionados con la biosíntesis de la pared celular, que se han sugerido como posibles contribuyentes a la susceptibilidad a los beta-lactámicos. Sin embargo, algunas de estas variaciones podrían estar relacionadas con resistencias a fármacos comúnmente utilizados en el tratamiento de la tuberculosis, lo cual sugiere la necesidad de ampliar la investigación en este aspecto.

La generación de información genómica de los aislados ecuatorianos contribuye al entendimiento de la asociación clínica de las variantes genéticas locales en los programas de vigilancia de la tuberculosis en el país.

Decoding Ecuadorian *Mycobacterium tuberculosis* Isolates: Unveiling Lineage-Associated Signatures in Beta-Lactamase Resistance via Pangenome Analysis

Gabriel Morey-León[1,2,*], Juan Carlos Fernández-Cadena[1,3], Derly Andrade-Molina[1,3], Luisa Berná[4,5,6,*]

[1] Facultad de Ciencias de la Salud. Universidad Espíritu Santo, Samborondón, Ecuador.

[2] Universidad de la República. Montevideo, Uruguay

[3] Laboratorio de Ciencias Ómicas. Universidad Espíritu Santo, Samborondón, Ecuador.

[4] Institut Pasteur de Montevideo, Montevideo, Uruguay.

[5]Laboratorio de Interacciones Hospedero-Patógeno, Unidad de Biología Molecular, Institut Pasteur de Montevideo, Montevideo, Uruguay

[6]Unidad de Genómica Evolutiva, Facultad de Ciencias, Universidad de la República, Uruguay.

*Corresponding author: garielmorey@uees.edu.ec

Abstract

Background: Tuberculosis (TB) is the second largest public health threat caused by pathogens. Understanding the transmission, virulence, and resistance profile of *Mycobacterium tuberculosis* (*Mtb*) is crucial for outbreak control. This study aimed to investigate the pangenome composition of 88 *Mtb* strains classified as L4 derived from Ecuador by whole-genome sequencing with Roary and BPGA pipelines. **Results:** Our analysis identified a core genome of approximately 3200 genes and a pangenome that differed in accessory and unique genes. Metabolism-related genes were the most representative in all partitions according to the COG database. However, differences were found in the metabolic pathways described by KEGG in all the lineages analyzed. Isolates from Ecuador had variations in genomics regions associated with beta-lactamase susceptibility, potentially leading to epistatic resistance to other drugs commonly used in TB treatment, warranting further investigation. **Conclusion:** Our findings provide valuable insights into the genetic diversity of Mtb in Ecuador. These insights may be associated with increasing adaptation within-host heterogeneity, variable latency periods, and reduced host damage, collectively contributing to disease spread.

The application of whole-genome sequencing is essential to elucidating the epidemiology of TB in the country.

keywords: Pangenome, core genome, tuberculosis, beta-lactamase, Ecuador, Surveillance

Background

Tuberculosis (TB) caused by *Mycobacterium tuberculosis* (*Mtb*) is a significant global health concern, with an estimated 10.6 million infections and 1.6 million deaths in 2021(1). *Mtb* encompasses a complex of genetically related bacteria, including *M. africanum*(2,3), which, with *Mtb,* are typical human pathogens, and other lineages that infect animals (*M. bovis, M. canetii, M. microti, M. caprae, M. pinnipedii,* and *M. orygis*)(4–7), which have evolved from common ancestors over centuries from different geographic areas and adapted to external conditions(8–10). Some studies have demonstrated the relationship of clinical isolates to specific genotypes and geographical regions, as well as their remarkable capacity for increased dissemination that sets them apart from ancestral lineages(11–14). This is particularly evident due to their higher virulence and shorter latency periods to change the active form.

Furthermore, it has been noted that even within modern lineages, which are known for their heightened virulence and rapid dissemination, not all isolates display identical characteristics(15). This diversity in behaviour is influenced by various external factors, including drug resistance, host heterogeneity, demographic factors, and the presence of dominant lineages that confer advantages in terms of dissemination and impact on the host(10,16). In Ecuador, it has been reported that lineage Euroamerican (L4) is the most predominant, and molecular epidemiology studies have shown that the population structure of *Mtb* is composed of LAM, X-type, and Haarlem sublineages(17–21).

Genomic approaches, such as whole-genome sequencing (WGS) of *Mtb* strains, have improved the understanding of composition analysis and provided invaluable information on gain or loss genes, evolutionary markers, and polymorphisms related to drug resistance, virulence, and sublineage patterns(22–28). With a high amount of information on genes from sequence isolates, the pangenome-based approach is the more convenient to discern a complete analysis, in search of unique (specific to one single strain), accessory (shared among two or more strains), and core genes (present in all strains), estimating the

diversity of genes and novel marker genes, especially local distribution associated(29,30), mostly associated with virulence and drug resistance.

Pangenome studies have been performed on different isolates from Peru, Mexico(31), Brazil, Argentina, Paraguay, and Colombia(32) from clinical samples of *M. tuberculosis* to understand the variation in terms of unique sequences among them and identify the importance of genes related to metabolism to adapted external conditions and latency periods. However, no extensive pangenome studies have been performed in Ecuador to reveal the composition of the circulating lineages in the country. A unique WGS study identified the presence of 4.3.2/3 (LAM) and 4.1.2 (Haarlem) sublineages in a small dataset of *Mtb* strains from Ecuador(33). In this study, we aimed to investigate the genomic composition of 88 *Mtb* strains classified as L4 derived from Ecuador by pangenome analysis. In addition, we explored the genetic variation in genomic regions associated with beta-lactamase susceptibility. Our findings provide valuable insights into the genetic diversity of *Mtb* in Ecuador and suggest the association of lineage with specific mutations related to beta-lactamase resistance in TB that drive the development of more effective TB control strategies tailored to the specific characteristics of the local *Mtb* population in the region.

2. Materials and Methods

2.1. *Mtb* samples, assembly, functional annotation, and pangenome analysis

This study included reads from 88 clinical isolates of *Mtb* collected from different provinces in Ecuador between 2019 and 2021, which are available in the Sequence Read Archive (SRA) database (http://www.ncbi.nlm.nih.gov/sra) under the PRJNA827129 BioProject. Raw reads were quality-checked using FastQC v0.11.9(34) and improved by Trimmomatic v0.38(35). Species confirmation and contamination screening were performed by Kraken v2(36). A Unicycler assembly pipeline(37) was used to assemble the high-quality reads with a 300 bp minimum contig size and polished with Pilon(38). The quality of the assemblies was evaluated using *Mtb* strain HR37v (NC_000962.3) as a reference genome in Quast v5.0.2.(39). Structural and functional annotation was performed using the Rapid Annotation using Subsystem Technology tool kit (RASTtk) in the Pathosystems Resource Integration Center (PATRIC) (https://www.patricbrc.org/). The pangenome analysis was performed with Panaroo(40), and the BPGA pipeline(41) from the GFF archives annotated by Prokka v1.14.16(42) using the H37Rv (NC_000962.3) *M.*

tuberculosis reference genome. Panaroo was performed with a minimum identity percentage for blastp of 95%, and gene detection in 99% of isolates was to be recorded as a core gene. Roary_stats, a custom R tool, was used to process the outputs. In the BPGA pipeline, the COG and KEEG databases were applied to define the functionality of the gene characterized. The pangenome was divided into three partitions: core, accessory, and cloud/unique. The core genes were present in 100% of the isolates, accessory in 3-99%, and cloud/unique in 1-2%.

2.2. Variant calling, lineage classification, and drug-resistance genes

Genetic relatedness among 88 *Mtb* genomes was processed with the MTBseq pipeline using the standard input parameters [(42)]. Briefly, BWA-mem and SAMtools algorithms were used for mapping the reads to the reference genome of *Mtb* (NC_000962.3); GATK v3 was applied for base call recalibration and realignment of reads around InDels, and SAMtools mpileup was used for variant calling. Good-quality genomes presented a minimum coverage mean of >20x, read depth DP<5, and reference genome coverage >95%. The sublineage classification and the genetic distance matrix among transmission groups were also obtained from MTBseq analysis. The TB-Profiler v4 pipeline was used to predict the canonical mutations in genes associated with resistance to first- and second-line drugs. We used bcftool as the caller algorithm within the TB-profiler. For beta-lactamase susceptibility analysis, 46 genes were selected because they were especially related to beta-lactamase resistance and cell wall biosynthesis.

3. Results

3.1. Assembly, functional annotation, and pangenome analysis

In this study, we conducted a comparative genome analysis of 88 isolates of Mtb from Ecuador. The results revealed an average genome size of 4.32 Mb (range 4.21 - 4.33) and a high GC content of 65.2-65.5%. The depth and coverage are approximately ~60.9X ± 23.6 and 99.8%. The analysis of single nucleotide variants (SNVs) revealed an average of approximately 788 ± 111 SNPs, along with approximately 51 ± 8 insertions and 47 ± 7 deletions. Notably, the majority of InDels were short (≤20 nt, as shown in Table 1). The distribution of SNPs encompassed approximately 105 ± 18 intergenic, 372 ± 47 nonsynonymous, and 232 ± 36 synonymous variations. Genome annotation highlighted the existence of approximately 4,294 ± 14 genes, consisting of 4,254 coding sequences (CDS) and 44 transfer RNA (tRNA) genes, with functional assignments

identified in approximately 3,378 ± 12 cases. A total of approximately 1,061 ± 4 proteins were annotated with Enzyme Commission (EC) numbers, while approximately 918 ± 4 exhibited Gene Ontology (GO) assignments. Moreover, approximately 814 ± 3 proteins were successfully linked to KEGG pathways. The assignment of approximately 1,961 ± 19 genes was carried out within subsystems, which represent clusters of proteins collaborating to execute specific biological processes or structural complexes (Table S1).

Table 1. Summary of detected genomic variants in lineages of 88 *M. tuberculosis* identified in this study.

Lineages	SNP	Deletions		Insertions	
		more than 20 nt	less than 20 nt	more than 20 nt	less than 20 nt
Euro-American	900	55	4	55	4
Haarlem	877	45	3	55	3
LAM	760	43	4	45	2
mainly T	494	33	2	32	1
S-type	760	37	4	47	3
X-type	863	45	6	55	3

SNP: Single nucleotide polymorphism

Pangenome analysis of the 88 Mtb isolates by Panaroo unveiled a pangenome consisting of 4,276 gene families, which comprised 3,304 core genes, 666 accessory genes, and 270 cloud genes (Figure 1A-D). The ratio between core and pangenome size was 0.78, meaning that the genome comprises more than three-quarters of the pangenome (78.1%), indicative of the lowest variability. The analysis of pathways showed that 19.2% of the annotated protein-coding genes were associated with amino acid metabolism, followed by genes related to carbohydrates (16.2%), lipid metabolism (13.6%), and xenobiotics, biodegradation, and metabolism (12.5%). The genes involved in amino acid biosynthesis were found to be conserved and essential for pathogenicity in bacteria, including Mtb (43). Through the BPGA pipeline, our analysis revealed that the pangenome consisted of an average of 4,397 genes. These genes were categorized into a core genome comprising 3,104 genes, ~659 accessory genes, and 767 unique genes. According to pangenome evaluation, the "b" value of 0.0861522 in the power-law regression model suggests that the pangenome of *Mycobacterium tuberculosis* is near to close.

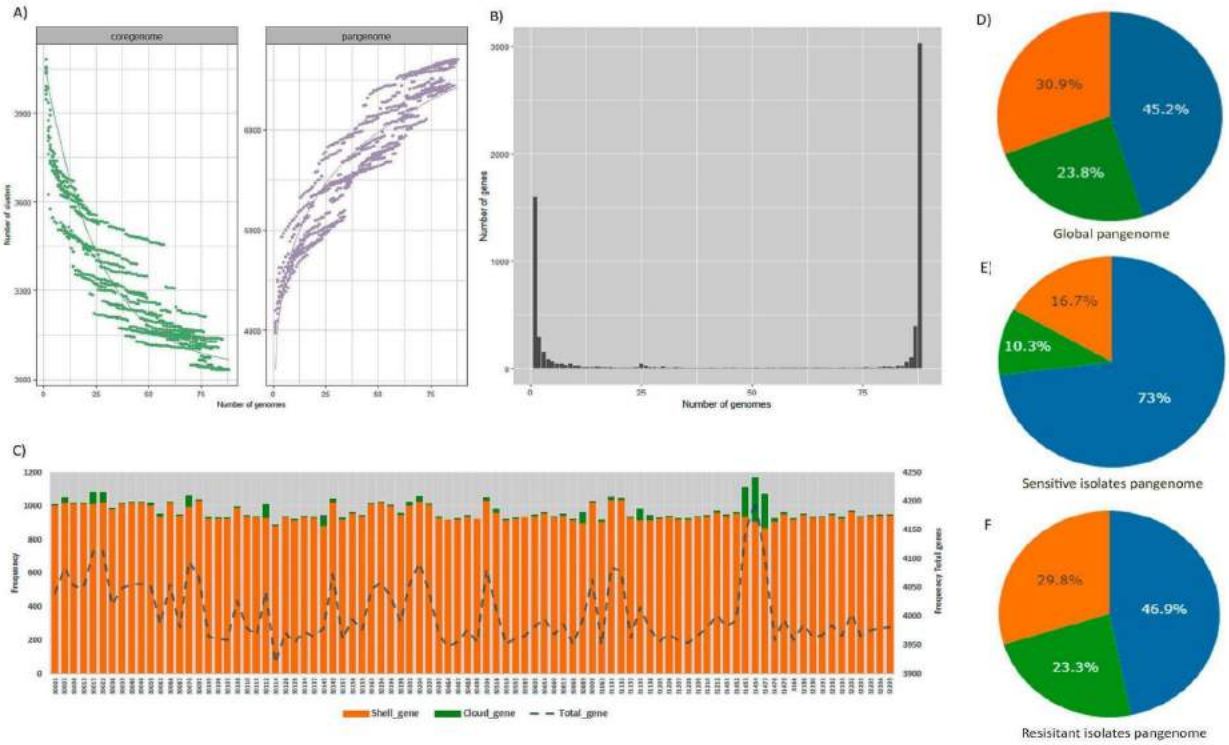

Figure 1. The pangenome of the 88 *Mtb* isolates from Ecuador. A) Rarefaction curves of the core genome (green) and the pangenome (light purple). B) Cluster frequency in the core genome. Each bar represents the core genome composition of an individual isolate, with the global core genome depicted by the final bar after analysis by Roary all isolates. C) Distribution of the accessory, cloud, and total genes. D) Global pangenome composition. E) Pangenome composition for sensitive isolates. F) Pangenome composition for resistant isolates. Colors in pies: core gene (blue), accessory gene (orange), cloud gene (green). The dotted line in panel C represents the total gene for each isolate.

Notably, the three *Mtb* isolates with the highest number of cloud genes belonged to the X-type (S1454 [n=261] and S1477 [n=200]) and LAM (S1453 [n=177]) sublineages (**Figure 1C**), and a similar pattern was identified by the BPGA pipeline in these isolates (S1454 [n=211], S1477 [n=163]), and S1453 [n=144]). Only 6.3% (40 out of 638) of the cloud genes in these isolates were assigned to COG (Clusters of Orthologous Genes), encompassing lipid transport and metabolism, secondary metabolites biosynthesis, transport, and catabolism, amino acid transport and metabolism, and energy production and conversion categories. Notably, 80% of the genes associated with lipid transport and metabolism were found in isolate S1454. In comparison, 75% of the genes associated with energy production and conversion were found in isolate S1477.

A functional analysis among different pangenome partitions based on COG categories globally showed that genes associated with metabolism (45.45% in the core genes to 35.27% in the accessory genes), as well as genes related to cellular processes and signalling (24.97% in the unique gene to 9.79% in the core genes),

presented significant global variations. Additionally, we determined that genes with general function prediction [R] genes (over 10% in all partitions) are the most recurrent, followed by genes associated with secondary metabolite biosynthesis, transport, and catabolism [Q] (10% in accessory and unique genes). Finally, genes associated with cell motility [N] were predominantly present in the accessory gene category.

Furthermore, according to functional assignments based on the KEGG pathways database, our genomes globally exhibited mostly genes associated with metabolism (core and accessory genes), whereas genes associated with environmental information processing were mostly unique genes. Within the core-genome gene distributions, amino acid metabolism (15.0%) had the highest representation, followed by carbohydrate metabolism (14.6%), overview (11.8%), and xenobiotic biodegradation and metabolism (8.9%). Among the accessory genes, carbohydrate metabolism (13.9%) and overview (10.0%) were the most prevalent. In the unique gene category, genes associated with the cellular community, digestive system, energy metabolism, immune system, infectious diseases, and signalling molecules and interactions had a similar distribution, each accounting for 8.7%.

3.2. Pangenome Diversity on Lineage of *Mycobacterium tuberculosis*.

When we analyzed the core genome composition based on the drug resistance pattern, the sensitive isolate (91.0% by Panaroo and 83.0% by BPGA) had a closer pangenome than the resistant isolates (80.1% by Panaroo and 60.7% by BPGA) (Figure 1E, F). By lineage, the LAM and X-type isolates had a higher number of cloud genes (463 and 563, respectively); however, by BPGA, the results differed slightly in the core genome, and the largest changes were found mostly in cloud/unique genes (Table 2).

Table 2. Pangenome composition according to drug resistance and lineage features of the 88 *Mtb* isolates from Ecuador.

		Panaroo pipeline				***BPGA pipeline***			
	n	**Core genes**	**Accessory gene**	**Cloud gene**	**Pangenome**	**Core genes**	**Accessory gene**	**Unique gene**	**Pangenome**
Total	88	3340	666	270	4276	3104	659	767	4397
Genomic resistance									
Resistant	64	3428	648	206	4282	3123	646	756	5145
Sensitive	24	3763	265	106	4134	3477	255	198	4188
Lineages									
LAM	39	3618	451	217	4286	3330	443	463	4649

X-type	21	3617	450	171	4238	3352	392	563	4655
S-type	10	3881	141	12	4034	3589	150	239	4095
Haarlem	9	3827	224	16	4067	3540	209	190	4146
Mainly-T	5	3818	254	0	4072	3618	112	194	3990
Euroamerican	4	3954	126	0	4080	3616	144	118	3985

A closer analysis of the distribution within the COG database and KEGG pathways reveals that the diversity in lineage comes mainly from accessory and unique genes. When analyzing the functional annotation of COG major categories for accessory genes, we found that Euroamerican isolates had the highest percentage of information storage and processing genes (29.14%), while the S-type lineage had a higher percentage of cellular processes and signalling genes (31.16%), and the X-type lineage exhibited the highest percentages of metabolism and poorly characterized genes (38.29% and 20.0%, respectively). In terms of unique genes, the S-type lineage showed the highest representation of information storage and processing genes (33.24%), followed by Euroamerican with cellular processes and signalling genes (30.30%). Metabolism had the highest representation in the LAM lineage (48.13%), and the X-type lineage had the highest unique poorly characterized genes (20.33%). Reviewing the distribution of COG functional categories, we found that genes related to [Q] secondary metabolite biosynthesis, transport, and catabolism genes (found in all six lineages) and [N] cell motility genes (presented predominantly in Haarlem, LAM, mainly T, S-type, and X-type) were the most prevalent, accounting for more than 10% in each lineage. Additionally, [R] General function prediction genes were exclusively present in LAM, S-type, and X-type lineages with the highest percentage (Figure 2, Table S2).

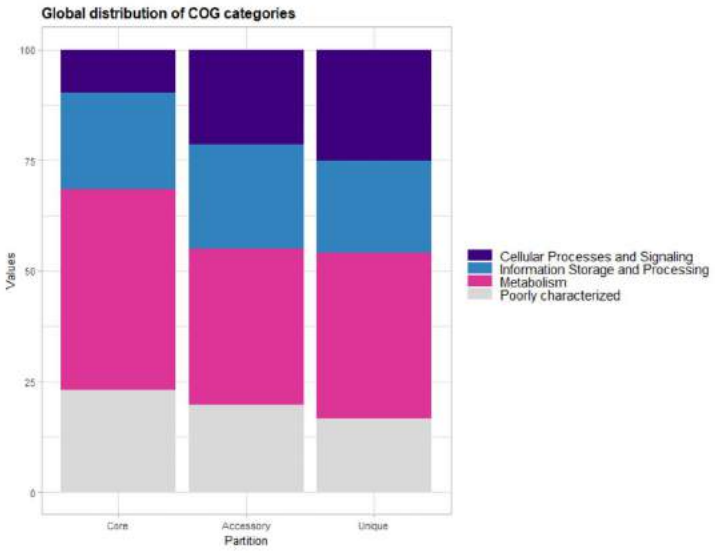

Figure 2. Distribution of major COG categories. Bars represent the percentage distribution of the functional COG annotations among the genes highly conserved in the core, accessory, and unique partitions from all lineages analyzed in this study.

In unique genes, the Euroamerican isolate exhibited a greater diversity, primarily composed of genes related to cell cycle control, cell division, chromosome partitioning, cell motility, intracellular trafficking, secretion, vesicular transport, energy production, and conversion (12.9% each). Additionally, a significant portion of their unique genes were associated with secondary metabolite biosynthesis, transport, and catabolism (21.65%). On the other hand, the X-type lineage had a notable representation of genes involved in carbohydrate transport and metabolism as well as secondary metabolite biosynthesis, transport, and catabolism (14.79% each). Furthermore, a proportion of their unique genes fell into the category of Function unknown (11.09%) (Figure 3, Table S3).

From Global Metabolic KEEG pathways, the mainly T isolate significantly contributed to the accessory partition. It had the highest percentage of genes related to Environmental Information Processing (33.3%), followed by Organismal Systems (25.0%), and Cellular Processes (12.5%). Metabolism had the highest representation, with percentages of 57.96% in the X-type lineage, 54.14% in LAM, and 53.57% in Haarlem. Human Diseases accounted for 13.64% of genes in the S-type lineage, and Genetic Information Processing represented 11.05% in LAM. For unique genes, all lineages exhibited a significant presence of genes related to Environmental Information Processing, Human Diseases, and Organismal Systems, followed by Cellular Processes (Euroamerican, Haarlem, LAM, and S-type). Metabolism genes were notably represented in the LAM, mainly-T, and S-type lineages. However, only the X-type lineage showed the highest percentage of genes associated with genetic information processing (15.38%) (Figure 4, Table S2).

However, a more detailed analysis of accessory genes, based on the KEGG pathway distribution, revealed that carbohydrate metabolism genes exhibited a higher level of diversity. The Haarlem lineage represented 14.3% carbohydrate metabolism genes, followed by LAM with 12.7% and X-type with 12.10%. Similarly, signal transduction genes displayed significant variability, with the mainly-T isolate having 20.8%, S-type at 13.6%, and Euroamerican at 11.5%. Notably, the mainly-T and S-type isolates lacked genes associated with carbohydrate metabolism (Figure 5a).

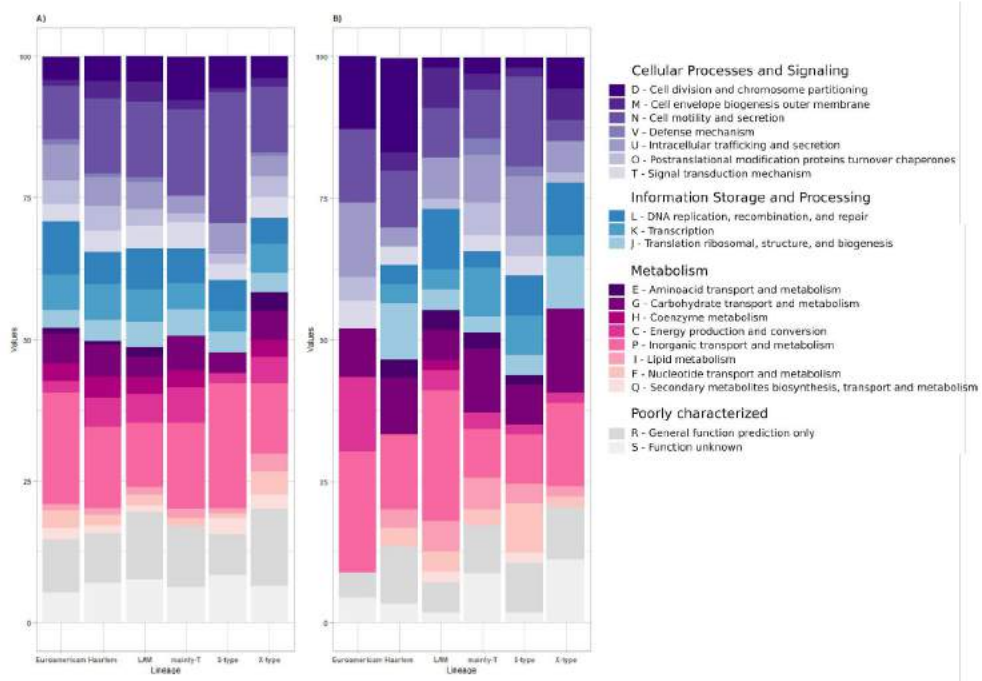

Figure 3. Distribution of Primary COG categories. Bars represent the percentage distribution of the functional annotations among the genes highly conserved on A) the accessory partition; B) the unique genes from all lineages analyzed in this study. **Functional annotation groups.** *Cellular Processes and Signalling*. D: Cell division and chromosome partitioning; M: Cell envelope biogenesis outer membrane; N: Cell motility and secretion; V: Defense mechanism; U: Intracellular trafficking and secretion; O: Posttranslational modification protein turnover chaperones; T: Signal transduction mechanism. *Information Storage and Processing*. L: DNA replication, recombination, and repair; K: Transcription; J: Translation ribosomal, structure, and biogenesis. *Metabolism*. E: Amino acid transport and metabolism; G: Carbohydrate transport and metabolism; H: Coenzyme metabolism; C: Energy production and conversion; P: Inorganic transport and metabolism; I: Lipid metabolism; F: Nucleotide transport and metabolism; Q: Secondary metabolite biosynthesis, transport and metabolism. *Poorly characterized*. S: Function unknown; R: General function prediction only.

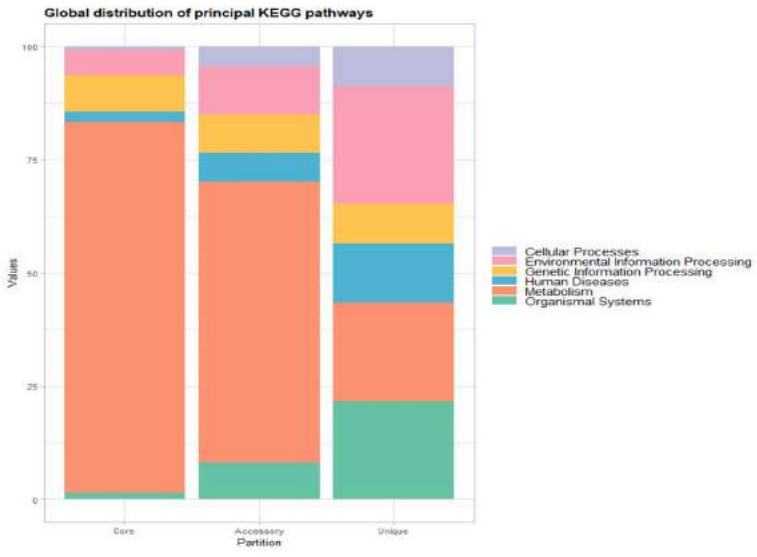

Figure 4. General distribution of KEGG pathways. Differences in the percentage of the functional KEGG pathways annotated in the core, accessory, and unique partitions from all lineages analyzed in this study.

Among the unique genes, infectious diseases were prevalent across almost all lineages (except mainly T), followed by the cellular community, digestive system, immune system, signalling molecules and interaction, and signal transduction. The Euroamerican and Haarlem isolates represented 11.3% of these categories, while the LAM lineage had 11.11%, and the S-type lineage had 11.43% (increasing to 14.29% in signal transduction genes). It is worth noting that several pathways in the case of accessory genes did not involve any genes (Figure 5b, Table S4).

3.3. Genetic diversity of the beta-lactamase resistance-associated genes

To identify the existence of mutations that may affect genes related to beta-lactamase susceptibility, we analyzed the 88 Ecuadorian *Mtb* genomes to identify the SNPs present in 46 genomics regions commonly identified as beta-lactamase resistance-associated and cell wall biosynthesis genes of the MTBC[(44)]. In all isolates, 107 SNPs were identified inside 46 genomic regions. The genes *murD* (132 SNPs), *ddlA* (86 SNPs), *rpfC* (30 SNPs), *rpfE* (99 SNPs), *ponA1* (95 SNPs), *dacB2* (29 SNPs), *ftsK* (51 SNPs), *Rv0192* (93 SNPs), and *Rv0008c* (104 SNPs) showed the highest accumulation of SNPs distributed into various substitutions among the characterized resistance profiles (Table 4 and Table S5). These genes are primarily associated with peptidoglycan biosynthesis. Typically, the MDR isolates presented the highest number of mutations. The lineages that frequently exhibited a higher number of SNPs were LAM and S-type. Six substitutions were present in all isolates, one synonymous (*Ala189Ala* in *eccA2*) and five nonsynonymous (*Arg247Gly* in *murD, Thr365Ala* in *ddlA, Arg126Gln* in *rpfE, Ser127Pro* in *Rv0192, and Ser145Pro* in *Rv0008c*).

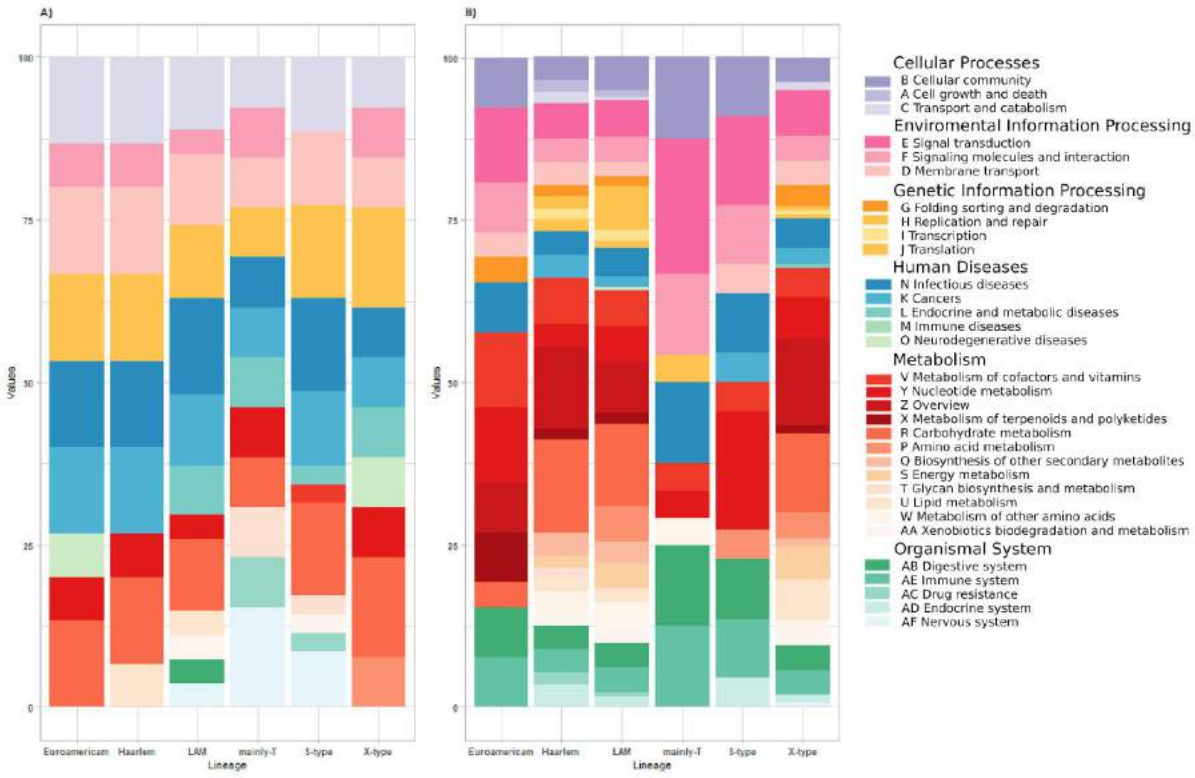

Figure 5. Distribution of primary KEGG pathways. A) Percentage of the functional annotations among the genes highly conserved among the unique genes from all lineages analyzed; B) Percentage of the functional annotations among the genes highly conserved among the accessory genes from all lineages analyzed in this study.

The *ftsk* gene (Rv2748c) displayed the highest number of substitutions, with six nonsynonymous and two synonymous substitutions. Among these, the *Met298Val* substitution was the most frequent (47.0%, 24/51) and was only identified in LAM isolates. In *murD* (Rv2155c), among the three nonsynonymous mutations, the most frequent mutation was *Arg247Gly*, which was predominantly present in MDR isolates (52/88 isolates), with LAM and S-type isolates accounting for 67% of the cases. On the other hand, *Phe76Leu* was exclusively present in S-type isolates. The *ponA1* (Rv0050) gene exhibited five substitutions, with *Pro631Ser* (43/95) and *Ala244Ala* (33/95) being the most recurrent substitutions.

Additionally, for substitutions distributed in all isolates, mutations were identified in different frequencies according to the lineages as follows: In the LAM isolates, more than 76% had mutations in genomic regions *pknA, mmaA4, hisI, rpfC, cut3, papA1, ponA1*, Rv0791c, Rv0948c, Rv1987, Rv2022c, Rv3057c, and Rv3365c. For the Haarlem isolates, 100% presented mutations in genomics regions *pknA, Rv1128c, pheT, hsdM, Rv3057c, cut3, eccA2, murD, Rv3915, and ponA1*, particularly a deleterious effect mutation in the eccA2 gene (Gln460*). Among S-type isolates, mutations in genomics regions *pheT, murD, rpfE, ftsK, ftsH,* and *Rv0008c* were found in more than 90% of isolates. Finally, in X-type isolates, mutations in *Rv0324, Rv1128, hsdM, cut3, eccA2, murD, and ponA1* were identified in 100% of them, but *lpqK, glmU, rpfB, and ponA1* (*Thr58Ala*) were identified in 76% of isolates.

Table 4. Principal synonymous and nonsynonymous mutations identified in genes related to □-lactamase resistance.

Function	Functional categories	Locus	gene name	AA change
	Regulatory proteins	Rv0015c	pknA	Ser385Arg (agc/agG)
	Regulatory proteins	Rv0324		Thr168Ala (act/Gct)
	Lipid metabolism	Rv0642c	mmaA4	Asn165Ser (aac/aGc)
	Conserved hypotheticals	Rv0791c	-	Ser100Cys (tcc/tGc)
	Intermediary metabolism and respiration	Rv0948c	-	Lys59Thr (aag/aCg)
	Insertion seqs and phages	Rv1128c	-	Glu270Gly (gaa/gGa)
	Intermediary metabolism and respiration	Rv1606	hisI	Thr99Ile (acc/aTc)
	Cell wall and cell processes	Rv1987	-	Ser36Asn (agt/aAt)
	Conserved hypotheticals	Rv2022c	-	Val118Ala (gtg/gCg)
	Information pathways	Rv2756c	hsdM	Leu306Pro (ctg/cCg)

	Intermediary metabolism and respiration	Rv3057c	-	Asp112Ala (gat/gCt)/
				His111His (cac/caT)
	Conserved hypotheticals	Rv3365c	-	Ala266Thr (gcg/Acg)
	Cell wall and cell processes	Rv3451	cut3	Gly209Asp (ggc/gAc)/
				Leu259Arg (ctg/cGg)
	Lipid metabolism	Rv3824c	papA1	Leu35Phe (ctt/Ttt)
	Cell wall and cell processes	Rv3884c	eccA2	Ala189Ala (gcc/gcG)
PG synthesis	Cell wall and cell processes	Rv2155c	murD	Arg247Gly (cgg/Ggg)
PG synthesis	Cell wall and cell processes	Rv2981c	ddlA	Thr365Ala (aca/Gca)
PG hydrolysis	Cell wall and cell processes	Rv1884c	rpfC	His16Arg (cac/cGc)
PG hydrolysis	Cell wall and cell processes	Rv2450c	rpfE	Arg126Gln (cgg/cAg)
PG assembly	Cell wall and cell processes	Rv0050	ponA1	Ala244Ala (gca/gcG)
PG assembly	Conserved hypotheticals	Rv0192	-	Ser127Pro (tcg/Ccg)
Cell division	Cell wall and cell processes	Rv0008c	-	Ser145Pro (tcc/Ccc)
Cell division	Cell wall and cell processes	Rv2748c	ftsK	Met298Val (atg/Gtg)

4. Discussion

Several studies have been conducted to identify the causes of specific clinical behavior in different isolates of *Mycobacterium tuberculosis*[(45)]. However, depending on the geographical location, different results have been obtained(46–48). The present study seeks to characterize the genetic composition of Ecuadorian isolates and ascertain whether this pangenome displays variations associated with lineages or specific resistance patterns. To our knowledge, this is the first comprehensive study analyzing the pangenome composition in Ecuadorian isolates. At the time of this study, a single investigation had been undertaken, involving 21 isolates[(33)], highlighting the necessity for an expanded collection of sequenced isolates from Ecuador.

Pangenome studies provide evidence of gene loss or gain events within a species, which depend on the environment in which they thrive [(49)]. In the case of *M. tuberculosis*, its environment is restricted, resulting in limited diversity[(50)]. Our analyses using Roary and BPGA reveal significant disparities in pangenome determination. Among the 88 *M. tuberculosis* isolates examined, the pangenome derived from the Roary algorithm encompasses 48% more accessory genes and unique genes compared to the results from the BPGA algorithm. Furthermore, the core genome constitutes 45% (Roary) or 70% (BPGA) of the total pangenome. These findings not only diverge from each other but also deviate from prior reports, which describe a core genome accounting for 25% [(51)], approximately 86% (52), or approximately 74%[(32)] of the pangenome. The Roary assignment suggests that the accessories gene will continue to increase.

Different studies attribute variations in the number of annotated and assigned genes within pangenome partitions to the different thresholds used by the algorithms for gene annotation, clustering, and classification[(29,31,53,54)] Therefore, it is important to standardize the processes for defining the pangenome and its partitions, especially for prokaryotes.

Functional analysis showed that among the lineages identified in our study, gene assignments were predominantly associated with metabolism (~45%) in the core genome, specifically lipid (~7.8%) and amino acid (~7.3%) transport and metabolism. In the accessory and unique genes, assignments were related to secondary metabolite biosynthesis, catabolism, and transport (~15.9%), particularly in the LAM and S-type lineages. This is consistent with other studies where major genes are associated with lipid metabolism[(32,52)]. In fact, lipid metabolism holds significant relevance for *M. tuberculosis* primarily due to its critical role as a component of the cellular membrane and mass, along with its involvement in vital processes such as cell invasion, evasion of the immune system, virulence, and growth retardation. On the other hand, the identification of accessory and unique genes mostly associated with biosynthesis, catabolism, and transport of secondary metabolites highlights the adaptive capacity of *M. tuberculosis* to the host or external environment[(45,55,56)]. These factors are associated with differences in transmissibility and virulence characteristics observed in the members of the *M. tuberculosis complex*, playing a crucial role in disease recurrence[(32,46)].

Studies have revealed that isolates from the LAM lineage (4.3.4.2) exhibit increased susceptibility to beta-lactams, while the Beijing and Haarlem lineages (4.1.2.1) display high resistance. Our results identified the highest number of mutations in isolates of LAM and S-type sublineages, suggesting that mutations in cell wall biosynthesis genes could be associated with higher resistance to beta-lactams, which should be further investigated if the inclusion of beta-lactams is included. No mutations in the blaC gene were detected. Mutations in Rpfs (resuscitation-promoting factors) related to PG hydrolases have been associated with increased outer membrane permeability and susceptibility to beta-lactams[(58,59)]. Similarly, a pattern was observed with substitutions in EccA2, an ESX-2 type VII secretion system component. Our findings suggest that Ecuadorian isolates related to Euroamerican and Haarlem sublineages had higher resistance to amoxicillin and meropenem because they had the substitution Glu215Gly. Conversely, mutations in lpqK, a conserved lipoprotein with similarity to PBPs, and RpfC evidence increased susceptibility in the strains. In our study, we identified the substitutions Glu67Lys and His16Arg in the lpqK and rpfC genes, respectively, mostly in LAM isolates. Considering the possible uses of beta-lactams in TB treatment, further investigation is warranted to identify the broadest susceptibility profile of this drug class. This will ensure that patients receive timely and appropriate treatment.

Declarations

Ethics approval and consent to participate.

Consent of publication

Not applicable

Data Availability

Raw reads sequences analyzed in this study can be found in the SRA database under the following accession numbers: PRJNA827129.

Funding

GM is a doctoral student in the PEDECIBA program. LB is a member of PEDECIBA and the Sistema Nacional de Investigadores (SNI) of ANII. The funders had no role in the study design, data collection, analysis, the decision to publish, or the preparation of the manuscript.

Author Contributions

GML conceived and designed the study, performed data analysis (bioinformatic processing of the raw sequencing data), and drafted the original manuscript. DAM and JCFC performed data analysis and wrote the draft manuscript. LB conceived and designed the study, wrote the main manuscript, and reviewed it. Funding acquisition was performed by GM and DAM. All the authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgements

Not applicable

References

1. World Health Organization. Global Tuberculosis Report, 2022 [Internet]. 2022 [citado 6 de noviembre de 2022]. Disponible en: https://www.who.int/teams/global-tuberculosis-programme/tb-reports

2. Jong BC de, Antonio M, Gagneux S. Mycobacterium africanum—Review of an Important Cause of Human Tuberculosis in West Africa. PLoS Negl Trop Dis. 28 de septiembre de 2010;4(9):e744.

3. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature. 11 de junio de 1998;393(6685):537-44.

4. Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, et al. The complete genome sequence of Mycobacterium bovis. Proc Natl Acad Sci U S A. 24 de junio de 2003;100(13):7877-82.

5. van Ingen J, Rahim Z, Mulder A, Boeree MJ, Simeone R, Brosch R, et al. Characterization of Mycobacterium orygis as M. tuberculosis Complex Subspecies. Emerg Infect Dis. abril de 2012;18(4):653-5.

6. Macedo R, Isidro J, Gomes MC, Botelho A, Albuquerque T, Sogorb A, et al. Animal-to-human transmission of Mycobacterium pinnipedii. Eur Respir J. 30 de julio de 2020;2000371.

7. Boritsch EC, Brosch R. Evolution of Mycobacterium tuberculosis: New Insights into Pathogenicity and Drug Resistance. Microbiol Spectr. 28 de octubre de 2016;4(5):10.1128/microbiolspec.tbtb2-0020-2016.

8. Napier G, Campino S, Merid Y, Abebe M, Woldeamanuel Y, Aseffa A, et al. Robust barcoding and identification of Mycobacterium tuberculosis lineages for epidemiological and clinical studies. Genome Med. diciembre de 2020;12(1):114.

9. Coll F, McNerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, et al. A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. Nat Commun. diciembre de 2014;5(1):4812.

10. Senghore M, Diarra B, Gehre F, Otu J, Worwui A, Muhammad AK, et al. Evolution of Mycobacterium tuberculosis complex lineages and their role in an emerging threat of multidrug resistant tuberculosis in Bamako, Mali. Sci Rep. diciembre de 2020;10(1):327.

11. Díaz Acosta CC, Russomando G, Candia N, Ritacco V, Vasconcellos SEG, de Berrêdo Pinho Moreira M, et al. Exploring the "Latin American Mediterranean" family and the RDRio lineage in Mycobacterium tuberculosis isolates from Paraguay, Argentina and Venezuela. BMC Microbiol. diciembre de 2019;19(1):131.

12. Ngabonziza JCS, Loiseau C, Marceau M, Jouet A, Menardo F, Tzfadia O, et al. A sister lineage of the Mycobacterium tuberculosis complex discovered in the African Great Lakes region. Nat Commun. diciembre de 2020;11(1):2917.

13. Gonzalo-Asensio J, Pérez I, Aguiló N, Uranga S, Picó A, Lampreave C, et al. New insights into the transposition mechanisms of IS6110 and its dynamic distribution between Mycobacterium tuberculosis Complex lineages. PLoS Genet. abril de 2018;14(4):e1007282.

14. David S, Duarte EL, Leite CQF, Ribeiro JN, Maio JN, Paixão E, et al. Implication of the RD(Rio) Mycobacterium tuberculosis sublineage in multidrug resistant tuberculosis in Portugal. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. octubre de 2012;12(7):1362-7.

15. Tatara MB, Perdigão J, Viveiros M, Kritski A, Silva KE da, Sacchi FPC, et al. Genetic Diversity and Molecular Epidemiology of Mycobacterium tuberculosis in Roraima State, Brazil. Am J Trop Med Hyg. 6 de agosto de 2019;101(4):774-9.

16. Shanmugam SK, Kumar N, Sembulingam T, Ramalingam SB, Selvaraj A, Rajendhiran U, et al. Mycobacterium tuberculosis Lineages Associated with Mutations and Drug Resistance in Isolates from India. Microbiol Spectr. 2022;10(3):e01594-21.

17. Arias APJ, Lahiguera MJ, Borrás R, Cardona CG, Silva MG, López MJV, et al. Comparative study of the genetic diversity of Mycobacterium tuberculosis Complex by Simplified Amplified Fragment Length Polymorphism and Mycobacterial Interspersed Repetitive Unit Variable Number Tandem Repeat Analysis. Rev Ecuat Med Cienc Biológicas [Internet]. 21 de mayo de 2018 [citado 4 de julio de 2022];39(1). Disponible en: https://remcb-puce.edu.ec/remcb/article/view/568

18. Zurita J, Espinel N, Barba P, Ortega-Paredes D, Zurita-Salinas C, Rojas Y, et al. Genetic diversity and drug resistance of Mycobacterium tuberculosis in Ecuador. Int J Tuberc Lung Dis. 1 de febrero de 2019;23(2):166-73.

19. Garzon-Chavez D, Zurita J, Mora-Pinargote C, Franco-Sotomayor G, Leon-Benitez M, Granda-Pardo JC, et al. Prevalence, Drug Resistance, and Genotypic Diversity of the Mycobacterium tuberculosis Beijing Family in Ecuador. Microb Drug Resist. julio de 2019;25(6):931-7.

20. Garzon-Chavez D, Garcia-Bereguiain MA, Mora-Pinargote C, Granda-Pardo JC, Leon-Benitez M, Franco-Sotomayor G, et al. Population structure and genetic diversity of Mycobacterium tuberculosis in Ecuador. Sci Rep. diciembre de 2020;10(1):6237.

21. Garcés E, Cifuentes L, Franco G, Romero-Sandoval N, Arias PJ. Study of the distribution of lineages of Mycobacterium tuberculosis in a prison in Guayaquil, Ecuador. 2023;

22. Yang T, Gan M, Liu Q, Liang W, Tang Q, Luo G, et al. SAM-TB: a whole genome sequencing data analysis website for detection of Mycobacterium tuberculosis drug resistance and transmission. Brief Bioinform. 24 de febrero de 2022;23(2):bbac030.

23. Lam C, Martinez E, Crighton T, Furlong C, Donnan E, Marais BJ, et al. Value of routine whole genome sequencing for Mycobacterium tuberculosis drug resistance detection. Int J Infect Dis. diciembre de 2021;113:S48-54.

24. Wang L, Yang J, Chen L, Wang W, Yu F, Xiong H. Whole-genome sequencing of Mycobacterium tuberculosis for prediction of drug resistance. Epidemiol Infect. 2022;150:e22.

25. Brown AC. Whole-Genome Sequencing of Mycobacterium tuberculosis Directly from Sputum Samples. En: Parish T, Kumar A, editores. Mycobacteria Protocols [Internet]. New York, NY: Springer US; 2021 [citado 18 de junio de 2022]. p. 459-80. (Methods in Molecular Biology). Disponible en: https://doi.org/10.1007/978-1-0716-1460-0_20

26. Torres Ortiz A, Coronel J, Vidal JR, Bonilla C, Moore DAJ, Gilman RH, et al. Genomic signatures of pre-resistance in Mycobacterium tuberculosis. Nat Commun. diciembre de 2021;12(1):7312.

27. Freschi L, Vargas R, Husain A, Kamal SMM, Skrahina A, Tahseen S, et al. Population structure, biogeography and transmissibility of Mycobacterium tuberculosis. Nat Commun. diciembre de 2021;12(1):6099.

28. Iketleng T, Lessells R, Dlamini MT, Mogashoa T, Mupfumi L, Moyo S, et al. Mycobacterium tuberculosis Next-Generation Whole Genome Sequencing: Opportunities and Challenges. Singh S, editor. Tuberc Res Treat. 9 de diciembre de 2018;2018:1298542.

29. Tonkin-Hill G, Corander J, Parkhill J. Challenges in prokaryote pangenomics. Microb Genomics. 2023;9(5):001021.

30. Ferrés I, Iraola G. Protocol for post-processing of bacterial pangenome data using Pagoo pipeline. STAR Protoc. diciembre de 2021;2(4):100802.

31. Negrete-Paz AM, Vázquez-Marrufo G, Gutiérrez-Moraga A, Vázquez-Garcidueñas MS. Pangenome Reconstruction of Mycobacterium tuberculosis as a Guide to Reveal Genomic Features Associated with Strain Clinical Phenotype. Microorganisms. junio de 2023;11(6):1495.

32. Hurtado-Páez U, Álvarez Zuluaga N, Arango Isaza RE, Contreras-Moreira B, Rouzaud F, Robledo J. Pan-genome association study of Mycobacterium tuberculosis lineage-4 revealed specific genes related to the high and low prevalence of the disease in patients from the North-Eastern area of Medellín, Colombia. Front Microbiol [Internet]. 2023 [citado 13 de junio de 2023];13. Disponible en: https://www.frontiersin.org/articles/10.3389/fmicb.2022.1076797

33. Morey-León G, Andrade-Molina D, Fernández-Cadena JC, Berná L. Comparative genomics of drug-resistant strains of Mycobacterium tuberculosis in Ecuador. BMC Genomics. 21 de diciembre de 2022;23(1):844.

34. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. 2017.

35. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 1 de agosto de 2014;30(15):2114-20.

36. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15(3):R46.

37. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 8 de junio de 2017;13(6):e1005595.

38. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS ONE. 19 de noviembre de 2014;9(11):e112963.

39. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 15 de abril de 2013;29(8):1072-5.

40. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 15 de noviembre de 2015;31(22):3691-3.

41. 41. Chaudhari NM, Gupta VK, Dutta C. BPGA- an ultra-fast pan-genome analysis pipeline. Sci Rep. 13 de abril de 2016;6:24373.

42. 42. Kohl TA, Utpatel C, Schleusener V, De Filippo MR, Beckert P, Cirillo DM, et al. MTBseq: a comprehensive pipeline for whole genome sequence analysis of Mycobacterium tuberculosis complex isolates. PeerJ. 13 de noviembre de 2018;6:e5895.

43. 43. Smith I. Mycobacterium tuberculosis Pathogenesis and Molecular Determinants of Virulence. Clin Microbiol Rev. 2003;16(3):463-96.

44. 44. Olivença F, Nunes A, Macedo R, Pires D, Silveiro C, Anes E, et al. Uncovering Beta-Lactam Susceptibility Patterns in Clinical Isolates of Mycobacterium tuberculosis through Whole-Genome Sequencing. Microbiol Spectr. 2022;10(4):e00674-22.

45. 45. Periwal V, Patowary A, Vellarikkal SK, Gupta A, Singh M, Mittal A, et al. Comparative Whole-Genome Analysis of Clinical Isolates Reveals Characteristic Architecture of Mycobacterium tuberculosis Pangenome. Hoshino Y, editor. PLOS ONE. 8 de abril de 2015;10(4):e0122979.

46. 46. Zakham F, Sironen T, Vapalahti O, Kant R. Pan and Core Genome Analysis of 183 Mycobacterium tuberculosis Strains Revealed a High Inter-Species Diversity among the Human Adapted Strains. Antibiotics. 28 de abril de 2021;10(5):500.

47. 47. Warner DF, Koch A, Mizrahi V. Diversity and disease pathogenesis in Mycobacterium tuberculosis. Trends Microbiol. 1 de enero de 2015;23(1):14-21.

48. 48. Gagneux S. Genetic Diversity in Mycobacterium tuberculosis. En: Pieters J, McKinney JD, editores. Pathogenesis of Mycobacterium tuberculosis and its Interaction with the Host Organism [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. p. 1-25. Disponible en: https://doi.org/10.1007/82_2013_329

49. 49. Rouli L, Merhej V, Fournier PE, Raoult D. The bacterial pangenome as a new tool for analysing pathogenic bacteria. New Microbes New Infect. 26 de junio de 2015;7:72-85.

50. 50. Achtman M. Evolution, Population Structure, and Phylogeography of Genetically Monomorphic Bacterial Pathogens. Annu Rev Microbiol. 2008;62(1):53-70.

51. 51. Dar HA, Zaheer T, Ullah N, Bakhtiar SM, Zhang T, Yasir M, et al. Pangenome Analysis of Mycobacterium tuberculosis Reveals Core-Drug Targets and Screening of Promising Lead Compounds for Drug Discovery. Antibiotics. 17 de noviembre de 2020;9(11):819.

52. 52. Yang T, Zhong J, Zhang J, Li C, Yu X, Xiao J, et al. Pan-Genomic Study of Mycobacterium tuberculosis Reflecting the Primary/Secondary Genes, Generality/Individuality, and the Interconversion Through Copy Number Variations. Front Microbiol [Internet]. 2018 [citado 20 de junio de 2022];9. Disponible en: https://www.frontiersin.org/article/10.3389/fmicb.2018.01886

53. 53. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol. 22 de julio de 2020;21:180.

54. 54. Tettelin H, Riley D, Cattuto C, Medini D. Comparative genomics: the bacterial pan-genome. Curr Opin Microbiol. 1 de octubre de 2008;11(5):472-7.

55. 55. Sapriel G, Brosch R. Shared Pathogenomic Patterns Characterize a New Phylotype, Revealing Transition toward Host-Adaptation Long before Speciation of Mycobacterium tuberculosis. Genome Biol Evol. 1 de agosto de 2019;11(8):2420-38.

56. 56. Kweon O, Kim SJ, Blom J, Kim SK, Kim BS, Baek DH, et al. Comparative functional pan-genome analyses to build connections between genomic dynamics and phenotypic evolution in polycyclic aromatic hydrocarbon metabolism in the genus Mycobacterium. BMC Evol Biol. 14 de febrero de 2015;15(1):21.

57. 57. Muzondiwa D, Hlanze H, Reva ON. The Epistatic Landscape of Antibiotic Resistance of Different Clades of Mycobacterium tuberculosis. Antibiotics. 15 de julio de 2021;10(7):857.

58. 58. Wivagg CN, Hung DT. Resuscitation-Promoting Factors Are Required for β-Lactam Tolerance and the Permeability Barrier in Mycobacterium tuberculosis. Antimicrob Agents Chemother. 14 de febrero de 2012;56(3):1591-4.

59. 59. Hett EC, Chao MC, Deng LL, Rubin EJ. A Mycobacterial Enzyme Essential for Cell Division Synergizes with Resuscitation-Promoting Factor. PLOS Pathog. 29 de febrero de 2008;4(2):e1000001.

Supplementary Materials: The following supporting information can be downloaded at Supplementary_tables: Table S1. Genomic and clinical data from Ecuadorian isolates; Table S2. Distribution of major COG and KEGG pathways from lineage and pangenome partitions; Table S3. Distribution of primary COG categories from lineage and pangenome partitions; Table S4. Distribution of primary

KEGG pathways from lineage and pangenome partitions; Table S5. Synonymous and nonsynonymous mutations identified in genes related to □-lactamase resistance.

DISCUSIÓN

La distribución de los diferentes linajes de *M. tuberculosis* se ha asociado a eventos migratorios en el tiempo. Basado en los procesos de descubrimiento y colonización de otros continentes, se ha evidenciado que la introducción de *M. tuberculosis* en las Américas podría estar relacionada con la colonización de 1490. Sin embargo, otros estudios sugieren un origen más antiguo, posiblemente relacionado con el consumo de animales silvestres contaminados, lo cual contribuyó a la adaptación humana. En América, se ha reportado una diversidad de linajes de *M. tuberculosis* asociados a áreas geográficas. El linaje L4 (Euroamericano) ha sido identificado como el más predominante, aunque se observa variación en la proporción de sublinajes entre diferentes países.

En Argentina existe un predominio de las cepas M y Ra(153). En Brasil, el linaje 4 varía entre el 25% y el 99%(154–156). En Colombia, se ha reportado una prevalencia del linaje 4 entre el 85 y el 99%(35,157,158). En México, se ha observado que el linaje 4 representa entre el 69 y el 98% de los casos(159,160) mientras que en Perú alcanza el 90%(161–163). Nuestros resultados, basados en el análisis de 88 aislados secuenciados, mostraron que los linajes más predominantes fueron LAM (39/88), X-type (21/88), S-type (10/88), Haarlem (9/88), mainly-T (5/88) y Euroamericano (4/88). Estos resultados son consistentes con investigaciones previas que utilizaron técnicas como MIRU-VNTR de 24-loci, Spoligotyping y PCR específica de SNP. Estos estudios también han informado un predominio de los sublinajes LAM, Haarlem y mainly-T, e incluso han detectado la presencia del linaje Beijing, lo cual resalta la importancia de aumentar la cantidad de aislados evaluados para una mejor comprensión de la diversidad genética de *M. tuberculosis*.

En microorganismos bacterianos la transferencia y preservación de genes es fundamental para procesos de adaptabilidad, supervivencia, virulencia además de procesos conjuntos de evolución(25). Además la identificación de la funcionalidad de los genes involucrados permite entender los diferentes grados de patogenicidad, virulencia o resistencia a fármacos entre los microorganismos estudiados(49,164). En el caso de *M. tuberculosis*, se ha demostrado que varios de sus procesos primordiales están asociados con las diferencias genéticas entre sus integrantes y que ciertos genes han

estado bajo selección positiva(165). Se sabe además, que la organización genómica y composición génica de *M. tuberculosis* es relativamente estable. Esto se debe a la ausencia de transferencia horizontal de genes y a su naturaleza clonal. Sin embargo, los análisis de core y pangenomas han mostrado una variabilidad más amplia(166–170). Los estudios de pangenoma evidencian los eventos de pérdida o ganancia de genes dentro de una especie lo cual depende del ambiente en el cual se desarrollan(171), en el caso de *M.tuberculosis*, su ambiente es restringido generando poca diversidad. Nuestros resultados mostraron que entre 88 aislados de *M. tuberculosis,* analizados se identificó estar compuesto por un core y pangenoma de 3340 y 4276 genes cuando se analiza por Panaroo, mientras que al emplear algoritmo para Análisis de Pangenoma Bacteriano (BPGA, por su siglas en inglés) se identificaron 3104 y 4397, respectivamente, considerándose que el core genoma representa el 78,1% (Panaroo) o 70,5% (BPGA) del pangenoma, los cuales difieren o se encuentran cercanos de estudios que identificaron el core genoma representó el 25%(151), ~86%(167), o ~74%(172) del pangenoma, diferentes estudios relacionan que las variaciones en la cantidad de genes anotados y asignadas dentro de las particiones del pangenoma se deben en los diferentes umbrales empleados por los algoritmos para anotación, clusterización y clasificación de los genes, por lo cual es importante estandarizar los proceso para la definición del pangenoma y sus particiones, en especial para procariotas(173,174).

Funcionalmente entre los linajes identificados en nuestro estudio se determinó, a nivel de core genoma, así como genes accesorios y únicos, la asignación de genes fue mayoritariamente relacionada con genes del Metabolismo (~45%), principalmente vinculados con metabolismo y transporte de lípidos (~7,8%), metabolismo y transporte de aminoácidos (~7,3%) en el core genoma, biosíntesis, catabolismo y transporte de metabolitos secundarios(~15,9%) en genes accesorios y únicos, especialmente en los linajes LAM y S-type, esto concuerda con otros estudios donde los principales genes están asociados con metabolismo de lípidos(167,172) lo cual permite entender además de la importancia como componente de la membrana y masa celular, su rol dentro del procesos como invasión celular, evasión del sistema inmune, virulencia y retraso del crecimiento; por otro lado la identificación de genes accesorios y únicos asociados con biosíntesis, catabolismo y transporte de metabolitos secundarios resalta las capacidad de adaptaciones de *M.tuberculosis* al huésped o del medio externo(170,175,176).

Además, nuestro análisis del pangenoma de las cepas de *M. tuberculosis* de Ecuador revela una característica diversidad genética dentro de los aislamientos circulantes, en especial de genes accesorios y único, los cuales dependiendo del linaje contribuiría en facilitar las interacciones con su huésped humano, particularmente en el contexto de la persistencia de micobacterias, competencia por nutrientes, evasión de inmunidad, lo cual estaría asociado a las diferencia en niveles de transmisibilidad, virulencia característico de los integrantes del linaje L4 de *M. tuberculosis* desempeñando un papel importante en la recurrencia de la enfermedad(168,172).

Aproximadamente 731 ± 100 polimorfismos de un solo nucleótido (SNPs), además de ~51 ± 7.6 Inserciones, y ~47± 7.0 delecciones, en la cual la mayoria de InDels fueron pequeños (≤20 nt) fueron identificados en los 88 aislados de nuestro estudio, encontrándose que los aislados S2192, S0516 y S0654 presentaron más de 1000 SNPs con mutaciones consideradas dentro de genes relacionados a resistencia debido sus interacciones epistáticas con otros genes(177), sin considerar las variaciones de los genes PE/PPE, por otro lado entre todos los linajes, los genes genes que presentaron más de 5 mutaciones en promedio fueron *plcA, ppsA, pks12, plcB y esxP*, se ha evidenciado que los genes del loci plcABC y esxP tienen una actividad de inmunomodulación siendo importantes para la patogenicidad de *M. tuberculosis*(178,179). mientras que ppsA está relacionado con resistencia al ácido pirazinoico, la forma activa de la pirazinamida(180), por otro lado pks12 está envuelta en la biosíntesis de mannosil-β-1-fosfomicocetido (MPM) producido principalmente por las micobacterias de crecimiento lento y se cree está involucrado en su patogenicidad(181). La modificación o inactivación que pueda generarse como consecuencia de las mutaciones en estos genes, podrían afectar la virulencia de *M. tuberculosis*, lo que llevaría posiblemente a presentar una enfermedad con menor grave para el huésped, lo que resulta en un daño pulmonar reducido y una progresión prolongada de la enfermedad, aumentando las posibilidades de transmisión.

Es bien conocido que el tratamiento de primera línea en *M. tuberculosis*, comprende la administración de fármacos como Rifampicina, Isoniazida, Etambutol y Pirazinamida, mientras que para la segunda línea se incluyen estreptomicina, las fluoroquinolonas y/o los aminoglucósidos inyectables (kanamicina, capreomicina, amikacina). Los diferentes esquemas de medicamentos, así como la baja adherencia al tratamiento, han conllevado que se desarrollen problemas de resistencia a estos fármacos

en diferentes niveles. La resistencia a drogas en *M. tuberculosis* está principalmente mediada por mutaciones en genes específicos. Sin embargo, la detección de estas mutaciones puede variar dependiendo de las pruebas moleculares utilizadas,y se debe tener en cuenta la posibilidad de heterorresistencia en algunos aislados(26,136,142,182). En nuestro estudio, identificamos 73 mutaciones presentes en 19 genes relacionados a la resistencia de 42 genes considerados. Es importante destacar que la presencia de deleciones es de vital importancia debido a su efecto sobre la funcionalidad génica y la transmisibilidad del patógeno(72,99,118,183–189).

En variados estudios, se han identificado genes asociados a la resistencia considerados canónicos, entre ellos, la región RRDR (rifampicin resistance-determining region) en el gen *rpoB* que contribuye a la detección de resistencia a rifampicina(183,190,191). Para Isoniazida, el gen *katG* frecuentemente se identifica en los aislados resistentes(72,73,131). Por su parte, mutaciones en el operón del gen *embCAB* y de la región promotora y del gen *pncA* están vinculadas con resistencias a etambutol(99–103) y pirazinamida(98,99,188,192,193). En nuestro estudio se identificó una alta prevalencia de mutaciones canónicas en los genes antes mencionados, siendo las más prevalentes las mutaciones S450L en *rpoB*, S315T en *katG*, M306I/V, D354A, G406A/D/S, y -16C>T en los genes *embA* y *embB* en el operón *embCAB* y H51R, H82R, y V125F en el gen *pncA*. Además, se identificaron mutaciones compensatorias que también han sido identificadas en aislados resistentes en otros estudios(79,81,194,195).

En relación con drogas de segunda línea, nuestro estudio reveló que los aislados resistentes a fluoroquinolonas presentaron principalmente mutaciones conocidas en el gen *gyrA* en la posición 90 y 94, y un aislado presentó mutaciones en la posición 89. Sin embargo,en la posición 499 en el gen *gyrB* se detectó una sustitución en un aislado identificado como sensible. No se detectaron mutaciones compensatorias que restauren la sensibilidad a las FQ. Estudios previos han reportado que los cambios en las posiciones 90, 94 y 499 están asociados a resistencia a todas las FQs(74,109,196,197). Sin embargo, la presencia de mutaciones relacionadas con aislados fenotípicamente sensibles podría deberse a puntos de corte incorrectos en las pruebas microbiológicas clásicas, heterorresistencia o la presencia de diferentes clones con variados perfiles de mutaciones(198,199). Particularmente se identificó un aislado resistente a Levofloxacina que no presentó ninguna mutación canónica en la

región QRDR, esto sugiere que la resistencia podría estar asociada con otros mecanismos, como por ejemplo la alteración en la bomba de flujo, la cual se ha relacionado con la adquisición de la resistencia(200).

En cuanto a la resistencia a estreptomicina, se observó que el 60.7% de los aislados estudiados presentaban mutaciones en el gen *rpsL*. La sustitución más común fue K43R, lo cual concuerda con hallazgos previos de otros estudios(109,118,201). Por otro lado, diferentes mutaciones en el gen *rrs* (que codifica para el 16S rRNA) han sido relacionadas con resistencia a estreptomicina y/o resistencia cruzada con aminoglucósidos (amikacina, kanamicina,capreomicina), en nuestro estudio identificamos cambios en las posiciones 514A>C, 517C>T, 888G>A, 906A>G y 1401A>G del gen *rrs*, estos cambios ha sido asociado con altos niveles de resistencia acorde a otros estudios(136,156,202–206), mientras que la variación eis:-12C>T está vinculado con resistencia a Kanamicina(186). Nuestro estudio confirma la importancia de monitorear mutaciones presentes en los genes *rpoB, katG*, *embB* y *pncA,* así como la relevancia de las mutaciones compensatorias, con el objeto de prevenir la emergencia de aislados MDR. También resalta la importancia de una detección temprana de mutaciones en la región QRDR, así como en genes *rpsL* y *rrs* u otros en genes relacionados que tienen un alto impacto en salud pública debido a su influencia en la salud del paciente, en los costos del tratamiento y la complejidad del mismo como son los XDR.

La persistencia de la infección por TB puede ser atribuida a procesos de reactivación de tuberculosis latente (LTB) o la reinfección con aislados resistentes autóctonos o de mezclas multinacionales, sin embargo, a pesar de que los aislados puedan presentar una alta relación genética, tienden a diferir acorde a su distribución geográfica, patrones de resistencia, transmisibilidad y virulencia(168). Debido a esta complejidad, el análisis de redes de dispersión usando la información generada mediante secuenciamiento del genoma completo proporciona datos que permite conectar los eventos de transmisión en TB, evidenciando la dinámica temporal y espacial del evento, en el cual si se incluyen datos epidemiológicos aportan a una mejor interpretación del movimiento de los aislados durante un brote epidemiológico (144,146,207,208). En nuestro estudio, empleando un umbral de ≤12 SNPs que indica eventos de transmisión cercanos, se pudo evidenciar, dentro de un análisis que involucró 503 aislados de diferentes países, que 17/88 aislados ecuatorianos, colectados entre 2019 y 2021,

presentaron clonalidad con aislados de *M. tuberculosis* caracterizados en pacientes colombianos y/o migrantes latinoamericanos en España entre 2014 y 2015. En algunos casos los aislados formaron parte de redes de transmisión (S2191) o se encontraban como parte de cluster genéticos (S0110, S2186, S0106), lo cual podría plantear los posibles punto de transmisión local luego de infectarse a partir de los aislados de migrantes ya sea por estancias en ambientes laborales principalmente ocupados por migrantes como son plantaciones, fábricas, restaurantes. Sorpresivamente algunos de los aislados involucrados en estas redes provenian de pacientes de sexo femenino, lo cual podría sugerir que pudo existir alguna relación sentimental que pudo facilitar la transmisión. Nuestros hallazgos aportan a la evidencia que la movilidad migratoria es un factor importante para la incidencia de TB en áreas donde la transmisión es limitada y también puede contribuir a la reactivación de la enfermedad (209,210). Además, estos hallazgos sugieren posibles eventos de transmisión transnacional que involucran a Ecuador y sus países fronterizos, así como destinos frecuentes de migración. La incorporación de WGS y análisis de redes en las intervenciones de vigilancia puede ayudar a identificar cuándo y dónde los pacientes adquirieron la TB dentro de una comunidad inmigrante y orientar las intervenciones de salud pública dirigidas a controlar la transmisión de la TB(211,212).

Diferentes estudios han evidenciado el potencial resolutivo que posee la secuenciación de genoma completo (WGS, por su siglas en inglés) dentro del estudio de diferentes microorganismos, en especial en *M. tuberculosis* para la detección de la diversidad genética asociada a linajes, resistencia a fármacos y dinámicas de transmisión(118,157,213). Sin embargo, para que su aplicación sea aceptada completamente se debe sobrellevar las diferencias que existen especialmente para la estandarización en la identificación de mutaciones no-canónicas asociadas a resistencia(214). En este contexto en 2021, la OMS actualizó las guías de los criterios para la identificación de mutaciones relacionadas y presentes en aislados mono, multi y extremo drogorresistente, con el objeto que pueda ser considerada dentro de los procesos de estandarización en los diferentes algoritmos de análisis de secuencias genómicas(215). Para evaluar la utilidad de los algoritmos, comparamos la exactitud, sensibilidad y especificidad de cinco herramientas bioinformáticas disponibles en la web para inferir la resistencia de *M. tuberculosis* frente al método microbiológico convencional. Nuestros resultados evidenciaron que de las cinco herramientas analizadas, globalmente PhyReSse, TB-rpofiler, Mykrobe, y SAM-TB

mostraron los mejores resultados (97.0% sensibilidad, 100.0% especificidad, y 97.7% exactitud), sin embargo, cuando se analiza la capacidad de detectar resistencia de forma individual a las diferentes drogas empleadas en el tratamientos, existe diferencias entre los resultados generados, probablemente se deba a los diferentes umbrales empleados que consideran para la caracterización de las mutaciones. Nuestros resultados resaltan la aplicabilidad de la secuenciación de genoma completo y análisis bioinformático especialmente los algoritmos disponibles en la web, debido a su facilidad de manejos para el estudio de la resistencia de tuberculosis dentro de los programas de vigilancia locales, pero también pone a consideración se tenga en cuenta los diferentes grados de confidencia que se aplican en la identificación de las diferentes mutaciones para una efectiva detección(216,217), especialmente hasta que la estandarización de criterios se apliquen en los algoritmos.

CONCLUSIONES

La diversidad filogenética de los aislados de *M. tuberculosis* en Ecuador está principalmente concentrada en el linaje Euroamericano (L4), con una destacada presencia de las familias LAM y X-type. No obstante, se detectó la presencia de otros cuatro linajes. La presencia del linaje Beijing y su relación con brotes de aislados pre-XDR y XDR, evidenciado en otros estudios, muestran la necesidad de ampliar el análisis de aislados de tuberculosis mediante secuenciamiento del genoma completo, para develar la real distribución de linajes circulantes en el país.

Se observa una amplia diversidad genética en los aislados obtenidos, con variaciones específicas en los patrones de resistencia a fármacos de primera y segunda línea de tratamiento. Estas variaciones son identificadas de manera más efectiva a través del secuenciamiento masivo, lo que permite un enfoque más preciso en el tratamiento de los pacientes. La identificación de aislados pre-XDR requiere una atención eficiente para evitar la dispersión e incremento de casos que afectaría el programa de vigilancia.

La disponibilidad de herramientas web para el estudio genético de *M. tuberculosis*, con actualizaciones basadas en las definiciones y consideraciones establecidas por la Organización Mundial de la Salud, sería beneficioso para estandarizar la implementación de procesos de vigilancia genómica de la tuberculosis en el país. Es fundamental capacitar al personal en el procesamiento de muestras de laboratorio y en el análisis e interpretación bioinformática de la información, lo que contribuiría al fortalecimiento del sistema de vigilancia nacional.

La movilidad migratoria hacia y desde Ecuador, genera un fuerte impacto en la vigilancia convencional de esta enfermedad. En este sentido la aplicación de herramientas genómicas basadas en el análisis del genoma completo, junto con datos epidemiológicos y migratorios, tanto a nivel interno como externo, podría ser de utilidad para identificar las posibles rutas de transmisión entre las poblaciones. Esto proporcionaría información sobre posibles puntos de ingreso de asilados de importancia en salud pública en la vigilancia de la tuberculosis, contribuyendo a los esfuerzos realizados en la vigilancia como aporte a la estrategia "End-TB 2030".

BIBLIOGRAFÍA

1. Bu Q, Qiang R, Fang L, Peng X, Zhang H, Cheng H. Global trends in the incidence rates of MDR and XDR tuberculosis: Findings from the global burden of disease study 2019. Front Pharmacol [Internet]. 2023 [citado 25 de junio de 2023];14. Disponible en: https://www.frontiersin.org/articles/10.3389/fphar.2023.1156249
2. Bagcchi S. WHO's Global Tuberculosis Report 2022. Lancet Microbe. 1 de enero de 2023;4(1):e20.
3. World Health Organization. Global Tuberculosis Report, 2022 [Internet]. 2022 [citado 6 de noviembre de 2022]. Disponible en: https://www.who.int/teams/global-tuberculosis-programme/tb-reports
4. TB disease burden [Internet]. [citado 25 de junio de 2023]. Disponible en: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022/tb-disease-burden
5. World Bank. World Bank Open Data. 2023 [citado 25 de junio de 2023]. Incidencia de tuberculosis mundial. Disponible en: https://data.worldbank.org
6. Jones AJ, Jones-López EC, Butler-Wu SM, Wilson ML, Rodman J, Flors L, et al. Impact of COVID-19 on diagnosis and testing for TB in a high-resource, low-burden setting. Int J Tuberc Lung Dis. septiembre de 2022;26(9):888-90.
7. Comella-del-Barrio P, De Souza-Galvão ML, Prat-Aymerich C, Domínguez J. Impact of COVID-19 on Tuberculosis Control. Arch Bronconeumol. 1 de abril de 2021;57:5-6.
8. Martins M, Carvalho L, Carvalho T, Gomes I. Impact of the COVID-19 pandemic on in-hospital diagnosis of tuberculosis in non-HIV patients. Pulmonology. 1 de noviembre de 2022;28(6):481-3.
9. Kumar R, Singhal M, Kumar D, Joshi A, Islam KM. Designing, Development, and Evaluation of an Informatics Platform for Enhancing Treatment Adherence in Latent Tuberculosis Infection Patients: A Study Protocol. BioMedInformatics. marzo de 2023;3(1):252-9.
10. Houben RMGJ, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. PLOS Med. 25 de octubre de 2016;13(10):e1002152.

11. Al Abri S, Kasaeva T, Migliori GB, Goletti D, Zenner D, Denholm J, et al. Tools to implement the World Health Organization End TB Strategy: Addressing common challenges in high and low endemic countries. Int J Infect Dis. marzo de 2020;92:S60-8.
12. Satyanarayana S, Thekkur P, M. V. Kumar A, Lin Y, A. Dlodlo R, Khogali M, et al. An Opportunity to END TB: Using the Sustainable Development Goals for Action on Socio-Economic Determinants of TB in High Burden Countries in WHO South-East Asia and the Western Pacific Regions. Trop Med Infect Dis. 18 de junio de 2020;5(2):101.
13. Nataraj V, Varela C, Javid A, Singh A, Besra GS, Bhatt A. Mycolic acids: deciphering and targeting the Achilles' heel of the tubercle bacillus. Mol Microbiol. octubre de 2015;98(1):7-16.
14. PaweŁczyk J, Kremer L. The Molecular Genetics of Mycolic Acid Biosynthesis. Microbiol Spectr. 22 de agosto de 2014;2(4):10.1128/microbiolspec.mgm2-0003-2013.
15. Dhar N, McKinney J, Manina G. Phenotypic Heterogeneity in Mycobacterium tuberculosis. Microbiol Spectr. 11 de noviembre de 2016;4(6):10.1128/microbiolspec.tbtb2-0021-2016.
16. Yusoof KA, García JI, Schami A, Garcia-Vilanova A, Kelley HV, Wang SH, et al. Tuberculosis Phenotypic and Genotypic Drug Susceptibility Testing and Immunodiagnostics: A Review. Front Immunol. 7 de julio de 2022;13:870768.
17. Beste DJV, Espasa M, Bonde B, Kierzek AM, Stewart GR, McFadden J. The Genetic Requirements for Fast and Slow Growth in Mycobacteria. PLoS ONE. 28 de abril de 2009;4(4):e5349.
18. Baker JJ, Johnson BK, Abramovitch RB. Slow growth of Mycobacterium tuberculosis at acidic pH is regulated by phoPR and host-associated carbon sources. Mol Microbiol. octubre de 2014;94(1):56-69.
19. Moopanar K, Nyide ANG, Senzani S, Mvubu NE. Clinical strains of Mycobacterium tuberculosis exhibit differential lipid metabolism-associated transcriptome changes in in vitro cholesterol and infection models. Pathog Dis. 1 de enero de 2023;81:ftac046.
20. Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, et al. High Functional Diversity in Mycobacterium tuberculosis Driven by Genetic Drift and Human Demography. PLOS Biol. 16 de diciembre de 2008;6(12):e311.

21. Firdessa R, Berg S, Hailu E, Schelling E, Gumi B, Erenso G, et al. Mycobacterial Lineages Causing Pulmonary and Extrapulmonary Tuberculosis, Ethiopia - Volume 19, Number 3—March 2013 - Emerging Infectious Diseases journal - CDC. [citado 3 de octubre de 2023]; Disponible en: https://wwwnc.cdc.gov/eid/article/19/3/12-0256_article
22. Blouin Y, Hauck Y, Soler C, Fabre M, Vong R, Dehan C, et al. Significance of the Identification in the Horn of Africa of an Exceptionally Deep Branching Mycobacterium tuberculosis Clade. PLOS ONE. 27 de diciembre de 2012;7(12):e52841.
23. Sanoussi CN, Affolabi D, Rigouts L, Anagonou S, de Jong B. Genotypic characterization directly applied to sputum improves the detection of Mycobacterium africanum West African 1, under-represented in positive cultures. PLoS Negl Trop Dis. 1 de septiembre de 2017;11(9):e0005900.
24. Menardo F, Rutaihwa LK, Zwyer M, Borrell S, Comas I, Conceição EC, et al. Local adaptation in populations of Mycobacterium tuberculosis endemic to the Indian Ocean Rim. F1000Research. 29 de marzo de 2021;10:60.
25. Brites D, Gagneux S. Co-evolution of Mycobacterium tuberculosis and Homo sapiens. Immunol Rev. marzo de 2015;264(1):6-24.
26. Senghore M, Diarra B, Gehre F, Otu J, Worwui A, Muhammad AK, et al. Evolution of Mycobacterium tuberculosis complex lineages and their role in an emerging threat of multidrug resistant tuberculosis in Bamako, Mali. Sci Rep. diciembre de 2020;10(1):327.
27. Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, Willery E, et al. Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage. Nat Genet. marzo de 2015;47(3):242-9.
28. Shuaib YA, Utpatel C, Kohl TA, Barilar I, Diricks M, Ashraf N, et al. Origin and Global Expansion of Mycobacterium tuberculosis Complex Lineage 3. Genes. 31 de mayo de 2022;13(6):990.
29. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature. 11 de junio de 1998;393(6685):537-44.

30. Alonso H, Samper S, Martín C, Otal I. Mapping IS6110 in high-copy number Mycobacterium tuberculosis strains shows specific insertion points in the Beijing genotype. BMC Genomics. 25 de junio de 2013;14(1):422.

31. Coros A, DeConno E, Derbyshire KM. 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. J Bacteriol. mayo de 2008;190(9):3408-10.

32. van Soolingen D, Hermans PW, de Haas PE, van Embden JD. Insertion element IS1081-associated restriction fragment length polymorphisms in Mycobacterium tuberculosis complex species: a reliable tool for recognizing Mycobacterium bovis BCG. J Clin Microbiol. julio de 1992;30(7):1772-7.

33. Nghiem MN, Nguyen BV, Nguyen ST, Vo TTB, Nong HV. A Simple, Single Triplex PCR of IS6110, IS1081, and 23S Ribosomal DNA Targets, Developed for Rapid Detection and Discrimination of Mycobacterium from Clinical Samples. J Microbiol Biotechnol. 2015;25(5):745-52.

34. Ates LS, Sayes F, Frigui W, Ummels R, Damen MPM, Bottai D, et al. RD5-mediated lack of PE_PGRS and PPE-MPTR export in BCG vaccine strains results in strong reduction of antigenic repertoire but little impact on protection. PLOS Pathog. 18 de junio de 2018;14(6):e1007139.

35. Cerezo I, Jiménez Y, Hernandez J, Zozio T, Murcia MI, Rastogi N. A first insight on the population structure of Mycobacterium tuberculosis complex as studied by spoligotyping and MIRU-VNTRs in Bogotá, Colombia. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. junio de 2012;12(4):657-63.

36. Rizvi SMS, Tarafder S, Anwar S, Perdigão J, Johora FT, Sattar H, et al. Circulating strains of Mycobacterium tuberculosis: 24 loci MIRU-VNTR analysis in Bangladesh. Infect Genet Evol. 1 de diciembre de 2020;86:104634.

37. Fieweger RA, Wilburn KM, VanderVen BC. Comparing the Metabolic Capabilities of Bacteria in the Mycobacterium tuberculosis Complex. Microorganisms. junio de 2019;7(6):177.

38. Warner DF. Mycobacterium tuberculosis Metabolism. Cold Spring Harb Perspect Med. abril de 2015;5(4):a021121.

39. Mashabela GT, de Wet TJ, Warner DF. Mycobacterium tuberculosis Metabolism. Microbiol Spectr. 26 de julio de 2019;7(4):10.1128/microbiolspec.gpp3-0067-2019.

40. Av-Gay Y, Everett M. The eukaryotic-like Ser/Thr protein kinases of Mycobacterium tuberculosis. Trends Microbiol. mayo de 2000;8(5):238-44.

41. Malhotra V, Okon BP, Clark-Curtiss JE. Mycobacterium tuberculosis Protein Kinase K Enables Growth Adaptation through Translation Control. J Bacteriol. agosto de 2012;194(16):4184-96.

42. Sukumar S, Krishnan A, Khan MKA. Protein kinases as antituberculosis targets: The case of thymidylate kinases. Front Biosci-Landmark. 1 de marzo de 2020;25(9):1636-54.

43. Prisic S, Husson RN. Mycobacterium tuberculosis Serine/Threonine Protein Kinases. Microbiol Spectr. octubre de 2014;2(5):10.1128/microbiolspec.MGM2-0006-2013.

44. Gago G, Diacovich L, Gramajo H. Lipid metabolism and its implication in mycobacteria-host interaction. Curr Opin Microbiol. febrero de 2018;41:36-42.

45. Quadri LEN. Biosynthesis of mycobacterial lipids by polyketide synthases and beyond. Crit Rev Biochem Mol Biol. 2014;49(3):179-211.

46. D'Souza C, Kishore U, Tsolaki AG. The PE-PPE Family of Mycobacterium tuberculosis: Proteins in Disguise. Immunobiology. 1 de marzo de 2023;228(2):152321.

47. Forrellad MA, Klepp LI, Gioffré A, Sabio Y García J, Morbidoni HR, Santangelo MDLP, et al. Virulence factors of the *Mycobacterium tuberculosis* complex. Virulence. enero de 2013;4(1):3-66.

48. Asensio JG, Maia C, Ferrer NL, Barilone N, Laval F, Soto CY, et al. The Virulence-associated Two-component PhoP-PhoR System Controls the Biosynthesis of Polyketide-derived Lipids in Mycobacterium tuberculosis*. J Biol Chem. 20 de enero de 2006;281(3):1313-6.

49. Beceiro A, Tomás M, Bou G. Antimicrobial Resistance and Virulence: a Successful or Deleterious Association in the Bacterial World? Clin Microbiol Rev. abril de 2013;26(2):185-230.

50. Ly A, Liu J. Mycobacterial Virulence Factors: Surface-Exposed Lipids and Secreted Proteins. Int J Mol Sci. 2 de junio de 2020;21(11):3985.

51. Pérez I, Campos-Pardos E, Díaz C, Uranga S, Sayes F, Vicente F, et al. The Mycobacterium tuberculosis PhoPR virulence system regulates expression of the universal second messenger

c-di-AMP and impacts vaccine safety and efficacy. Mol Ther - Nucleic Acids. 8 de marzo de 2022;27:1235-48.

52. Espinoza V. Prevención, diagnóstico, tratamiento y control de la Tuberculosis [Internet]. Ministerio de Salud Pública; 2018. Disponible en: https://www.salud.gob.ec/wp-content/uploads/2018/03/GP_Tuberculosis-1.pdf
53. World Health Organization. WHO consolidated guidelines on tuberculosis: module 1: prevention: tuberculosis preventive treatment [Internet]. Geneva, Switzerland: World Health Organization; 2020 [citado 29 de junio de 2023] p. 55 pp. Disponible en: https://www.who.int/publications-detail-redirect/9789240001503
54. Timmins GS, Deretic V. Mechanisms of action of isoniazid. Mol Microbiol. 2006;62(5):1220-7.
55. Khan SR, Manialawy Y, Siraki AG. Isoniazid and host immune system interactions: A proposal for a novel comprehensive mode of action. Br J Pharmacol. diciembre de 2019;176(24):4599-608.
56. Campbell EA, Korzheva N, Mustaev A, Murakami K, Nair S, Goldfarb A, et al. Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. Cell. 23 de marzo de 2001;104(6):901-12.
57. Nusrath Unissa A, Hanna LE. Molecular mechanisms of action, resistance, detection to the first-line anti tuberculosis drugs: Rifampicin and pyrazinamide in the post whole genome sequencing era. Tuberculosis. 1 de julio de 2017;105:96-107.
58. Zhang Y, Shi W, Zhang W, Mitchison D. Mechanisms of Pyrazinamide Action and Resistance. Microbiol Spectr. 18 de julio de 2014;2(4):10.1128/microbiolspec.mgm2-0023-2013.
59. Schubert K, Sieger B, Meyer F, Giacomelli G, Böhm K, Rieblinger A, et al. The Antituberculosis Drug Ethambutol Selectively Blocks Apical Growth in CMN Group Bacteria. mBio. 7 de febrero de 2017;8(1):e02213-16.
60. Cavalcante SC, Durovni B, Barnes GL, Souza FBA, Silva RF, Barroso PF, et al. Community-randomized trial of enhanced DOTS for tuberculosis control in Rio de Janeiro, Brazil. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. febrero de 2010;14(2):203-9.
61. Mirzayev F, Viney K, Linh NN, Gonzalez-Angulo L, Gegia M, Jaramillo E, et al. World Health Organization recommendations on the treatment of drug-resistant tuberculosis, 2020 update. Eur

Respir J. 3 de junio de 2021;57(6):2003300.

62. Reygaert WC. An overview of the antimicrobial resistance mechanisms of bacteria. AIMS Microbiol. 26 de junio de 2018;4(3):482-501.
63. Abushaheen MA, Muzaheed, Fatani AJ, Alosaimi M, Mansy W, George M, et al. Antimicrobial resistance, mechanisms and its clinical significance. Dis Mon. 1 de junio de 2020;66(6):100971.
64. Poulton NC, Rock JM. Unraveling the mechanisms of intrinsic drug resistance in Mycobacterium tuberculosis. Front Cell Infect Microbiol. 17 de octubre de 2022;12:997283.
65. Nasiri MJ, Haeili M, Ghazi M, Goudarzi H, Pormohammad A, Imani Fooladi AA, et al. New Insights in to the Intrinsic and Acquired Drug Resistance Mechanisms in Mycobacteria. Front Microbiol [Internet]. 2017 [citado 30 de junio de 2023];8. Disponible en: https://www.frontiersin.org/articles/10.3389/fmicb.2017.00681
66. Kempker RR, Kipiani M, Mirtskhulava V, Tukvadze N, Magee MJ, Blumberg HM. Acquired Drug Resistance in Mycobacterium tuberculosis and Poor Outcomes among Patients with Multidrug-Resistant Tuberculosis. Emerg Infect Dis. junio de 2015;21(6):992-1001.
67. Andries K, Villellas C, Coeck N, Thys K, Gevers T, Vranckx L, et al. Acquired Resistance of Mycobacterium tuberculosis to Bedaquiline. PLOS ONE. 10 de julio de 2014;9(7):e102135.
68. Brammachary U, Ramachandra V, Palavesam S, Kapalamurthy VRC, Muralidhar A, Muthaiah M, et al. Mechanisms and Action of Drug Resistance on Mycobacterium tuberculosis [Internet]. IntechOpen; 2022 [citado 30 de junio de 2023]. Disponible en: https://www.intechopen.com/online-first/85046
69. Gygli SM, Loiseau C, Jugheli L, Adamia N, Trauner A, Reinhard M, et al. Prisons as ecological drivers of fitness-compensated multidrug-resistant Mycobacterium tuberculosis. Nat Med. julio de 2021;27(7):1171-7.
70. Viney K, Linh NN, Gegia M, Zignol M, Glaziou P, Ismail N, et al. New definitions of pre-extensively and extensively drug-resistant tuberculosis: update from the World Health Organization. Eur Respir J. abril de 2021;57(4):2100361.
71. Marney MW, Metzger RP, Hecht D, Valafar F. Modeling the structural origins of drug resistance to isoniazid via key mutations in Mycobacterium tuberculosis catalase-peroxidase, KatG. Tuberc

Edinb Scotl. enero de 2018;108:155-62.

72. Pinhata JMW, Brandao AP, Mendes F de F, Rabello MC da S, Ferrazoli L, de Oliveira RS. Correlating genetic mutations with isoniazid phenotypic levels of resistance in Mycobacterium tuberculosis isolates from patients with drug-resistant tuberculosis in a high burden setting. Eur J Clin Microbiol Infect Dis. 1 de diciembre de 2021;40(12):2551-61.

73. Bakhtiyariniya P, Khosravi AD, Hashemzadeh M, Savari M. Detection and characterization of mutations in genes related to isoniazid resistance in Mycobacterium tuberculosis clinical isolates from Iran. Mol Biol Rep [Internet]. 2 de abril de 2022 [citado 18 de junio de 2022]; Disponible en: https://link.springer.com/10.1007/s11033-022-07404-2

74. Penn-Nicholson A, Georghiou SB, Ciobanu N, Kazi M, Bhalla M, David A, et al. Detection of isoniazid, fluoroquinolone, ethionamide, amikacin, kanamycin, and capreomycin resistance by the Xpert MTB/XDR assay: a cross-sectional multicentre diagnostic accuracy study. Lancet Infect Dis. 1 de febrero de 2022;22(2):242-9.

75. Mvelase NR, Pillay M, Sibanda W, Ngozo JN, Brust JCM, Mlisana KP. rpoB Mutations Causing Discordant Rifampicin Susceptibility in Mycobacterium tuberculosis: Retrospective Analysis of Prevalence, Phenotypic, Genotypic, and Treatment Outcomes. Open Forum Infect Dis. 1 de abril de 2019;6(4):ofz065.

76. Mboowa G, Namaganda C, Ssengooba W. Rifampicin resistance mutations in the 81 bp RRDR of rpoB gene in Mycobacterium tuberculosis clinical isolates using Xpert®MTB/RIF in Kampala, Uganda: a retrospective study. BMC Infect Dis. 4 de septiembre de 2014;14(1):481.

77. Portelli S, Myung Y, Furnham N, Vedithi SC, Pires DEV, Ascher DB. Prediction of rifampicin resistance beyond the RRDR using structure-based machine learning approaches. Sci Rep. 22 de octubre de 2020;10(1):18120.

78. Koch A, Mizrahi V, Warner DF. The impact of drug resistance on Mycobacterium tuberculosis physiology: what can we learn from rifampicin? Emerg Microbes Infect. 1 de enero de 2014;3(1):1-11.

79. Ma P, Luo T, Ge L, Chen Z, Wang X, Zhao R, et al. Compensatory effects of M. tuberculosis rpoB mutations outside the rifampicin resistance-determining region. Emerg Microbes Infect. 1 de

enero de 2021;10(1):743-52.

80. Lahiri N, Shah RR, Layre E, Young D, Ford C, Murray MB, et al. Rifampin Resistance Mutations Are Associated with Broad Chemical Remodeling of Mycobacterium tuberculosis. J Biol Chem. 1 de julio de 2016;291(27):14248-56.
81. Conkle-Gutierrez D, Ramirez-Busby S, Gorman B, Elghraoui A, Hoffner S, Elmaraachli W, et al. Novel and Reported Compensatory Mutations in *rpoABC* Associate Specifically with Predominant *Mycobacterium tuberculosis* Rifampicin Resistance Marker *rpoB* :S450L [Internet]. Microbiology; 2022 feb [citado 5 de noviembre de 2022]. Disponible en: http://biorxiv.org/lookup/doi/10.1101/2022.02.22.481565
82. Vargas AP, Rios AA, Grandjean L, Kirwan DE, Gilman RH, Sheen P, et al. Determination of Potentially Novel Compensatory Mutations in rpoC Associated with Rifampin Resistance and rpoB Mutations in Mycobacterium tuberculosis Clinical Isolates from Peru. Int J Mycobacteriology. 2020;9(2):26.
83. Wang S, Zhou Y, Zhao B, Ou X, Xia H, Zheng Y, et al. Characteristics of compensatory mutations in the rpoC gene and their association with compensated transmission of Mycobacterium tuberculosis. Front Med. 1 de febrero de 2020;14(1):51-9.
84. Yun YJ, Lee JS, Yoo JC, Cho E, Park D, Kook YH, et al. Patterns of rpoC Mutations in Drug-Resistant Mycobacterium tuberculosis Isolated from Patients in South Korea. Tuberc Respir Dis. julio de 2018;81(3):222-7.
85. Xu Z, Zhou A, Wu J, Zhou A, Li J, Zhang S, et al. Transcriptional Approach for Decoding the Mechanism of rpoC Compensatory Mutations for the Fitness Cost in Rifampicin-Resistant Mycobacterium tuberculosis. Front Microbiol [Internet]. 2018 [citado 30 de junio de 2023];9. Disponible en: https://www.frontiersin.org/articles/10.3389/fmicb.2018.02895
86. Allana S, Shashkina E, Mathema B, Bablishvili N, Tukvadze N, Shah NS, et al. pncA Gene Mutations Associated with Pyrazinamide Resistance in Drug-Resistant Tuberculosis, South Africa and Georgia. Emerg Infect Dis. marzo de 2017;23(3):491-5.
87. Barco P, Cardoso RF, Hirata RDC, Leite CQF, Pandolfi JR, Sato DN, et al. pncA mutations in pyrazinamide-resistant Mycobacterium tuberculosis clinical isolates from the southeast region of

Brazil. J Antimicrob Chemother. 1 de noviembre de 2006;58(5):930-5.

88. Cuevas-Córdoba B, Xochihua-González SO, Cuellar A, Fuentes-Domínguez J, Zenteno-Cuevas R. Characterization of pncA gene mutations in pyrazinamide-resistant Mycobacterium tuberculosis isolates from Mexico. Infect Genet Evol. 1 de octubre de 2013;19:330-4.
89. Li K, Yang Z, Gu J, Luo M, Deng J, Chen Y. Characterization of pncA Mutations and Prediction of PZA Resistance in Mycobacterium tuberculosis Clinical Isolates From Chongqing, China. Front Microbiol [Internet]. 2021 [citado 3 de noviembre de 2022];11. Disponible en: https://www.frontiersin.org/articles/10.3389/fmicb.2020.594171
90. Shi D, Zhou Q, Xu S, Zhu Y, Li H, Xu Y. Pyrazinamide Resistance and pncA Mutation Profiles in Multidrug Resistant Mycobacterium Tuberculosis. Infect Drug Resist. 30 de agosto de 2022;15:4985-94.
91. Shrestha D, Maharjan B, Thapa J, Akapelwa ML, Bwalya P, Chizimu JY, et al. Detection of Mutations in pncA in Mycobacterium tuberculosis Clinical Isolates from Nepal in Association with Pyrazinamide Resistance. Curr Issues Mol Biol. septiembre de 2022;44(9):4132-41.
92. Pandey B, Grover S, Tyagi C, Goyal S, Jamal S, Singh A, et al. Molecular principles behind pyrazinamide resistance due to mutations in panD gene in Mycobacterium tuberculosis. Gene. 25 de abril de 2016;581(1):31-42.
93. Zhang S, Chen J, Shi W, Liu W, Zhang W, Zhang Y. Mutations in panD encoding aspartate decarboxylase are associated with pyrazinamide resistance in Mycobacterium tuberculosis. Emerg Microbes Infect. 1 de enero de 2013;2(1):1-5.
94. Ramirez-Busby SM, Rodwell TC, Fink L, Catanzaro D, Jackson RL, Pettigrove M, et al. A Multinational Analysis of Mutations and Heterogeneity in PZase, RpsA, and PanD Associated with Pyrazinamide Resistance in M/XDR Mycobacterium tuberculosis. Sci Rep. 19 de junio de 2017;7(1):3790.
95. Tan Y, Hu Z, Zhang T, Cai X, Kuang H, Liu Y, et al. Role of pncA and rpsA Gene Sequencing in Detection of Pyrazinamide Resistance in Mycobacterium tuberculosis Isolates from Southern China. J Clin Microbiol. 21 de diciembre de 2020;52(1):291-7.
96. Gu Y, Yu X, Jiang G, Wang X, Ma Y, Li Y, et al. Pyrazinamide resistance among

multidrug-resistant tuberculosis clinical isolates in a national referral center of China and its correlations with pncA, rpsA, and panD gene mutations. Diagn Microbiol Infect Dis. 1 de marzo de 2016;84(3):207-11.

97. Werngren J, Mansjö M, Glader M, Hoffner S, Davies Forsman L. Detection of Pyrazinamide Heteroresistance in Mycobacterium tuberculosis. Antimicrob Agents Chemother. 17 de agosto de 2021;65(9):10.1128/aac.00720-21.

98. Whitfield MG, Engelthaler DM, Allender C, Folkerts M, Heupink TH, Limberis J, et al. Comparative Performance of Genomic Methods for the Detection of Pyrazinamide Resistance and Heteroresistance in Mycobacterium tuberculosis. J Clin Microbiol. 2022;60(1):e01907-21.

99. Li M chao, Chen R, Lin S qiang, Lu Y, Liu H can, Li G lian, et al. Detecting Ethambutol Resistance in Mycobacterium tuberculosis Isolates in China: A Comparison Between Phenotypic Drug Susceptibility Testing Methods and DNA Sequencing of embAB. Front Microbiol [Internet]. 2020 [citado 3 de noviembre de 2022];11. Disponible en: https://www.frontiersin.org/articles/10.3389/fmicb.2020.00781

100. Sun Q, Xiao T yang, Liu H can, Zhao X qin, Liu Z guang, Li Y nan, et al. Mutations within embCAB Are Associated with Variable Level of Ethambutol Resistance in Mycobacterium tuberculosis Isolates from China. Antimicrob Agents Chemother. 21 de diciembre de 2017;62(1):e01279-17.

101. Xu Y, Jia H, Huang H, Sun Z, Zhang Z. Mutations Found in embCAB, embR, and ubiA Genes of Ethambutol-Sensitive and -Resistant Mycobacterium tuberculosis Clinical Isolates from China. BioMed Res Int. 2015;2015:951706.

102. Wang J, Zhao W, Liu R, Huo F, Dong L, Xue Y, et al. Rapid Detection of Ethambutol-Resistant Mycobacterium tuberculosis from Sputum by High-Resolution Melting Analysis in Beijing, China . Infect Drug Resist. 20 de octubre de 2020;13:3707-13.

103. Madrazo-Moya CF, Cancino-Muñoz I, Cuevas-Córdoba B, González-Covarrubias V, Barbosa-Amezcua M, Soberón X, et al. Whole genomic sequencing as a tool for diagnosis of drug and multidrug-resistance tuberculosis in an endemic region in Mexico. PLOS ONE. 5 de junio de 2019;14(6):e0213046.

104. Chauhan DS, Sharma R, Parashar D, Das R, Sharma P, Singh AV, et al. Rapid detection of ethambutol-resistant Mycobacterium tuberculosis in clinical specimens by real-time polymerase chain reaction hybridisation probe method. Indian J Med Microbiol. 2018;36(2):211-6.

105. Miotto P, Zhang Y, Cirillo DM, Yam WC. Drug resistance mechanisms and drug susceptibility testing for tuberculosis. Respirology. 2018;23(12):1098-113.

106. Meier A, Sander P, Schaper KJ, Scholz M, Böttger EC. Correlation of molecular resistance mechanisms and phenotypic resistance levels in streptomycin-resistant Mycobacterium tuberculosis. Antimicrob Agents Chemother. noviembre de 1996;40(11):2452-4.

107. Murray JF, Schraufnagel DE, Hopewell PC. Treatment of Tuberculosis. A Historical Perspective. Ann Am Thorac Soc. diciembre de 2015;12(12):1749-59.

108. Germovsek E, Barker CI, Sharland M. What do I need to know about aminoglycoside antibiotics? Arch Dis Child - Educ Pract. 1 de abril de 2017;102(2):89-93.

109. Sun H, Zeng J, Li S, Liang P, Zheng C, Liu Y, et al. Interaction between rpsL and gyrA mutations affects the fitness and dual resistance of Mycobacterium tuberculosis clinical isolates against streptomycin and fluoroquinolones. Infect Drug Resist. 27 de marzo de 2018;11:431-40.

110. Villellas C, Aristimuño L, Vitoria MA, Prat C, Blanco S, García de Viedma D, et al. Analysis of Mutations in Streptomycin-Resistant Strains Reveals a Simple and Reliable Genetic Marker for Identification of the Mycobacterium tuberculosis Beijing Genotype. J Clin Microbiol. julio de 2013;51(7):2124-30.

111. Bwalya P, Yamaguchi T, Solo ES, Chizimu JY, Mbulo G, Nakajima C, et al. Characterization of Mutations Associated with Streptomycin Resistance in Multidrug-Resistant Mycobacterium tuberculosis in Zambia. Antibiotics. octubre de 2021;10(10):1169.

112. Shafipour M, Shirzad-Aski H, Mohammadzadeh A, Ghazvini K, Zamani S, Koohi PM, et al. Evaluation of Mutations Related to Streptomycin Resistance in Mycobacterium tuberculosis Clinical Isolates. Curr Microbiol. 8 de octubre de 2022;79(11):343.

113. Wong SY, Lee JS, Kwak HK, Via LE, Boshoff HIM, Barry CE. Mutations in gidB Confer Low-Level Streptomycin Resistance in Mycobacterium tuberculosis▽. Antimicrob Agents

Chemother. junio de 2011;55(6):2515-22.

114. Pandey B, Grover S, Goyal S, Jamal S, Singh A, Kaur J, et al. Novel missense mutations in gidB gene associated with streptomycin resistance in Mycobacterium tuberculosis: insights from molecular dynamics. J Biomol Struct Dyn. 2 de enero de 2019;37(1):20-35.

115. Rodríguez-García Á, Mares-Alejandre RE, Muñoz-Muñoz PLA, Ruvalcaba-Ruiz S, González-Sánchez RA, Bernáldez-Sarabia J, et al. Molecular Analysis of Streptomycin Resistance Genes in Clinical Strains of Mycobacterium tuberculosis and Biocomputational Analysis of the MtGidB L101F Variant. Antibiotics. julio de 2021;10(7):807.

116. Wang Y, Li Q, Gao H, Zhang Z, Liu Y, Lu J, et al. The roles of rpsL, rrs, and gidB mutations in predicting streptomycin-resistant drugs used on clinical Mycobacterium tuberculosis isolates from Hebei Province, China. Int J Clin Exp Pathol. 1 de julio de 2019;12(7):2713-21.

117. Sreevatsan S, Pan X, Stockbauer KE, Williams DL, Kreiswirth BN, Musser JM. Characterization of rpsL and rrs mutations in streptomycin-resistant Mycobacterium tuberculosis isolates from diverse geographic localities. Antimicrob Agents Chemother. abril de 1996;40(4):1024-6.

118. Rocha DMGC, Magalhães C, Cá B, Ramos A, Carvalho T, Comas I, et al. Heterogeneous Streptomycin Resistance Level Among Mycobacterium tuberculosis Strains From the Same Transmission Cluster. Front Microbiol [Internet]. 2021 [citado 18 de junio de 2022];12. Disponible en: https://www.frontiersin.org/article/10.3389/fmicb.2021.659545

119. Rocha DMGC, Viveiros M, Saraiva M, Osório NS. The Neglected Contribution of Streptomycin to the Tuberculosis Drug Resistance Problem. Genes. diciembre de 2021;12(12):2003.

120. Infección ósea por micobacterium fortuitum (no tuberculosa) | Revista Ecuatoriana de Ortopedia y Traumatología. 4 de noviembre de 2021 [citado 28 de junio de 2023]; Disponible en: http://revistacientificaseot.com/index.php/revseot/article/view/163

121. Chafla SJC, Macias LGC, Trujillo KAA, Sailema RIM, Olovacha IAM. Micobacteriosis pulmonar por Mycobacterium kansasii. Cienc Al Serv Salud Nutr. 28 de marzo de 2018;8(2):13-22.

122. Nasre-Nasser R, Muñoz-López G, Nasre-Nasser R, Muñoz-López G. Infección por Mycobacterium scrofulaceum en un niño de las Islas Galápagos - Ecuador. Rev Chil Infectol.

2018;35(6):705-9.

123. Moreno-Izquierdo C, Zurita J, Contreras-Yametti FI, Jara-Palacios MA. Mycobacterium abscessus subspecies abscessus infection associated with cosmetic surgical procedures: Case series. IDCases. 1 de enero de 2020;22:e00992.

124. Gill CM, Dolan L, Piggott LM, McLaughlin AM. New developments in tuberculosis diagnosis and treatment. Breathe. marzo de 2022;18(1):210149.

125. Siddiqi S, Ahmed A, Asif S, Behera D, Javaid M, Jani J, et al. Direct Drug Susceptibility Testing of Mycobacterium tuberculosis for Rapid Detection of Multidrug Resistance Using the Bactec MGIT 960 System: a Multicenter Study. J Clin Microbiol. 21 de diciembre de 2020;50(2):435-40.

126. Kim SJ. Drug-susceptibility testing in tuberculosis: methods and reliability of results. Eur Respir J. 1 de marzo de 2005;25(3):564-9.

127. Gomathi NS, Kumar V. Reliability of Mycobacterial Growth Indicator Tube (MGIT) 960 for the detection of isoniazid resistance in a tuberculosis endemic setting. Indian J Med Res. marzo de 2014;139(3):471-3.

128. Mokaddas E, Ahmad S, Eldeen HS, Al-Mutairi N. Discordance between Xpert MTB/RIF Assay and Bactec MGIT 960 Culture System for Detection of Rifampin-Resistant Mycobacterium tuberculosis Isolates in a Country with a Low Tuberculosis (TB) Incidence. J Clin Microbiol. abril de 2015;53(4):1351-4.

129. Thosani P, Agarwal A, Karade S, Sen S, Katoch CS, Jindamwar P, et al. Assessment of resistance to second line anti-tubercular drugs by line probe assay at a tertiary care hospital. 19 de junio de 2022 [citado 18 de junio de 2022]; Disponible en: https://www.mjdrdypv.org/preprintarticle.asp?id=336821;type=0

130. Iem V, Chittamany P, Suthepmany S, Siphanthong S, Somphavong S, Kontogianni K, et al. Pooling sputum for Xpert MTB/RIF and Xpert Ultra testing during the Covid-19 pandemic in Lao People's Democratic Republic. PLOS Glob Public Health. 4 de abril de 2022;2(4):e0000116.

131. Rahman SMM, Ather MdF, Nasrin R, Hoque MA, Khatun R, Rahman T, et al. Performance of WHO-Endorsed Rapid Tests for Detection of Susceptibility to First-Line Drugs in Patients with Pulmonary Tuberculosis in Bangladesh. Diagnostics. 5 de febrero de 2022;12(2):410.

132. Simieneh A, Tadesse M, Kebede W, Gashaw M, Abebe G. Combination of Xpert® MTB/RIF and DetermineTM TB-LAM Ag improves the diagnosis of extrapulmonary tuberculosis at Jimma University Medical Center, Oromia, Ethiopia. PLOS ONE. 3 de febrero de 2022;17(2):e0263172.

133. Tantivitayakul P, Ruangchai W, Juthayothin T, Smittipat N, Disratthakit A, Mahasirimongkol S, et al. Homoplastic single nucleotide polymorphisms contributed to phenotypic diversity in Mycobacterium tuberculosis. Sci Rep. diciembre de 2020;10(1):8024.

134. Yang T, Gan M, Liu Q, Liang W, Tang Q, Luo G, et al. SAM-TB: a whole genome sequencing data analysis website for detection of Mycobacterium tuberculosis drug resistance and transmission. Brief Bioinform. 24 de febrero de 2022;23(2):bbac030.

135. Lam C, Martinez E, Crighton T, Furlong C, Donnan E, Marais BJ, et al. Value of routine whole genome sequencing for Mycobacterium tuberculosis drug resistance detection. Int J Infect Dis. diciembre de 2021;113:S48-54.

136. Wang L, Yang J, Chen L, Wang W, Yu F, Xiong H. Whole-genome sequencing of *Mycobacterium tuberculosis* for prediction of drug resistance. Epidemiol Infect. 2022;150:e22.

137. Brown AC. Whole-Genome Sequencing of Mycobacterium tuberculosis Directly from Sputum Samples. En: Parish T, Kumar A, editores. Mycobacteria Protocols [Internet]. New York, NY: Springer US; 2021 [citado 18 de junio de 2022]. p. 459-80. (Methods in Molecular Biology). Disponible en: https://doi.org/10.1007/978-1-0716-1460-0_20

138. Torres Ortiz A, Coronel J, Vidal JR, Bonilla C, Moore DAJ, Gilman RH, et al. Genomic signatures of pre-resistance in Mycobacterium tuberculosis. Nat Commun. diciembre de 2021;12(1):7312.

139. Freschi L, Vargas R, Husain A, Kamal SMM, Skrahina A, Tahseen S, et al. Population structure, biogeography and transmissibility of Mycobacterium tuberculosis. Nat Commun. diciembre de 2021;12(1):6099.

140. Bogaerts B, Delcourt T, Soetaert K, Boarbi S, Ceyssens PJ, Winand R, et al. A Bioinformatics Whole-Genome Sequencing Workflow for Clinical Mycobacterium tuberculosis Complex Isolate Analysis, Validated Using a Reference Collection Extensively Characterized with Conventional Methods and In Silico Approaches. J Clin Microbiol. 31 de marzo de 2021;59(6):e00202-21.

141. Shea J, Halse TA, Lapierre P, Shudt M, Kohlerschmidt D, Van Roey P, et al. Comprehensive Whole-Genome Sequencing and Reporting of Drug Resistance Profiles on Clinical Cases of Mycobacterium tuberculosis in New York State. J Clin Microbiol. junio de 2017;55(6):1871-82.

142. Coll F, McNerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. Genome Med. diciembre de 2015;7(1):51.

143. Glynn JR, Khan P, Mzembe T, Sichali L, Fine PEM, Crampin AC, et al. Contribution of remote M.tuberculosis infection to tuberculosis disease: A 30-year population study. PLOS ONE. 27 de enero de 2023;18(1):e0278136.

144. Yassine E, Galiwango R, Ssengooba W, Ashaba F, Joloba ML, Zalwango S, et al. Assessing a transmission network of Mycobacterium tuberculosis in an African city using single nucleotide polymorphism threshold analysis. MicrobiologyOpen. 16 de junio de 2021;10(3):e1211.

145. Xu G, Mao X, Wang J, Pan H. Clustering and recent transmission of Mycobacterium tuberculosis in a Chinese population. Infect Drug Resist. 6 de marzo de 2018;11:323-30.

146. Alaridah N, Hallbäck ET, Tångrot J, Winqvist N, Sturegård E, Florén-Johansson K, et al. Transmission dynamics study of tuberculosis isolates with whole genome sequencing in southern Sweden. Sci Rep. 20 de marzo de 2019;9(1):4931.

147. Chang Campos C. Manual de Normas y Procedimientos para el control de la Tuberculosis en Ecuador. Minist Salud Pública. 2010;2da edición:336 pp.

148. Mora-Pinargote C, Garzon-Chavez D, Franco-Sotomayor G, Leon-Benitez M, Granda-Pardo JC, Trueba G, et al. Country-wide rapid screening for the Mycobacterium tuberculosis Beijing sublineage in Ecuador using a single-nucleotide polymorphism-polymerase chain reaction method. Int J Mycobacteriology. 2019;8(4):366-70.

149. Garzon-Chavez D, Zurita J, Mora-Pinargote C, Franco-Sotomayor G, Leon-Benitez M, Granda-Pardo JC, et al. Prevalence, Drug Resistance, and Genotypic Diversity of the Mycobacterium tuberculosis Beijing Family in Ecuador. Microb Drug Resist. julio de 2019;25(6):931-7.

150. Zurita J, Espinel N, Barba P, Ortega-Paredes D, Zurita-Salinas C, Rojas Y, et al. Genetic diversity

and drug resistance of Mycobacterium tuberculosis in Ecuador. Int J Tuberc Lung Dis. 1 de febrero de 2019;23(2):166-73.

151. Garcés E, Cifuentes L, Franco G, Romero-Sandoval N, Arias PJ. Study of the distribution of lineages of Mycobacterium tuberculosis in a prison in Guayaquil, Ecuador. 2023;

152. Garzon-Chavez D, Garcia-Bereguiain MA, Mora-Pinargote C, Granda-Pardo JC, Leon-Benitez M, Franco-Sotomayor G, et al. Population structure and genetic diversity of Mycobacterium tuberculosis in Ecuador. Sci Rep. 10 de abril de 2020;10:6237.

153. Díaz Acosta CC, Russomando G, Candia N, Ritacco V, Vasconcellos SEG, de Berrêdo Pinho Moreira M, et al. Exploring the "Latin American Mediterranean" family and the RDRio lineage in Mycobacterium tuberculosis isolates from Paraguay, Argentina and Venezuela. BMC Microbiol. diciembre de 2019;19(1):131.

154. Tatara MB, Perdigão J, Viveiros M, Kritski A, Silva KE da, Sacchi FPC, et al. Genetic Diversity and Molecular Epidemiology of Mycobacterium tuberculosis in Roraima State, Brazil. Am J Trop Med Hyg. 6 de agosto de 2019;101(4):774-9.

155. Hadi SA, Kolte IV, Brenner EP, Cunha E a. T, Simonsen V, Ferrazoli L, et al. Identification of a predominant genotype of Mycobacterium tuberculosis in Brazilian indigenous population. Sci Rep. 13 de enero de 2021;11(1):1224.

156. Salvato RS, Reis AJ, Schiefelbein SH, Gómez MAA, Salvato SS, da Silva LV, et al. Genomic-based surveillance reveals high ongoing transmission of multi-drug-resistant Mycobacterium tuberculosis in Southern Brazil. Int J Antimicrob Agents. 1 de octubre de 2021;58(4):106401.

157. Sánchez-Corrales L, Tovar-Aguirre OL, Galeano-Vanegas NF, Jiménez PAC, Martínez-Vega RA, Maldonado-Londoño CE, et al. Phylogenomic analysis and Mycobacterium tuberculosis antibiotic resistance prediction by whole-genome sequencing from clinical isolates of Caldas, Colombia. PLOS ONE. 7 de octubre de 2021;16(10):e0258402.

158. Ferro BE, Nieto LM, Rozo JC, Forero L, van Soolingen D. Multidrug-Resistant Mycobacterium tuberculosis, Southwestern Colombia. Emerg Infect Dis. julio de 2011;17(7):1259-62.

159. Munro-Rojas D, Fernandez-Morales E, Zarrabal-Meza J, Martínez-Cazares MT, Parissi-Crivelli

A, Fuentes-Domínguez J, et al. Genetic diversity of drug and multidrug-resistant Mycobacterium tuberculosis circulating in Veracruz, Mexico. PLOS ONE. 15 de marzo de 2018;13(3):e0193626.

160. Lefort B, Pérez-Martínez D, Viveros D, Zenteno-Cuevas R, Lefort B, Pérez-Martínez D, et al. Diversidad genética de Mycobacterium tuberculosis circulando en la jurisdicción V de Jalapa, Veracruz, México. Rev Chil Infectol. octubre de 2021;38(5):639-46.

161. Santos-Lazaro D, Gavilan RG, Solari L, Vigo AN, Puyen ZM. Whole genome analysis of extensively drug resistant Mycobacterium tuberculosis strains in Peru. Sci Rep. diciembre de 2021;11(1):9493.

162. Sheen P, Couvin D, Grandjean L, Zimic M, Dominguez M, Luna G, et al. Genetic Diversity of Mycobacterium tuberculosis in Peru and Exploration of Phylogenetic Associations with Drug Resistance. PLOS ONE. 24 de junio de 2013;8(6):e65873.

163. Taype CA, Agapito JC, Accinelli RA, Espinoza JR, Godreuil S, Goodman SJ, et al. Genetic diversity, population structure and drug resistance of Mycobacterium tuberculosis in Peru. Infect Genet Evol. 1 de abril de 2012;12(3):577-85.

164. Naor-Hoffmann S, Svetlitsky D, Sal-Man N, Orenstein Y, Ziv-Ukelson M. Predicting the pathogenicity of bacterial genomes using widely spread protein families. BMC Bioinformatics. 24 de junio de 2022;23(1):253.

165. Chiner-Oms Á, López MG, Moreno-Molina M, Furió V, Comas I. Gene evolutionary trajectories in *Mycobacterium tuberculosis* reveal temporal signs of selection. Proc Natl Acad Sci. 26 de abril de 2022;119(17):e2113600119.

166. Moreno-Molina M, Shubladze N, Khurtsilava I, Avaliani Z, Bablishvili N, Torres-Puente M, et al. Genomic analyses of Mycobacterium tuberculosis from human lung resections reveal a high frequency of polyclonal infections. Nat Commun. 11 de mayo de 2021;12(1):2716.

167. Yang T, Zhong J, Zhang J, Li C, Yu X, Xiao J, et al. Pan-Genomic Study of Mycobacterium tuberculosis Reflecting the Primary/Secondary Genes, Generality/Individuality, and the Interconversion Through Copy Number Variations. Front Microbiol [Internet]. 2018 [citado 20 de junio de 2022];9. Disponible en: https://www.frontiersin.org/article/10.3389/fmicb.2018.01886

168. Zakham F, Sironen T, Vapalahti O, Kant R. Pan and Core Genome Analysis of 183

Mycobacterium tuberculosis Strains Revealed a High Inter-Species Diversity among the Human Adapted Strains. Antibiotics. 28 de abril de 2021;10(5):500.

169. Dar HA, Zaheer T, Ullah N, Bakhtiar SM, Zhang T, Yasir M, et al. Pangenome Analysis of Mycobacterium tuberculosis Reveals Core-Drug Targets and Screening of Promising Lead Compounds for Drug Discovery. Antibiotics. 17 de noviembre de 2020;9(11):819.

170. Periwal V, Patowary A, Vellarikkal SK, Gupta A, Singh M, Mittal A, et al. Comparative Whole-Genome Analysis of Clinical Isolates Reveals Characteristic Architecture of Mycobacterium tuberculosis Pangenome. Hoshino Y, editor. PLOS ONE. 8 de abril de 2015;10(4):e0122979.

171. Rouli L, Merhej V, Fournier PE, Raoult D. The bacterial pangenome as a new tool for analysing pathogenic bacteria. New Microbes New Infect. 26 de junio de 2015;7:72-85.

172. Hurtado-Páez U, Álvarez Zuluaga N, Arango Isaza RE, Contreras-Moreira B, Rouzaud F, Robledo J. Pan-genome association study of Mycobacterium tuberculosis lineage-4 revealed specific genes related to the high and low prevalence of the disease in patients from the North-Eastern area of Medellín, Colombia. Front Microbiol [Internet]. 2023 [citado 7 de julio de 2023];13. Disponible en: https://www.frontiersin.org/articles/10.3389/fmicb.2022.1076797

173. Tonkin-Hill G, Corander J, Parkhill J. Challenges in prokaryote pangenomics. Microb Genomics. 2023;9(5):001021.

174. Negrete-Paz AM, Vázquez-Marrufo G, Gutiérrez-Moraga A, Vázquez-Garcidueñas MS. Pangenome Reconstruction of Mycobacterium tuberculosis as a Guide to Reveal Genomic Features Associated with Strain Clinical Phenotype. Microorganisms. junio de 2023;11(6):1495.

175. Sapriel G, Brosch R. Shared Pathogenomic Patterns Characterize a New Phylotype, Revealing Transition toward Host-Adaptation Long before Speciation of Mycobacterium tuberculosis. Genome Biol Evol. 1 de agosto de 2019;11(8):2420-38.

176. Kweon O, Kim SJ, Blom J, Kim SK, Kim BS, Baek DH, et al. Comparative functional pan-genome analyses to build connections between genomic dynamics and phenotypic evolution in polycyclic aromatic hydrocarbon metabolism in the genus Mycobacterium. BMC Evol Biol. 14 de febrero de 2015;15(1):21.

177. Muzondiwa D, Hlanze H, Reva ON. The Epistatic Landscape of Antibiotic Resistance of Different Clades of Mycobacterium tuberculosis. Antibiotics. 15 de julio de 2021;10(7):857.

178. Wang X, Chen S, Xu Y, Zheng H, Xiao T, Li Y, et al. Identification and evaluation of the novel immunodominant antigen Rv2351c from Mycobacterium tuberculosis. Emerg Microbes Infect. junio de 2017;6(6):e48.

179. Mustafa AS. Immunological Characterization of Proteins Expressed by Genes Located in Mycobacterium tuberculosis-Specific Genomic Regions Encoding the ESAT6-like Proteins. Vaccines. 7 de enero de 2021;9(1):27.

180. Gopal P, Tasneen R, Yee M, Lanoix JP, Sarathy J, Rasic G, et al. In Vivo-Selected Pyrazinoic Acid-Resistant Mycobacterium tuberculosis Strains Harbor Missense Mutations in the Aspartate Decarboxylase PanD and the Unfoldase ClpC1. ACS Infect Dis. 14 de julio de 2017;3(7):492-501.

181. Jackson M. The Mycobacterial Cell Envelope—Lipids. Cold Spring Harb Perspect Med. octubre de 2014;4(10):a021105.

182. Sanchini A, Jandrasits C, Tembrockhaus J, Kohl TA, Utpatel C, Maurer FP, et al. Improving tuberculosis surveillance by detecting international transmission using publicly available whole genome sequencing data. Eurosurveillance. 14 de enero de 2021;26(2):1900677.

183. Guo Y, Cao X, Yang J, Wu X, Liu Y, Wan B, et al. Rifampin-resistance-associated mutations in the rifampin-resistance-determining region of the rpoB gene of Mycobacterium tuberculosis clinical isolates in Shanghai, PR China. J Med Microbiol. 70(3):001317.

184. Hameed HA, Tan Y, Islam MM, Lu Z, Chhotaray C, Wang S, et al. Detection of Novel Gene Mutations Associated with Pyrazinamide Resistance in Multidrug-Resistant Mycobacterium tuberculosis Clinical Isolates in Southern China . Infect Drug Resist. 22 de enero de 2020;13:217-27.

185. Ang MLT, Zainul Rahim SZ, de Sessions PF, Lin W, Koh V, Pethe K, et al. EthA/R-Independent Killing of Mycobacterium tuberculosis by Ethionamide. Front Microbiol [Internet]. 2017 [citado 3 de noviembre de 2022];8. Disponible en: https://www.frontiersin.org/articles/10.3389/fmicb.2017.00710

186. Sowajassatakul A, Prammananan T, Chaiprasert A, Phunpruch S. Overexpression of eis without a mutation in promoter region of amikacin- and kanamycin-resistant Mycobacterium tuberculosis clinical strain. Ann Clin Microbiol Antimicrob. 16 de julio de 2018;17(1):33.

187. Chaiyachat P, Chaiprasert A, Nonghanphithak D, Smithtikarn S, Kamolwat P, Pungrassami P, et al. Whole-genome analysis of drug-resistant Mycobacterium tuberculosis reveals novel mutations associated with fluoroquinolone resistance. Int J Antimicrob Agents. 1 de septiembre de 2021;58(3):106385.

188. Gomes LC, Campino S, Marinho CRF, Clark TG, Phelan JE. Whole genome sequencing reveals large deletions and other loss of function mutations in Mycobacterium tuberculosis drug resistance genes. Microb Genomics. 10 de diciembre de 2021;7(12):000724.

189. Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, Hannan M, et al. Functional and evolutionary genomics of Mycobacterium tuberculosis: Insights from genomic deletions in 100 strains. Proc Natl Acad Sci. 6 de abril de 2004;101(14):4865-70.

190. Kardan-Yamchi J, Amini S, Hamzelou G, Rahimi Foroushani A, Ghodousi A, Cirillo DM, et al. Assessment of the GenoType MTBDRsl VER 2.0 compared to the phenotypic drug susceptibility testing and whole genome sequencing for the rapid detection of resistance to fluoroquinolone and second-line injectable drugs among rifampicin-resistant Mycobacterium tuberculosis isolates. Arch Microbiol. 1 de septiembre de 2021;203(7):3989-96.

191. Cox H, Goig GA, Salaam-Dreyer Z, Dippenaar A, Reuter A, Mohr-Holland E, et al. Whole-Genome Sequencing Has the Potential To Improve Treatment for Rifampicin-Resistant Tuberculosis in High-Burden Settings: a Retrospective Cohort Study. J Clin Microbiol. 16 de marzo de 2022;60(3):e02362-21.

192. Soetaert K, Ceyssens PJ, Boarbi S, Bogaerts B, Delcourt T, Vanneste K, et al. Retrospective evaluation of routine whole genome sequencing of Mycobacterium tuberculosis at the Belgian National Reference Center, 2019. Acta Clin Belg. 3 de septiembre de 2022;77(5):853-60.

193. Shrestha D, Maharjan B, Thida Oo NA, Isoda N, Nakajima C, Suzuki Y. Molecular analysis of streptomycin-resistance associating genes in Mycobacterium tuberculosis isolates from Nepal. Tuberculosis. diciembre de 2020;125:101985.

194. Alame Emane AK, Guo X, Takiff HE, Liu S. Drug resistance, fitness and compensatory mutations in Mycobacterium tuberculosis. Tuberculosis. 1 de julio de 2021;129:102091.

195. Valafar SJ. Systematic Review of Mutations Associated with Isoniazid Resistance Points to Continuing Evolution and Subsequent Evasion of Molecular Detection, and Potential for Emergence of Multidrug Resistance in Clinical Strains of Mycobacterium tuberculosis. Antimicrob Agents Chemother. 17 de febrero de 2021;65(3):e02091-20.

196. He W, Tan Y, Liu C, Wang Y, He P, Song Z, et al. Drug-Resistant Characteristics, Genetic Diversity, and Transmission Dynamics of Rifampicin-Resistant Mycobacterium tuberculosis in Hunan, China, Revealed by Whole-Genome Sequencing. Microbiol Spectr. 16 de febrero de 2022;10(1):e01543-21.

197. Kabir S, Tahir Z, Mukhtar N, Sohail M, Saqalein M, Rehman A. Fluoroquinolone resistance and mutational profile of gyrA in pulmonary MDR tuberculosis patients. BMC Pulm Med. 11 de mayo de 2020;20(1):138.

198. Rinder H, Mieskes KT, Löscher T. Heteroresistance in Mycobacterium tuberculosis. Int J Tuberc Lung Dis. 1 de abril de 2001;5(4):339-45.

199. Aung YW, Faksri K, Sangka A, Tomanakan K, Namwat W. Heteroresistance of Mycobacterium tuberculosis in the Sputum Detected by Droplet Digital PCR. Biology. abril de 2023;12(4):525.

200. El Garch F, Lismond A, Piddock LJV, Courvalin P, Tulkens PM, Van Bambeke F. Fluoroquinolones induce the expression of patA and patB, which encode ABC efflux pumps in Streptococcus pneumoniae. J Antimicrob Chemother. octubre de 2010;65(10):2076-82.

201. Cohen KA, Stott KE, Munsamy V, Manson AL, Earl AM, Pym AS. Evidence for Expanding the Role of Streptomycin in the Management of Drug-Resistant Mycobacterium tuberculosis. Antimicrob Agents Chemother. 20 de agosto de 2020;64(9):e00860-20.

202. Conkle-Gutierrez D, Kim C, Ramirez-Busby SM, Modlin SJ, Mansjö M, Werngren J, et al. Rare alternative second line injectable drug resistance markers identified by gene-wise genome wide association in *M. tuberculosis* with unexplained resistance [Internet]. Genomics; 2021 nov [citado 6 de noviembre de 2022]. Disponible en: http://biorxiv.org/lookup/doi/10.1101/2021.11.23.469801

203. Matsui T, Pinhata JMW, Rabello MC da S, Brandão AP, Ferrazoli L, Leão SC, et al. Frequency of first and second-line drug resistance-associated mutations among resistant Mycobacterium tuberculosis clinical isolates from São Paulo, Brazil. Mem Inst Oswaldo Cruz. 2020;115:e200055.

204. Li Q, Gao H, Zhang Z, Tian Y, Liu T, Wang Y, et al. Mutation and Transmission Profiles of Second-Line Drug Resistance in Clinical Isolates of Drug-Resistant Mycobacterium tuberculosis From Hebei Province, China. Front Microbiol [Internet]. 2019 [citado 6 de noviembre de 2022];10. Disponible en: https://www.frontiersin.org/articles/10.3389/fmicb.2019.01838

205. Coll F, Phelan J, Hill-Cawthorne GA, Nair MB, Mallard K, Ali S, et al. Genome-wide analysis of multi- and extensively drug-resistant Mycobacterium tuberculosis. Nat Genet. febrero de 2018;50(2):307-16.

206. Ali A, Hasan Z, McNerney R, Mallard K, Hill-Cawthorne G, Coll F, et al. Whole Genome Sequencing Based Characterization of Extensively Drug-Resistant Mycobacterium tuberculosis Isolates from Pakistan. PLOS ONE. 26 de febrero de 2015;10(2):e0117771.

207. López-Cortés A, Zambrano AK, Guevara-Ramírez P, Echeverría BA, Guerrero S, Cabascango E, et al. Clinical, genomics and networking analyses of a high-altitude native American Ecuadorian patient with congenital insensitivity to pain with anhidrosis: a case report. BMC Med Genomics. 17 de agosto de 2020;13:113.

208. Cudahy PG, Andrews JR, Bilinski A, Dowdy DW, Mathema B, Menzies NA, et al. Spatially targeted screening to reduce tuberculosis transmission in high incidence settings: a systematic review and synthesis. Lancet Infect Dis. marzo de 2019;19(3):e89-95.

209. Cancino-Muñoz I, López MG, Torres-Puente M, Villamayor LM, Borrás R, Borrás-Máñez M, et al. Population-based sequencing of Mycobacterium tuberculosis reveals how current population dynamics are shaped by past epidemics. Lewnard J, Franco E, Meehan CJ, Walker TM, editores. eLife. 26 de julio de 2022;11:e76605.

210. Meumann EM, Horan K, Ralph AP, Farmer B, Globan M, Stephenson E, et al. Tuberculosis in Australia's tropical north: a population-based genomic epidemiological study. Lancet Reg Health – West Pac [Internet]. 1 de octubre de 2021 [citado 17 de abril de 2023];15. Disponible en: https://www.thelancet.com/journals/lanwpc/article/PIIS2666-6065(21)00138-3/fulltext

211. Pescarini JM, Simonsen V, Ferrazoli L, Rodrigues LC, Oliveira RS, Waldman EA, et al. Migration and tuberculosis transmission in a middle-income country: a cross-sectional study in a central area of São Paulo, Brazil. BMC Med. 30 de abril de 2018;16(1):62.

212. Sanabria GE, Sequera G, Aguirre S, Méndez J, dos Santos PCP, Gustafson NW, et al. Phylogeography and transmission of Mycobacterium tuberculosis spanning prisons and surrounding communities in Paraguay. Nat Commun. 19 de enero de 2023;14(1):303.

213. Colijn C, Cohen T. Whole-genome sequencing of Mycobacterium tuberculosis for rapid diagnostics and beyond. Lancet Respir Med. 1 de enero de 2016;4(1):6-8.

214. Goldfeder RL, Priest JR, Zook JM, Grove ME, Waggott D, Wheeler MT, et al. Medical implications of technical accuracy in genome sequencing. Genome Med. 2 de marzo de 2016;8(1):24.

215. World Health Organization. Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance [Internet]. Geneva, Switzerland: World Health Organization; 2021 [citado 19 de agosto de 2022]. 86 p. Disponible en: https://www.who.int/publications-detail-redirect/9789240028173

216. Mahé P, El Azami M, Barlas P, Tournoud M. A large scale evaluation of TBProfiler and Mykrobe for antibiotic resistance prediction in Mycobacterium tuberculosis. PeerJ. 1 de mayo de 2019;7:e6857.

217. van Beek J, Haanperä M, Smit PW, Mentula S, Soini H. Evaluation of whole genome sequencing and software tools for drug susceptibility testing of Mycobacterium tuberculosis. Clin Microbiol Infect. enero de 2019;25(1):82-6.