



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



Tuberculosis pulmonar: un abordaje global hacia la caracterización de cepas aisladas y la respuesta del hospedero

Lic. Bioq. Joaquín Ignacio Hurtado Gutiérrez

Tesis de maestría en Ciencias Biológicas
Opción Biología Celular y Molecular, PEDECIBA

Laboratorio de interacciones hospedero-patógeno / UBM

Institut Pasteur de Montevideo.

Marzo 2022, Montevideo Uruguay

Tuberculosis pulmonar: un abordaje global hacia la caracterización de cepas aisladas y la respuesta del hospedero

Lic. Joaquín Ignacio Hurtado Gutiérrez

Tesis presentada con el objetivo de obtener el título de Magister en Ciencias Biológicas, opción Biología Celular y Molecular, en el marco del Programa PEDECIBA

Tutor: Dr. Gonzalo Greif

Tribunal: Dra. Andrea Villarino (presidenta)
 Dr. Juan Pablo Tosar
 Dr. Darío García de Viedma

Tesis realizada en el Laboratorio de Interacciones Hospedero-Patógeno / UBM, Institut Pasteur de Montevideo, entre abril de 2019 y diciembre de 2021

AGRADECIMIENTOS

En primer lugar, quiero agradecer a mi tutor de maestría Gonzalo Greif por la confianza, por la independencia que me dió durante todo este trayecto, por enseñarme todo lo que sé, por ser un gran orientador y un amigo.

A Carlos Robello por abrirme las puertas de su laboratorio y por estar siempre dispuesto al diálogo franco y honesto.

A toda la UBM que fue mi segunda casa durante la tesina de grado y durante esta maestría. Me llevo grandes compañeros y compañeras que son una inspiración para mi, y también me llevo buenos amigos y amigas.

Agradezco a las fuentes de financiación que me permitieron llevar adelante mis estudios: A la Agencia Nacional de Investigación e Innovación, al Programa de Desarrollo de las Ciencias Básicas y al Fondo COVID-19 del Ministerio de Educación y Cultura.

Por último, las personas más importantes que verdaderamente son sostén de estos proyectos y de la vida:

a mi madre, a mi padre, a mis hermanos, a mis amigos,

Gracias!.

RESUMEN

La tuberculosis (TB) es una de las primeras causas de muerte asociadas a un solo agente infeccioso en el mundo. En el año 2020 se registraron 10,4 millones de casos nuevos y 1,3 millones de muertes. La enfermedad causada por la bacteria *Mycobacterium tuberculosis* es prevenible y curable. Sin embargo, en las últimas décadas, Uruguay viene experimentado un aumento lento pero sostenido en la incidencia de los casos, opuesto a la tendencia mundial y regional. Existen siete linajes descritos dentro de *Mycobacterium tuberculosis*, de los cuales algunos son asociados con un peor pronóstico y mayor ocurrencia de resistencia a fármacos.

Distintos métodos moleculares se han utilizado a lo largo del tiempo para caracterizar genótipicamente las distintas cepas. El avance tecnológico en materia de secuenciación masiva, su mayor disponibilidad y bajo costo, hacen que esta herramienta sea cada día más utilizada no sólo a nivel de investigación sino también en el diagnóstico.

En este trabajo de maestría, se desarrolló una metodología basada en secuenciación masiva, como método de tipificación y búsqueda de resistencia a fármacos de cepas de *M. tuberculosis*. Para ello, se realizó la búsqueda de polimorfismos de nucleótido simple en genes que discriminan los diferentes linajes y al mismo tiempo, se analizó la presencia de mutaciones en las regiones asociadas a resistencia a fármacos. Se realizó la puesta a punto de esta metodología en aislados de una población del sistema penitenciario y se compararon con cepas circulantes en la comunidad. Como resultado se observó la presencia de una cepa particular representada de forma significativa en el sistema penitenciario con muy baja circulación a nivel comunitario (linaje X). Por otra parte, se observaron mutaciones de baja asociación a resistencia a algunos fármacos anti-tuberculosos.

Conocer los linajes que circulan en la población es información importante para identificar la aparición de brotes, detectar resistencia y prevenir el aumento de casos. Pero no sólo un seguimiento epidemiológico exhaustivo y con toda la información posible es necesario para controlar esta enfermedad, el desarrollo de un diagnóstico temprano mejoraría enormemente las posibilidades de éxito en el tratamiento del paciente con TB. Como la confirmación del cuadro clínico implica en muchos casos realizar un cultivo de la bacteria, insumiendo varias semanas, contar con un marcador molecular de enfermedad a pie de cama resultaría valioso

para acortar estos plazos. Por lo que, profundizando en la respuesta del hospedero, se realizó la búsqueda de ARNs pequeños circulantes en suero de una población de pacientes con TB grave que requirieron internación en Cuidados Intensivos. Los ARNs pequeños circulantes en sangre, son actualmente utilizados como biomarcadores en el diagnóstico de otras enfermedades. En este trabajo de maestría, se realizó la comparación de esta población de ARNs pequeños circulantes en suero de pacientes con TB grave y un grupo control buscando en particular, identificar micro-ARNs (miARNs) capaces de discriminar entre los grupos y entre estadios tempranos y avanzados de la enfermedad. Del análisis de la abundancia diferencial de las secuencias obtenidas, encontramos tres miARNs con posible valor pronóstico de la evolución de la enfermedad y se generaron datos que permiten considerar otros ARNs pequeños (tRNAs, ncRNA, snoRNAs, etc) para mejorar la evaluación de estos pacientes.

Palabras clave: Tuberculosis, tipificación, ARNs pequeños, diagnóstico molecular.

TABLA DE CONTENIDOS

1. INTRODUCCIÓN	6
1.1. Generalidades de la bacteria	6
1.2. Métodos de tipificación	8
1.3. Generalidades de la enfermedad.	13
1.4. Situación epidemiológica actual de la tuberculosis	15
1.5. Tratamiento contra la tuberculosis	17
1.6. Tuberculosis resistente	19
1.7. Situación epidemiológica de la tuberculosis en Uruguay	20
1.8. Tuberculosis en la Unidad de Cuidados Intensivos	22
1.9. Marcadores moleculares en el hospedero	23
2. OBJETIVOS	25
2.1. Objetivos generales	25
2.2. Objetivos específicos	25
3. MATERIALES Y MÉTODOS	26
3.1. Caracterización de aislados de <i>Mycobacterium tuberculosis</i> .	26
3.1.1. Obtención de muestras	26
3.1.2. Selección de genes para tipificación de cepas y búsqueda resistencia	27
3.1.3. Amplificación en múltiplex	29
3.1.4. Preparación de bibliotecas para secuenciación.	32
3.1.5. Análisis de los datos	34
3.2. Búsqueda de marcadores moleculares en suero	35
3.2.1. Población de estudio	35
3.2.2. Extracción de ARN	36
3.2.3. Secuenciación de ARNs pequeños circulantes en suero	36
3.2.4. Análisis de datos	38
3.2.5. Validación de miARNs	39
4. RESULTADOS	42
PARTE A - Caracterización de aislados en población de personas privadas de libertad	42
4.1. Amplificación de pools	42
4.2. Análisis primario de secuencias obtenidas	44
4.3. Mapeo de secuencias	46

4.4.	<i>Clasificación de linajes</i>	49
4.5.	<i>Resistencia a fármacos antituberculosis</i>	52
	<i>PARTE B - Búsqueda de marcadores con valor pronóstico de la enfermedad</i>	55
4.6.	<i>ARNs pequeños en suero (miARN)</i>	55
4.7.	<i>Secuenciación de bibliotecas para miARNs</i>	55
4.8.	<i>Abundancia diferencial de microARNs</i>	58
4.9.	<i>Validación de miARNs candidatos para identificación de TB</i>	63
5.	DISCUSIÓN Y CONCLUSIONES	66
	<i>PARTE A- Tipificación de aislados y búsqueda de resistencia</i>	66
5.1.	<i>Tipificación de cepas de una población de personas privadas de libertad</i>	66
5.2.	<i>Resistencia en aislados del sistema penitenciario</i>	69
5.3.	<i>Puesta a punto y consideraciones finales</i>	71
	<i>PARTE B- Búsqueda de marcadores moleculares de tuberculosis</i>	72
5.4.	<i>miARNs como marcadores moleculares de enfermedad</i>	72
5.5.	<i>Abundancia diferencial de miARNs</i>	73
5.6.	<i>Validación de los miARNs</i>	75
6.	PERSPECTIVAS	77
7.	BIBLIOGRAFÍA	79
	ANEXO I	85

1. INTRODUCCIÓN

1.1. Generalidades de la bacteria

Mycobacterium tuberculosis (Mtb) es un bacilo que carece de motilidad, es ácido-alcohol resistente y débilmente Gram-positivo. Microscópicamente se observan como bastones rectos o curvos de 1 a 4 μm de largo. Mtb o bacilo de Koch, llamado así en honor a su descubridor (en 1882) Robert Koch, presenta una pared celular única compuesta por ácidos micólicos la cual juega un rol importante en la función celular, así como en su patogenicidad. La compleja pared celular le confiere características de extrema hidrofobicidad, resistencia a la deshidratación y a la acidez o alcalinidad y se encuentra conservada dentro del género (Sakamoto, 2012).

Mtb es un patógeno obligado de humanos que pertenece al género *Mycobacterium* del filo *Actinobacteria*, que comprende cerca de 170 especies distintas. El género se puede dividir a su vez en micobacterias de crecimiento rápido y de crecimiento lento (Figura 1). Mtb se encuentra entre las de crecimiento lento junto con otras especies causantes de enfermedades en humanos y animales, agrupadas bajo el nombre del Complejo *Mycobacterium Tuberculosis* (MTBC) (Bañuls et al., 2015).

Existe evidencia que sugiere que el MTBC emergió como un solo grupo clonal de un pool de cepas recombinantes de micobacterias dando lugar al ancestro *Mycobacterium canetti*. Este linaje presenta una morfología inusual para lo que conocemos hoy como micobacterias; es muy poco frecuente aislarlo de infecciones humanas y no se ha establecido hasta la fecha un link epidemiológico en la transmisión humano-humano. Se cree que es una micobacteria potencialmente ambiental, con una alta tasa de recombinación y transferencia horizontal de genes (Boritsch et al., 2016). Sin embargo, esta capacidad se perdió más adelante en la descendencia del MTBC, que experimentó una expansión clonal y una diferenciación menor, pero que de alguna forma ganó mayor capacidad de transmisibilidad y de virulencia (Orgeur & Brosch, 2018).

El MTBC incluye especies de *Mycobacterium* con una muy alta similitud en la secuencia de la subunidad 16S del ARN ribosomal. Esta particular similitud va en línea con la hipótesis del origen de un mismo ancestro común, aunque difieren en el tropismo por el hospedero, fenotipo y patogenicidad. Dentro del MTBC se encuentran los patógenos exclusivos de humanos (*Mycobacterium tuberculosis*, *Mycobacterium africanum*), pero también patógenos exclusivos de roedores (*Mycobacterium microti*), leones marinos (*Mycobacterium pinnipedi*), de ovejas y cabras (*Mycobacterium caprae*). También, en este grupo se encuentra *Mycobacterium bovis* que es patógeno preferentemente de bovinos, aunque en menor medida puede infectar humanos.



Figura 1. Árbol filogenético del género *Mycobacterium* basado en la secuencia 16S ARNr. A la izquierda se encuentra el árbol con expansión de especie de crecimiento lento, donde se recuadra en rojo al *Complejo de Mycobacterium tuberculosis*. A la derecha se encuentra el árbol con expansión de especie de crecimiento rápido, que en general no se encuentran asociadas a enfermedades en humanos. Imagen modificada de Tortoli E. et al, Nontuberculous Mycobacteria, 2019.

Dentro de la especie *M. tuberculosis* se describen siete linajes diferentes. Los linajes 5 y 6 (L5 y L6 respectivamente) son los más ancestrales dentro del MTBC que infectan humanos, y los linajes L1 y L7 divergieron antes de la delección de una región particular del genoma

(región TbD1). Esta deleción da origen a los linajes más modernos o más recientes en la evolución (L2, L3 y L4) (Figura 2).

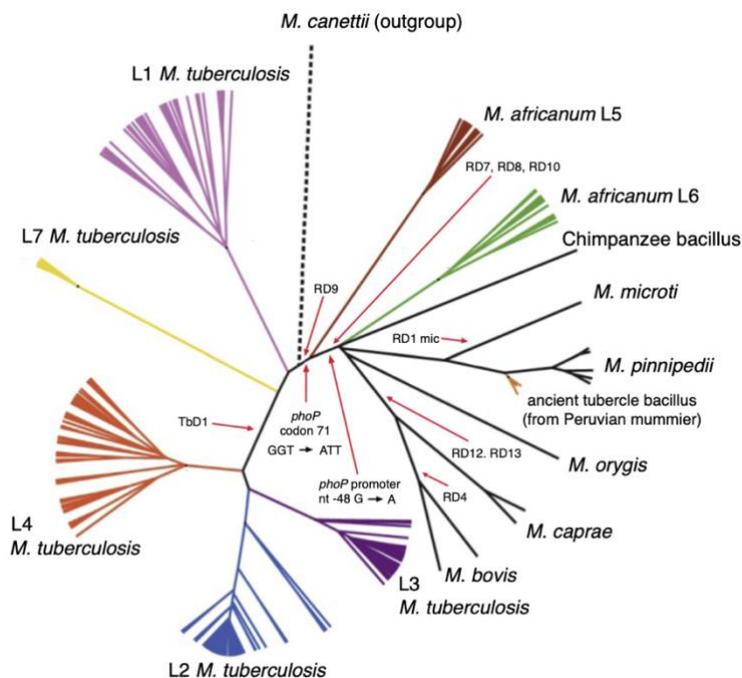


Figura 2. Árbol filogenético de los diferentes linajes que componen el MTBC, basado en SNPs de 261 genomas de micobacterias. El linaje L1 corresponde a Este-Indio-Africano (EAI), L2 a las cepas Beijín, L3 a las cepas de Asia Central (CAS), L4 a las cepas Euro-Americanas (T, Haarlem, LAM, S, X), L5 a *M. africanum* 1 y L6 a *M. africanum* 2. Por último, L7 corresponde estrictamente a las regiones de Etiopía y el Cuerno de África. Imagen tomada de Michael O. et al, Microbiology, 2018.

1.2. Métodos de tipificación

Si bien es sabido que los miembros del MTBC son parientes muy cercanos y comparten un muy alto porcentaje de identidad en su secuencia nucleotídica, también se han reportado variantes asociadas con peor pronóstico y mayor gravedad (Ford et al., 2013). Uno de los avances fundamentales de las últimas décadas para la visualización de esta diversidad ha sido el desarrollo de técnicas de secuenciación de genoma completo.

El MTBC ha evolucionado entre varios linajes adaptados a humanos y linajes que se adaptaron a animales domésticos y salvajes. Los linajes adaptados a humanos son patógenos obligados y no existen reservorios animales o ambientales para estos. Los linajes L1, L2, L3, L4 y L7 pertenecen a *Mycobacterium tuberculosis* propiamente dicho y los linajes L5 y L6 a *M. africanum* (Figura 3).

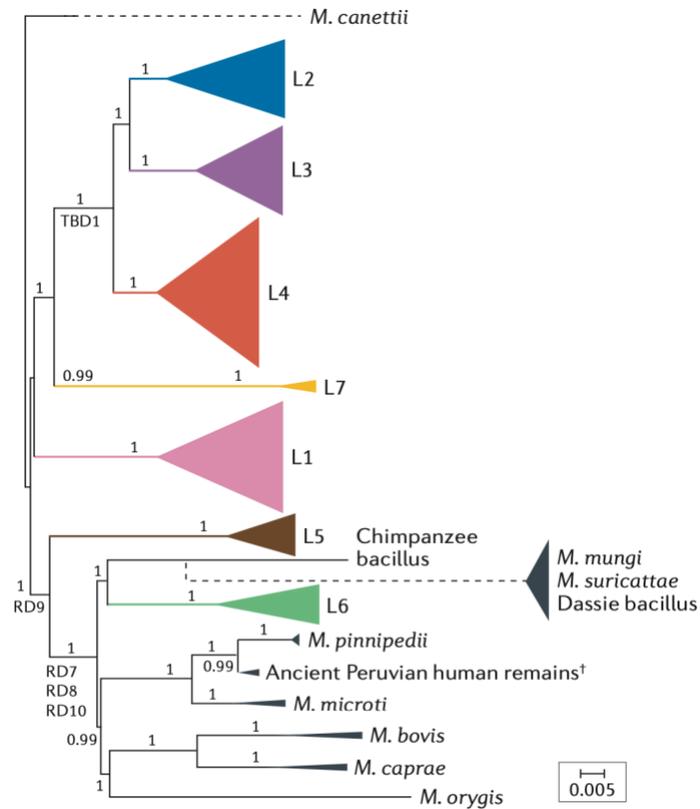


Figura 3. Árbol filogenético del MTBC. Filogenia basada en el genoma completo del Complejo *Mycobacterium tuberculosis* con raíz en *Mycobacterium canettii*. Los triángulos representan una simplificación de las ramas de los principales linajes. Los linajes específicos de humanos se encuentran en color y aquellos linajes adaptados a animales en gris. El largo de la escala representa el número de sustituciones por sitio nucleotídico. Imagen tomada de Gagneux S. et al. Nature, 2019.

Algunos linajes se encuentran más restringidos geográficamente que otros y se hipotetiza que estas variantes pueden estar mejor adaptadas a la población humana local (Gagneux, 2018). Los linajes L5 y L6 están restringidos al oeste africano mientras que L2 y L4 son los más dispersos a nivel mundial, siendo el linaje 4 el más extendido. Este linaje además puede ser subdividido en múltiples sublinajes: L4.1.1 o X, L4.1.2 o Haarlem, L4.3 o LAM, L4.6.2 o Cameroon, etc. (Figura 4).

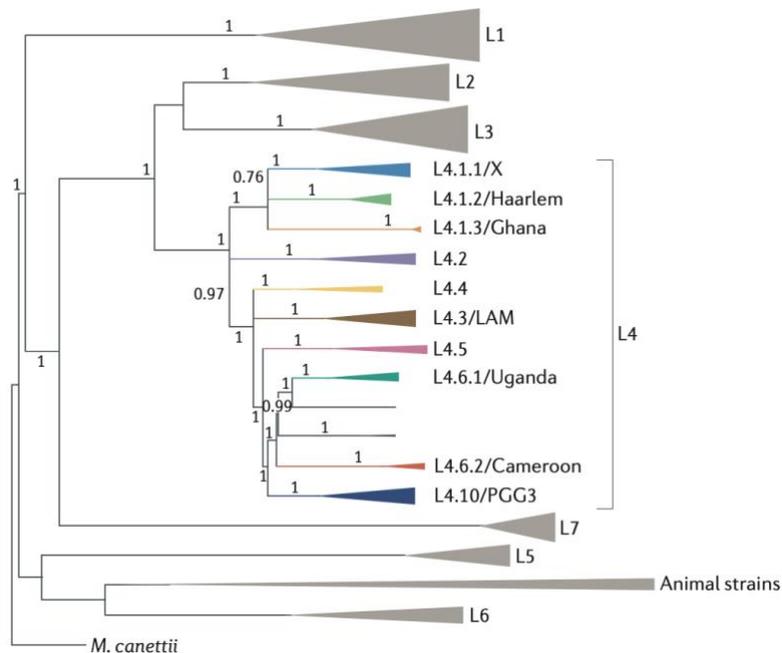


Figura 4. Árbol filogenético del Complejo de *Mycobacterium tuberculosis* con amplificación al linaje 4 que se subdivide en al menos diez sub-linajes. Imagen tomada de Gagneux S. et al. Nature, 2019.

Diferentes métodos moleculares se han utilizado para identificar los principales linajes de *M. tuberculosis* y así poder establecer contactos epidemiológicos y seguimiento de brotes. La primera técnica utilizada consiste en la digestión enzimática del cromosoma de ADN circular, y posteriormente, mediante una sonda marcada, determinar la presencia de la región IS6110. Esta región tiene una tasa de evolución mayor comparada con el resto del genoma y en base al número de copias se pueden clasificar los diferentes linajes (van Embden et al., 1993). El análisis de IS6110 se utilizó ampliamente a partir de la década del 90 y si bien permitió mejorar la capacidad de agrupamiento de los aislados, contaba con algunas desventajas como el bajo poder de discriminación de cepas cuando los números de copias de IS eran bajos.

Posteriormente se desarrollaron técnicas basadas en PCR como el “spoligotyping”. Este método se basa en la amplificación de un locus particular llamado repetido directo (DR). Esta región se conforma de sucesivas repeticiones de DRs (36 pb de largo) interespaciadas por secuencias únicas no repetidas (de entre 34 a 41 pb). A través de una hibridación en membrana, con oligonucleótidos correspondientes a las secuencias únicas espaciadoras, se determina la presencia o ausencia de cada región (Figura 5). De acuerdo a la cantidad de espaciadores que tenga una cepa se la clasifica en las diferentes familias. Por ejemplo, la cepa de referencia de *Mtb* H37Rv, lleva ese nombre ya que contiene 37 de los 43 espaciadores posibles. El spoligotyping es eficiente en cuanto a la tipificación de las cepas de MTBC pudiendo, por ejemplo, diferenciar entre *M. tuberculosis* y *M. bovis*, así como los diferentes linajes dentro de *Mtb* (Kamerbeek et al., 1997).

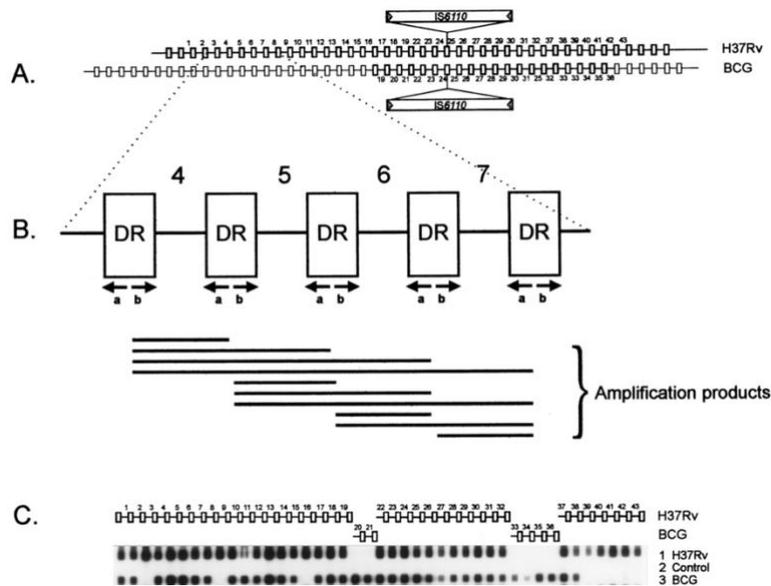


Figura 5. Esquematización de la estructura del locus DR (repetido directo). (A) En rectángulos se representan los locus DR para los cromosomas de *Mycobacterium tuberculosis* y *Mycobacterium bovis* BCG que tienen 48 y 41 DRs respectivamente. Entre los DR se encuentran las regiones espaciadoras que son secuencias únicas y no repetitivas. *Mycobacterium bovis* BCG tiene 6 espaciadores mientras que *Mycobacterium tuberculosis* tiene 37. (B) Esquematización para la amplificación de las regiones espaciadoras por PCR. Los amplicones tienen diferentes largos y los pares de cebadores pueden comenzar la amplificación en cualquier DR. (C) Patrones de hibridación o “spoligotyping” de las regiones espaciadoras por sonda en membrana. El orden de los espaciadores es en el mismo orden que se encuentran en el genoma. Imagen tomada de Kamerbeek J. et al, Journal of Clinical Microbiology, 1997.

Más adelante, a principios del 2000, luego de la obtención de la secuencia del genoma de referencia (Cole et al., 1998), se desarrolló un segundo método basado en PCR que consta del estudio de minisatélites. Se trata de un análisis en alta resolución de repetidos en tándem de número variable (VNTR: variable number of tandem repeats) de elementos genéticos de Mtb llamados unidades repetidas interespaciadas de micobacterias (MIRUs: mycobacteria interspead repeated units) (Supply et al., 1997). Los MIRU tienen entre 50 a 100 nucleótidos de largo y pueden presentar entre 0 a 25 repeticiones de cada unidad. Brevemente, se amplifican dichas regiones, y se evalúan los tamaños de los productos de dicha amplificación (Figura 6). Como el tamaño está dado por el número de copias para cada MIRU, se puede establecer un código de números para describir cuántos repetidos hay para cada locus. La tipificación a través de esta técnica fue mejorando, pasando de 12, 15 y finalmente 24 loci que es lo que se utiliza como técnica *gold standar* en la actualidad.

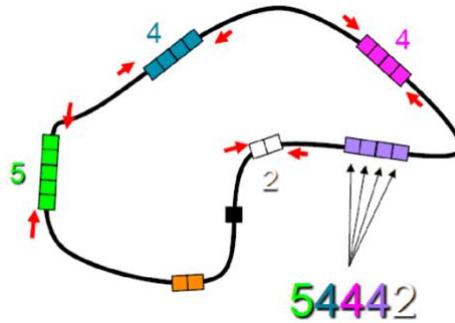


Figura 6. Esquema para la amplificación de los MIRU. Cada color representa un locus y cada cuadrado la cantidad de repetidos (que se especifica en números). De esta manera cada cepa forma un código numérico único. En rojo se esquematizan los pares de cebadores para cada locus. Imagen tomada de Takiff H, Mooc Institut Pasteur Paris, 2018.

El desarrollo de las técnicas de secuenciado masivo llevó a un crecimiento exponencial de genomas completos (WGS: whole genome sequencing) y revolucionó la caracterización de la diversidad genética de miles de organismos. El uso de WGS en la caracterización de las cepas de Mtb tiene un impacto directo en la salud pública. Permite detectar y predecir variantes con resistencia fenotípica a antibióticos contra la tuberculosis y por lo tanto influir en las decisiones clínicas, así como también, poder realizar una vigilancia epidemiológica de los linajes circulantes y de los mecanismos de resistencia.

El avance en estas tecnologías permitió desarrollar técnicas basadas en secuenciación profunda de genes objetivo y secuenciación de todo el genoma. Ambas aproximaciones se basan en los mismos principios y equipamiento, pero de acuerdo al objetivo que se busca se puede optar por una u otra. Por ejemplo, si se quisieran detectar mutaciones puntuales en genes de resistencia a fármacos, la secuenciación profunda en estos genes podría ser la mejor alternativa para el diagnóstico, siendo posible realizarla directamente de muestras clínicas. Esto resulta en un ahorro económico, pero fundamentalmente de tiempo ya que se ahorraría el tiempo que insume el cultivo bacteriano. Por el contrario, para realizar una secuenciación de todo el genoma se debe partir de ADN de alta calidad, y en este caso es necesario esperar el cultivo. Sin embargo, la secuenciación genómica resulta una mejor aproximación para realizar una filogenia completa, además de poder identificar otras mutaciones o variantes no reportadas aún.

Los diferentes métodos de tipificación molecular son utilizados para definir cadenas de transmisión o agrupamientos de cepas de Mtb en base a la observación de motivos genéticos en sus genomas. Por ejemplo, la presencia o ausencia de regiones espaciadoras con spoligotyping o la cantidad de repetidos de un loci de los 24-MIRU-VNTR descritos, hacen posible generar una filogenia de las cepas analizadas. La epidemiología molecular trata de

definir el hilo epidemiológico entre un caso y otro y para ello resulta importante conocer los tiempos de transmisión entre ellos. Esto es posible ya que los genomas acumulan mutaciones con el tiempo, lo que significa, a grandes rasgos, que genomas similares provienen de un ancestro común cercano, mientras que, si presentan marcadas diferencias, mayor es la distancia que los separa. La clave en la elucidación de estos tiempos es encontrar motivos genéticos que varíen lo suficiente y sean heredables, pero al mismo tiempo mantengan cierto grado de conservación. En este sentido las técnicas de WGS tienen un alto poder discriminatorio cuando se trata de analizar las dinámicas de transmisión, y es de particular importancia en organismos cuyas “distancias genéticas” son cercanas (Meehan et al., 2018).

En 2014, Coll et al. desarrollaron un método basado en WGS para el estudio de los polimorfismos de nucleótido simple (SNP: single nucleotide polymorphism) como marcadores de variación genética para el análisis filogenético de cepas de Mtb (Coll et al., 2014). Se identificaron 62 SNPs canónicos en genes esenciales bajo selección negativa, que resuelven la diversidad genómica de cada cepa del MTBC. Los genomas acumulan mutaciones con el tiempo y a través de estudios retrospectivos basados en WGS se puede determinar la tasa de mutación. En el caso de Mtb se reportaron 0,3 a 0,5 SNPs por genoma por año (1×10^{-7} sustituciones por nucleótido por año), lo que es una tasa baja comparada con otras especies de bacterias (Barbier & Wirth, 2016). Este método de tipificación tiene un poder de discriminación de los aislados más potente que las técnicas basadas en PCR. Para el caso del spoligotyping se ha determinado que no es capaz de discriminar eventos de transmisión con menos de 200 años de antigüedad, ya que las regiones del genoma que son estudiadas son altamente conservadas (Meehan et al., 2018).

1.3. Generalidades de la enfermedad.

Mtb es el agente etiológico de la tuberculosis (TB), una enfermedad infecciosa que produce en la actualidad, cerca de 10.4 millones de infecciones en humanos y que causó la muerte de 1.3 millones de personas en 2020 (Geneva: World Health Organization, 2021), superando al HIV como el agente infeccioso causante de más muertes a nivel mundial en el año 2019 y sólo superado en 2020 y 2021 por el SARS-CoV-2. La tuberculosis se transmite de persona a persona a través de partículas expelidas en la tos y afecta principalmente a adultos mayores de 45 años, siendo más frecuente en hombres que mujeres. Esta enfermedad afecta principalmente los pulmones, cuando las pequeñas partículas de Mtb logran llegar hasta el tracto respiratorio bajo donde son fagocitadas por los macrófagos y células dendríticas residentes. Sin embargo, también puede afectar otros órganos. Dentro de los pulmones se dispara una respuesta inflamatoria local con el objetivo de activar la respuesta inmune innata y reclutar células inmunes al sitio de infección. De esta manera se forma un agregado de varios tipos de células (macrófagos, neutrófilos, células T, etc) llamado granuloma (Figura 7), que es un marcador histopatológico de la enfermedad (Bussi & Gutierrez, 2019). Uno de los mecanismos más importantes en la patogenicidad de la bacteria, es la inhibición en la

maduración del fagosoma-lisosoma, que se produce a través de la expresión de un complejo multiproteico llamado ESX-1. De esta forma, la bacteria puede inmunomodular a la célula fagocítica, sobrevivir y reproducirse dentro de ella. Por el contrario, el bacilo atenuado de Calmette y Guérin (BCG) utilizado en la vacuna para la TB, perdió parte de los genes que codifican para el sistema de secreción de ESX-1, debido a una delección de la región RD1, por lo que carece de efecto patológico (Orgeur & Brosch, 2018).

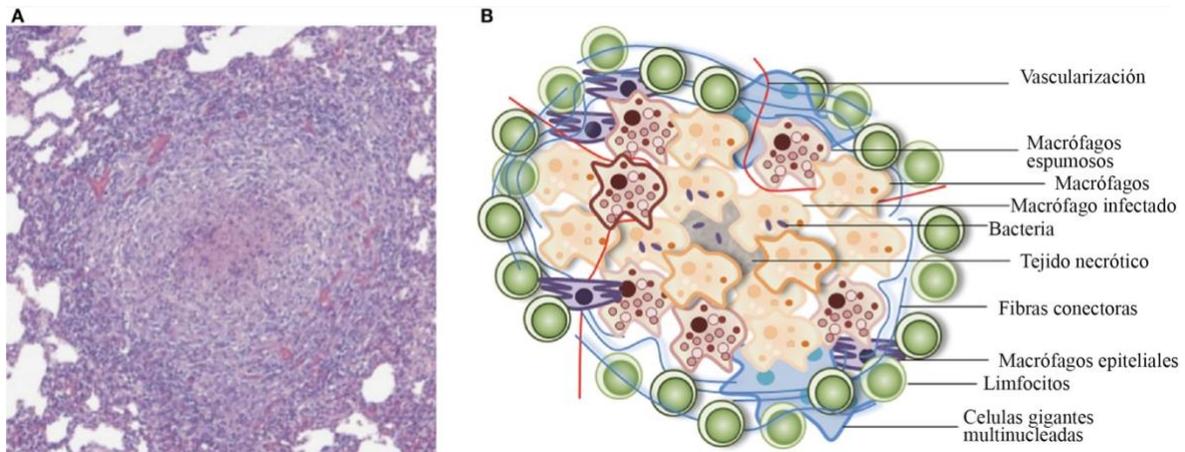


Figura 7. Arquitectura de un granuloma. (A) Muestra histológica obtenida de tejido pulmonar de cerdo con tinción hematoxilina-eosina. Granuloma representativo con centro necrótico. (B) Esquema de los componentes celulares que forman parte del granuloma. Imagen modificada de Guirado et al. *Frontiers in Immunology*, 2013.

La mayoría de las personas que se infectan con *M. tuberculosis* son asintomáticas y se estima que cerca de un cuarto de la población mundial se encuentra infectada. En la mayoría de los casos la infección es contenida y la bacteria entra en estado de latencia (LTBI: latent tuberculosis infection) y solamente entre 5-10% de personas infectadas desarrollarán la enfermedad activa (Figura 8). La probabilidad de la activación de la enfermedad aumenta notoriamente entre personas con VIH y también con comorbilidades como desnutrición, diabetes, fumadores y alcoholistas.

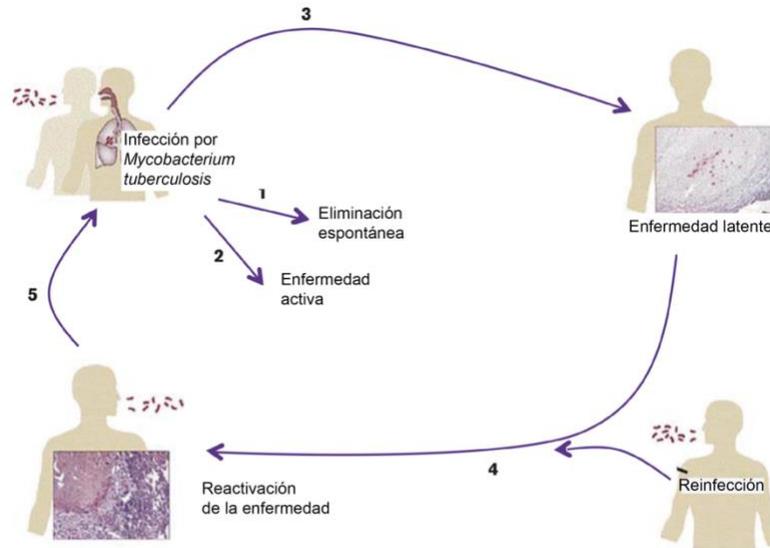


Figura 8. Ciclo de infección de Mtb. El contagio se da por inhalación de aerosoles y pueden ocurrir varios escenarios: (1) eliminación inmediata de la bacteria por el sistema inmune, (2) progresión a una infección activa de TB, (3) la infección no progresa a la enfermedad, sino que es contenida en el hospedero y entra en fase de latencia, (4) en algún momento de la vida del individuo, la TB en latencia puede progresar a la enfermedad o el individuo puede re-infectarse (5). En TB activa se puede dar la diseminación y la transmisión. Imagen modificada de Bañuls A. et al. Journal of Medical Microbiology, 2015

La tuberculosis es curable y prevenible. De las personas con tuberculosis activa cerca del 85% son curadas a través de un tratamiento que implica la utilización de cuatro fármacos llamados de primera línea: isoniazida (INH), rifampicina (RIF), etambutol (EMB) y pirazinamida (PZA) administrados de forma diaria durante 6 meses. Sin este tratamiento la mortalidad aumenta significativamente (Geneva: World Health Organization, 2021).

1.4. Situación epidemiológica actual de la tuberculosis

Actualmente la TB sigue siendo un problema de salud importante a nivel mundial. La distribución de la incidencia a nivel mundial varía ampliamente entre países, observándose una muy alta tasa de incidencia en casi todo el continente africano y el sureste asiático (Figura 9). Según la Organización Mundial de la Salud (OMS), en 2020 hubo 57 países con un promedio estimado de menos de 10 casos cada 100.000 habitantes, lo que los colocaría en una posición cercana a la eliminación de la enfermedad.

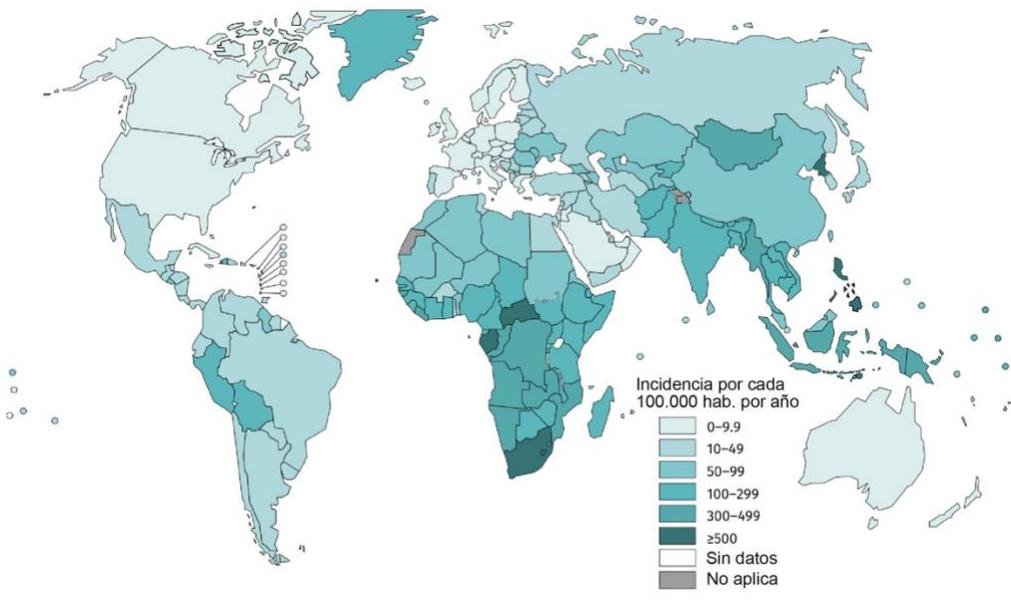


Figura 9. Incidencia estimada de TB en 2020. Casos cada 100.000 habitantes. Imagen tomada del reporte de la OMS de 2021.

Si bien la tendencia en la mortalidad se encontraba en descenso en los últimos 15 años (Figura 10), según la OMS, las muertes por esta enfermedad aumentaron en 2020 y se equipararon a datos del 2017 (1.3 millones de muertes en el mundo). Estos datos resultan todavía más preocupantes si se considera que la principal herramienta diagnóstica se basa en una técnica de microscopía con cerca de 120 años de antigüedad y que hay una sola vacuna de baja eficacia en adultos, desarrollada hace 100 años (la vacuna BCG). Los fármacos, administrados para tratamiento, desde hace más de 40 años, son efectivos, pero tienen una alta tasa de no adherencia debido al tratamiento prolongado y los efectos secundarios (Gagneux, 2018). Como agravante de esta situación, resulta preocupante no solo la cantidad de casos reportados anualmente, sino el aumento sostenido en la incidencia de resistencia a estos fármacos (Liu et al., 2018).

Además, la situación se ve agravada debido a los dos últimos años de pandemia causada por el SARS-CoV-2, revirtiéndose varios años de esfuerzos para controlar la TB. Uno de los impactos más importantes de esta pandemia, fue el subregistro de nuevos diagnósticos en 2020, que implicó una baja del 18% a nivel mundial (Geneva: World Health Organization, 2021). La reducción del acceso al diagnóstico y al tratamiento de la TB aumentó las muertes por esta enfermedad pasando de un estimado de 1.2 millones en 2019 a 1.3 millones en 2020. También el acceso al tratamiento para la TB resistente fue afectado por el colapso de los sistemas sanitarios, estimándose que 1 de cada 3 pacientes interrumpió el tratamiento.

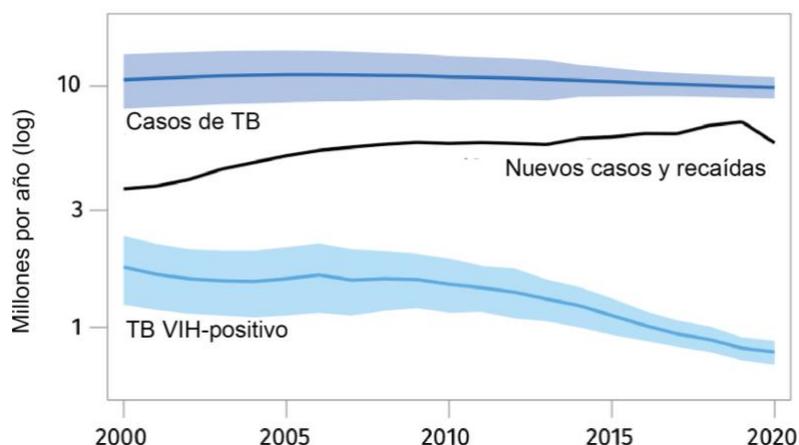


Figura 10. Incidencia de la TB a nivel mundial. Las áreas coloreadas representan los intervalos de incertidumbre. Imagen modificada del reporte anual de la OMS.

Desde 2014, la OMS ha establecido lineamientos para la estrategia del fin de la tuberculosis a nivel mundial. Para ello es necesario alcanzar ciertos hitos y objetivos para frenar la incidencia de la TB y las muertes por esta enfermedad. Para el año 2025 se debería reducir en un 75% las muertes y un 50% la incidencia respecto al 2015, para cumplir con el objetivo de alcanzar en 2035 un 95% de reducción en las muertes y un 90% en la incidencia. Los tres pilares en que se sostiene la estrategia son: (1) cuidado y prevención centrados en el paciente, diagnósticos más tempranos y testeo de la susceptibilidad a drogas; (2) compromiso político y adecuación de recursos para la prevención y el cuidado de la TB; (3) intensificar la investigación y la innovación para desarrollar nuevas herramientas e intervenciones.

1.5. Tratamiento contra la tuberculosis

El tratamiento de la TB requiere que sea prolongado y continuado en el tiempo para eliminar completamente los bacilos y evitar recaídas. También, resulta de gran importancia la combinación de varias drogas para asegurarse que aquellos bacilos que sean naturalmente resistentes a alguna, puedan ser eliminados. Los fármacos antituberculosos clásicamente se clasificaban en dos grupos. Los fármacos de primera línea: INH, RIF, EMB y PZA, y los de segunda línea, utilizados en recaídas o re-tratamientos, principalmente en TB resistentes: fluoroquinolonas como levofloxacin, moxifloxacin, ofloxacin y drogas inyectables como estreptomycin, amikacin, kanamicin y capreomicin. Actualmente la clasificación de los fármacos se realiza de acuerdo a 5 grupos que se dividen según un criterio de efectividad, accesibilidad y costos de cada fármaco (Tabla 1).

Tabla 1. Clasificación de fármacos antituberculosos

Grupo 1	<i>Fármacos de primera línea.</i> Isoniazida, rifampicina, etambutol, pirazinamida
Grupo 2	<i>Fluoroquinolonas</i> Levofloxacin, moxifloxacin, ofloxacin
Grupo 3	<i>Inyectables</i> Estreptomycin, amikacin, kanamicin, capreomicin
Grupo 4	<i>Drogas de segunda línea de menor efectividad.</i> Cicloserina, etionamida/protionamida.
Grupo 5	<i>Drogas de efectividad menor o escasa experiencia clínica.</i> Clofazimina, linezolid, isoniacida a altas dosis, amoxiciclina-clavulánico, carbapenémicos, claritromicina

La INH es una pro-droga que difunde dentro de la bacteria y que es activada por la enzima catalasa-peroxidasa bacteriana codificada por el gen *katG*. El mecanismo de acción de INH, implica el bloqueo de la proteína transportadora *inhA* que inhibe específicamente la síntesis de ácidos micólicos (Jena et al., 2014). La resistencia a esta droga suele ocurrir debido a mutaciones puntuales y altamente selectivas en el gen *katG*, pero también se observan resistencias con mutaciones en el gen *inhA*.

Por otra parte, la RIF es un bactericida que interfiere en la síntesis del ARN mensajero al unirse y bloquear la ARN polimerasa. Esta última es codificada por los genes *rpoA*, *rpoB*, *rpoC* y *rpoD*. Mutaciones en una región de 72 nucleótidos de la subunidad beta (codificada por el gen *rpoB*), da como resultado resistencia a rifampicina. En particular las mutaciones en este gen se concentran en una región altamente localizada entre los codones 507 a 533 (Fontalvo Rivera & Gómez Camargo, 2014).

El tratamiento aconsejado para enfermos con TB que no han sido tratados previamente con drogas antituberculosas, recomienda la utilización de 4 drogas: isoniazida, rifampicina, pirazinamida y etambutol por 2 meses seguido de IZH + RIF por dos meses más (4 meses en total) (CHLA-EP, 2016). Este esquema es supervisado y se realizan baciloscopías para evaluar la negativización del paciente.

Para los casos de recaídas, recuperación de enfermos que perdieron seguimiento o fracasos de tratamiento, se vuelve a iniciar el esquema anterior además de realizar test de sensibilidad a drogas. Cabe mencionar que dada la baja prevalencia de multi-resistencia (resistencia a INH y RIF) de Uruguay, se debe esperar por el test de sensibilidad y un estudio profundo de la historia clínica y bacteriológica antes de volver a iniciar el tratamiento. En países con

prevalencias mayores, las consideraciones frente a casos de recaídas o fracasos del tratamiento son otras.

1.6. Tuberculosis resistente

La resistencia a fármacos contra la TB es una crisis sanitaria a nivel mundial. En el año 2019 se reportaron 600.000 casos resistentes a rifampicina y/o isoniazida en todo el mundo. Las cepas resistentes se pueden clasificar como cepas multidrogo-resistentes (MDR-TB) cuando presentan resistencia a rifampicina e isoniazida, que son los fármacos más efectivos contra la enfermedad, o de resistencia extendida (XDR-TB) cuando, además, presentan resistencia a fármacos de segunda línea como las fluoroquinolonas, amikacina, capreomicina o kanamicina, etc. En el año 2020 se reportaron 25.681 casos de XDR-TB en el mundo, manteniendo una tendencia al alza en la última década. La resistencia tiene una distribución geográfica desigual, siendo mayor en aquellos países de bajos recursos económicos. La multidrogo-resistencia es una enfermedad que surge principalmente del mal uso de los fármacos anti-tuberculosis y es sabido que es más probable su transmisión entre hospederos que la aparición espontánea de nuevos casos (Rajendran et al., 2020).

Las tecnologías de secuenciado masivo han permitido conocer de manera muy completa las secuencias de las diferentes cepas de Mtb. En este sentido, se han descrito marcadores genotípicos de resistencia y se encuentran reportados los niveles de resistencia a determinados fármacos asociados a mutaciones puntuales. Dichos niveles se expresan como concentraciones mínimas inhibitorias (MIC: minimum inhibitory concentrations) y dan intervalos de confianza para la asociación de mutaciones y resistencia fenotípica. En la tabla 2 se resume a modo de ejemplo, algunas de las mutaciones con mayor grado de confianza para la asociación de resistencia fenotípica para INH y RIF. Cuando mayor el grado de confianza, más probable es que la mutación en un gen de resistencia provoque la resistencia fenotípica. Por ejemplo, una de las mutaciones más frecuentes es un cambio aminoacídico de una serina en la posición 315 por una treonina (S315T) en el gen *katG*, dando lugar a la resistencia a isoniazida, en promedio, en 64% de los casos (Lempens. et al., 2018). También es recurrente la mutación en la posición 15 del promotor del gen *inhA*. En conjunto estas mutaciones explican cerca del 84% de la resistencia fenotípica a la isoniazida. Del mismo modo, mutaciones en la región de los codones 507-533 del gen *rpoB* explican cerca del 96% de resistencia a rifampicina, y constituye un verdadero “hotspot” de resistencia (Dookie et al., 2018).

Tabla 2. Lista de mutaciones y grado de confianza asociada a resistencia fenotípica

<i>Fármaco</i>	<i>Gen</i>	<i>Mutación</i>	<i>Grado de confianza</i>	<i>Evidencia de MIC (aislados clínicos)</i>
INH	katG	S315N	Alto	0,8<MIC<12,8 mg/L ¹
INH	katG + inhA	S315T y c-15t	Alto y medio alto	MIC ≥ 1 mg/L ²
RIF	rpoB	Q513K	Alto	MIC ≥ 8 mg/L ³
RIF	rpoB	D516V	Alto	MIC ≥ 8 mg/L ³

¹ (Lempens et al., 2018), ² (Palomino & Martin, 2014), ³ (Berrada et al., 2016)

Nuevos fármacos como la bedaquilina, delamanid y linezolid se han empezado a utilizar para tratar los casos de tuberculosis resistente. Sin embargo, el acceso a estos nuevos tratamientos resulta limitado por su costo elevado y una duración de varios meses con varios fármacos que producen efectos secundarios. Por lo que, culminar el tratamiento se dificulta sobre todo en situaciones socioeconómicas desfavorables (Dookie et al., 2018). La búsqueda por nuevos fármacos anti-tuberculosis, así como tratamientos más efectivos y vacunas nuevas se ha intensificado en las últimas décadas en un esfuerzo para controlar la aparición de resistencia. Existen varios fármacos con nuevos inhibidores o moléculas objetivo en distintas fases de ensayo clínico. Hasta el año 2021 existían 25 drogas en distintas fases, de las cuales 16 son entidades químicas nuevas. También, hasta el mismo año existían 14 vacunas candidatas con efecto profiláctico para la TB en ensayos clínicos, de las cuales cuatro se encuentran entrando en fase III (Weerasuriya et al., 2020). Además, hubo avances en cuanto a las terapias dirigidas al hospedero (HDT: host-directed therapies), que buscan aumentar la respuesta inmune y reducir o modular el proceso de inflamación, para así, de esta forma complementar la terapia con fármacos anti-tuberculosos (Tiberi et al., 2018).

1.7. Situación epidemiológica de la tuberculosis en Uruguay

Desde la década del 80 hasta fines de la década del 90 la TB en Uruguay se encontraba con la posibilidad de ingresar en una situación cercana a su eliminación. Esto fue debido a la puesta en marcha del Programa Nacional de Tuberculosis de la Comisión Honoraria para la Lucha Antituberculosa y Enfermedades Prevalentes (CHLA-EP) que tiene como cometidos la prevención, diagnóstico, tratamiento y vigilancia epidemiológica de la TB en nuestro país. Posteriormente, a partir del 2000, las cifras de incidencia se estabilizaron y a partir de 2006 la situación epidemiológica comenzó a cambiar nuevamente. En el período 2006-2021 hubo un aumento lento, pero progresivo de los casos de TB en el país.

Según el informe nacional de la situación de la tuberculosis en Uruguay 2020, que realiza la CHLA-EP, Uruguay reportó 27,4 casos en 100.000 habitantes. Esto coloca a nuestro país por

encima del promedio en el continente americano, que en los últimos 20 años ve una tendencia descendente. La nueva situación epidemiológica de la TB, coloca a Uruguay dentro de los países de mediana prevalencia. Por otra parte, es importante mencionar que debido a la pandemia de COVID-19, podría estar ocurriendo un subregistro de los casos y abandono de tratamiento y consultas médicas.

Dentro del país, los departamentos que reportan mayor número de casos son Montevideo, Canelones y Maldonado (Figura 11). Se estima que aproximadamente el 87% de los casos corresponden a tuberculosis pulmonares, afectando mayoritariamente a hombres adultos mayores de 45 años. En 2020 la letalidad se ubicó en un 11% de los casos. Los grupos de riesgo asociados con mayores tasas de TB se encuentran bien identificados y se destacan los contactos con convivientes (2.474 casos/100.000 personas), personas con VIH (813/100.000) y personas privadas de libertad (639/100.000). La población de privados de libertad (PPL) es un grupo de especial interés, ya que las condiciones de hacinamiento, poca ventilación y condiciones de infraestructura inadecuadas, así como la baja adherencia a los tratamientos pueden complejizar la situación de la TB cuando estas personas retornan a la comunidad. Se ha reportado en otros estudios una asociación positiva entre el aumento de la población de PPL y la incidencia de la TB, así como los casos MDR-TB. Estas instituciones podrían comportarse como un amplificador de la enfermedad, aumentando el riesgo de diseminación en la comunidad (Warren et al., 2018).

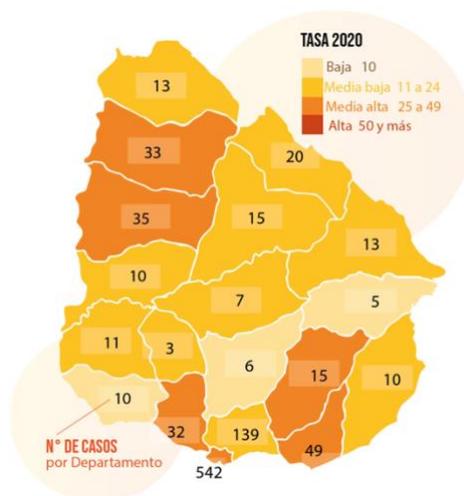


Figura 11. Mapa de la incidencia de TB en Uruguay. Imagen tomada del Informe Nacional de la Situación de la Tuberculosis en Uruguay, CHLA-EP 2020.

En cuanto a la resistencia, cabe mencionar, que en el año 2020 se reportaron 6 casos de mono-resistencia a INH, que se encuentra dentro del rango de los últimos 10 años, y 12 casos de

resistencia a RIF, siendo el triple que lo reportado en 2019. Dos de esos casos fueron diagnosticados como MDR-TB presentando resistencia a ambos fármacos.

Siguiendo con los lineamientos establecidos por la OMS en la estrategia del fin de la TB, se proyectaba un descenso del 20% de los casos para 2020, por el contrario, hubo un aumento del 10%. Esta situación tampoco fue inadvertida dentro del sistema de salud que ha observado un aumento en la cantidad y severidad de los casos que ingresan a las Unidades de Cuidados Intensivos (UCI) especializadas en la tuberculosis en nuestro país. A pesar de ello, algunos valores como la cobertura del tratamiento, la tasa de éxito, cobertura de vacuna para menores de 5 años y seguimiento de contactos permanecen muy altas y cerca de la meta establecida para 2025.

1.8. Tuberculosis en la Unidad de Cuidados Intensivos

En línea con la tendencia observada en el país, en los últimos años se evidenció una mayor cantidad de casos de tuberculosis grave que ingresaban a la UCI del Hospital Español – ASSE (centro de referencia de estos casos). La tuberculosis en cuidados intensivos es asociada con fallas respiratorias y multiorgánicas que requieren medidas de cuidado avanzadas y en varios casos ventilación mecánica invasiva. En Uruguay la mortalidad por TB entre los pacientes admitidos en la UCI es elevada y se ubica por arriba del 50% (Lanoix et al., 2014), (Muthu et al., 2018).

Para el control de la tuberculosis resulta de importancia conocer cuáles son los linajes circulantes en la población ya que las distintas familias inciden en el desarrollo clínico de la enfermedad (Hernández-Pando et al., 2003), en la respuesta a la vacuna y en la predisposición al surgimiento de resistencia a fármacos (Ford et al., 2013), (Kendall et al., 2015). En este mismo sentido, se conformó un grupo interinstitucional entre el Hospital Español, el Institut Pasteur de Montevideo y la CHLA-EP para evaluar la hipótesis de que el aumento de casos graves de TB estaba asociada a un linaje específico más virulento (Hurtado et al., 2021).

Dicho trabajo fue propuesto y financiado por la ANII al Fondo Sectorial de Salud 2018, y se genotiparon 60 aislados obtenidos de pacientes críticos en comparación con 95 aislados de la comunidad. La caracterización de estos aislados se realizó a través de las técnicas de spoligotyping y MIRU-VNTR. En dicho trabajo observamos una mayor presencia de la cepa Haarlem (53,3%) en comparación con las cepas circulantes en la comunidad (34,7%), lo que podría ser indicativo de una mayor virulencia de esta cepa en particular. Sin embargo, no fue posible descartar el aporte de las comorbilidades de los pacientes graves en la mortalidad observada.

1.9. Marcadores moleculares en el hospedero

El mayor acceso a tecnologías de secuenciación, también permitió el acercamiento al estudio del ARN y ADN extracelular, que circula principalmente en vesículas extracelulares o formando asociación con lipoproteínas por el torrente sanguíneo (Tosar et al., 2015). Las vesículas extracelulares son formadas por diferentes tipos de células, especialmente bajo estrés por daño y son importantes en la comunicación celular a distancia. Asimismo, la variación en la concentración de los ARNs o ADNs circulantes en el organismo (tanto dentro de las diferentes vesículas extracelulares o formando parte de complejos con proteínas), es reflejo de estados metabólicos o procesos de enfermedad particulares. Detectar estos cambios resulta de gran importancia para la búsqueda de marcadores moleculares de enfermedad, sobre todo en muestras biológicas de fácil acceso como sangre, suero u orina.

Los microARNs (miARN) son moléculas pequeñas de ARN no codificante, de entre 18 a 24 nucleótidos de largo, que tienen un rol muy importante en la regulación post-traduccional de proteínas, pudiendo afectar diferentes procesos, como por ejemplo la respuesta inmune. Los miARNs son transcritos principalmente por la ARN polimerasa II dando lugar a un precursor primario: pri-miARN. Este precursor es procesado por una proteína nuclear llamada Drosha, dando lugar al pre-miARN, que luego es exportado al citoplasma para nuevamente ser procesado por una proteína citoplasmática: Dicer (Figura 12).

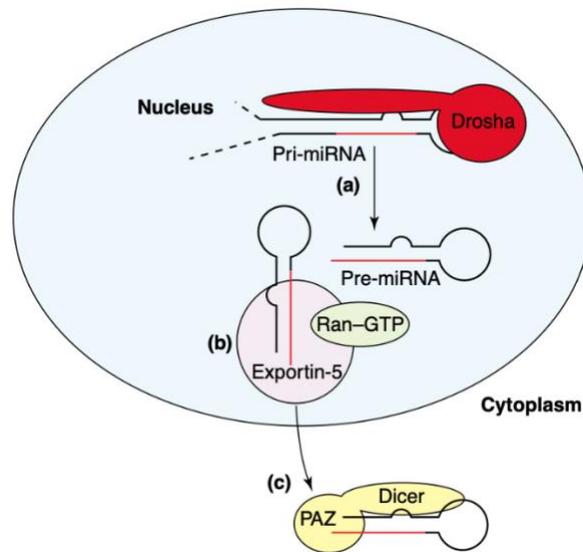


Figura 12. Modelo de la biogénesis de los miARNs. (a) los pri-miARN son transcritos en el núcleo y son procesados por Drosha dando lugar a los pre-miARN. (b) esta estructura es reconocida por una proteína de exportación del núcleo que la libera al citoplasma. (c) el precursor es cortado por Dicer que probablemente utilice el dominio PAZ que reconoce específicamente dos nucleótidos del extremo 3' y se une. Dicer corta al precursor formando el miARN maduro de ~22 nucleótidos. Imagen tomada de Murchison E.P., et al. Curr. Opin. Cell. Biol. 2004

El dúplex de ARN pequeño resultante es unido a la proteína Argonauta, que retiene una sola de las hebras y en conjunto con otros cofactores forman el complejo efector, llamado complejo de silenciamiento inducido por ARN (RISC: RNA-induced silencing complex) (Syeda et al., 2020). Es este complejo el que efectúa el rol regulatorio al unirse por complementariedad al extremo 3'UTR (UTR: untranslated región) de ARN mensajero (ARNm) objetivo reprimiendo su traducción o degradándolo. Cada miARN puede suprimir varios ARNms y un ARNm puede ser regulado por varios miARNs (Sinigaglia et al., 2020).

La expresión de estos miARN responde al contexto celular. La caracterización de los miARN en contexto de enfermedad se ha utilizado para la búsqueda de firmas moleculares características de esas patologías, como importantes herramientas diagnósticas y de pronóstico. En particular para el caso de la TB, existe bibliografía de la caracterización de miARN circulantes en suero en infección y en la enfermedad, con el objetivo de realizar diagnóstico, ya que resulta en un método más rápido que el cultivo bacteriano, con el potencial de suplantar el ensayo de tuberculina o la prueba de interferón gamma, para distinguir entre TB activa y TB latente (Walzl et al., 2018). En ese trabajo, los autores realizan un estudio de la expresión de miARNs en suero entre individuos con tuberculosis activa, individuos con tuberculosis latente y entre controles sanos vacunados y no vacunados con BCG, observando que los distintos grupos tienen perfiles de abundancia de miARNs diferentes y logran validar el hsa-miR-196b y hsa-miR-376c como potenciales marcadores de TB activa (Zhang et al., 2014)

Las ciencias ómicas enfocadas en la clínica de la TB estudian marcadores o firmas moleculares en fluidos o extractos biológicos, como formas de llegar a un diagnóstico precoz y más personalizado. En este trabajo realizamos la búsqueda de miARNs en suero de una población de pacientes con TB grave internados en una Unidad de Cuidados Intensivos. Para ello se realizó un análisis de abundancia diferencial contra un grupo control, como forma de identificar posibles marcadores moleculares de enfermedad.

=

2. OBJETIVOS

2.1. *Objetivos generales*

Generar un método de caracterización genotípica de cepas de tuberculosis por medio de secuenciación masiva y buscar marcadores moleculares con valor diagnóstico y pronóstico en pacientes graves con tuberculosis.

2.2. *Objetivos específicos*

Para poder cumplir el objetivo de caracterización de cepas nos planteamos los siguientes objetivos específicos:

1. Búsqueda de regiones génicas capaces de clasificar los diferentes linajes de *M. tuberculosis*
2. Selección de genes asociados a resistencia a fármacos
3. Puesta a punto de una PCR múltiplex para la amplificación de genes seleccionados de linaje y de resistencia.
4. Secuenciación y análisis de polimorfismos de nucleótido simples en las secuencias obtenidas.
5. Caracterizar una población de aislados en el sistema penitenciario y comparación con cepas aisladas de la comunidad.

Respecto a la búsqueda de marcadores moleculares de la enfermedad se plantearon los siguientes objetivos específicos:

1. Obtención de muestras de suero de una población de pacientes graves diagnosticados con tuberculosis pulmonar ingresados en la UCI.
2. Obtención de ARN total y secuenciación de ARNs pequeños circulantes en suero.
3. Análisis bioinformáticos de cuantificación de miRNAs circulantes.
4. Obtención de miRNAs candidatos con valor diagnóstico y pronóstico de la enfermedad.
5. Validación de los miRNAs candidatos mediante RT-qPCR.

3. MATERIALES Y MÉTODOS

3.1. Caracterización de aislados de *Mycobacterium tuberculosis*.

3.1.1. Obtención de muestras

La obtención del ADN genómico se realizó en el marco del proyecto FMV_3_2018_1_148367: “Generación de un sistema económico de tipificación molecular de cepas de *Mycobacterium tuberculosis* y su implementación en cepas aisladas en el país”. El laboratorio de biología molecular de la CHLA-EP, centraliza toda la bacteriología de la tuberculosis en el país y realiza los cultivos en medio Löwestein-Jensen a partir de diferentes muestras biológicas (esputo, aspirado traqueal, lavados broncoalveolares, etc). De los cultivos positivos se extrae el ADN mediante métodos estandarizados con el kit Genolyse (HAIN, Alemania) y se conserva a -20 °C en el Biobanco de la CHLA-EP.

Se obtuvieron 49 ADNs provenientes de aislados del sistema penitenciario que fueron diagnosticados con tuberculosis durante el año 2019 (Tabla 3). El diagnóstico de TB pulmonar se realizó ante la evidencia de un cuadro clínico y radiológico y posteriormente confirmación por cultivo bacteriano.

Tabla 3. ADN bacteriano aislado de personas privadas de libertad. El resultado de la baciloscopía se informa como negativo si no existe la presencia de bacilos ácido-alcohol resistentes (BAAR) en 100 campos. Se considera positivo <20 cuando se observan entre 1 a 9 BAAR/100 campos, positivo + entre 10 y 99 BAAR/100 campos, positivo ++ entre 1 a 10 BAAR/campo y positivo +++ 10 BAAR/campo. En la columna Recaída se indica con un 1 si la persona fue previamente tratada con cura clínica y vuelve a infectarse. Asimismo, en la columna VIH se indica con 1 si la persona fue diagnosticada con esta infección. En algunos casos no se encontró información de recaídas ni diagnóstico de VIH.

ID	Procedencia	F. nac.	Sexo	Departamento	Baciloscopía	Recaída	VIH
1	Cárcel Central	07/05/1994	M	Montevideo	Positivo ++	0	0
2	Punta de Rieles	18/09/1991	M	Montevideo	Positivo ++	-	-
3	Cárcel Central	27/04/1999	M	Paysandú	Positivo +	0	0
4	Cárcel Central	09/06/1987	M	Montevideo	Positivo ++	-	-
5	Cárcel Central	30/11/1984	M	Montevideo	Positivo <20	0	0
6	CP Durazno	21/07/1976	F	Durazno	Positivo <20	0	0
7	Cárcel Central	06/08/1959	M	Montevideo	Positivo +++	0	0
8	Cárcel Central	14/07/1995	M	Montevideo	Positivo ++	0	0
9	Cárcel Central	14/09/1996	M	Montevideo	Positivo +	0	0
10	Cárcel Central	21/05/1973	M	Montevideo	Positivo <20	0	0
11	Cárcel Central	17/10/1965	M	Montevideo	Positivo +	0	0
12	Cárcel Central	14/12/1986	M	Montevideo	Positivo <20	0	0
13	Punta de Rieles	08/12/1992	M	Montevideo	Positivo +++	0	0
14	Santiago Vázquez	15/09/1993	M	Montevideo	Positivo <20	0	0
15	Cárcel Central	25/07/1993	M	Montevideo	Positivo ++	0	0

16	Cárcel Central	21/07/1989	M	Montevideo	Positivo +	-	-
17	Cárcel Central	18/02/1997	M	Montevideo	Positivo <20	0	0
18	Cárcel Central	22/07/1991	M	Montevideo	Positivo +++	0	0
19	Cárcel Central	28/06/1979	M	Montevideo	Positivo <20	0	0
20	Cárcel Central	14/12/1980	M	Montevideo	Positivo <20	0	0
21	Cárcel Central	01/04/1946	M	Montevideo	Positivo <20	1	0
22	Cárcel Central	30/07/1982	M	Montevideo	Positivo +	0	0
23	CNR - Centro	28/07/1994	F	Montevideo	Positivo <20	0	0
24	CNR - Centro	08/10/1971	F	Montevideo	Positivo <20	-	-
25	Punta de Rieles	28/04/1998	M	Montevideo	Positivo +++	0	0
26	Cárcel Central	04/04/1976	M	Canelones	Positivo <20	0	0
27	Cárcel Central	30/01/1973	M	Montevideo	Positivo +++	0	0
28	Santiago Vázquez	01/01/1995	M	Montevideo	Positivo ++	0	0
29	Cárcel Central	24/07/1994	M	Montevideo	Positivo +	0	0
30	Cárcel Central	12/07/1978	M	Montevideo	Positivo <20	0	0
31	Canelones - U	22/01/1987	M	Montevideo	Positivo <20	1	0
32	CNR - Centro	10/07/1988	F	Montevideo	Positivo ++	0	0
33	Canelones - U	27/03/1992	M	Canelones	Positivo +	0	0
34	Punta de Rieles	09/01/1989	M	Montevideo	Positivo <20	-	-
35	Cárcel Central	02/03/1996	M	Montevideo	Positivo +++	0	0
36	Cárcel Central	28/11/1984	M	Montevideo	Positivo +++	0	0
37	Punta de Rieles	20/01/1985	M	Montevideo	Positivo <20	0	0
38	Punta de Rieles	05/04/1981	M	Montevideo	Positivo +++	1	0
39	Cárcel Central	13/12/1990	M		Positivo <20	0	0
40	Cárcel Central	17/06/1983	M	Montevideo	Positivo +++	0	0
41	Canelones - U	29/07/2000	M	Canelones	Positivo <20	0	0
42	Cárcel Central	05/02/1990	M	Montevideo	Positivo <20	0	1
43	Cárcel Central	10/06/1978	M	Montevideo	Positivo <20	0	0
44	Punta de Rieles	01/06/1987	M	Montevideo	Positivo +++	0	0
45	Santiago Vázquez	04/08/1984	M	Montevideo	Positivo +	0	0
46	Santiago Vázquez	01/02/1988	M	Montevideo	Positivo <20	1	0
47	Cárcel Central	19/06/1980	M	Montevideo	Positivo <20	0	0

3.1.2. Selección de genes para tipificación de cepas y búsqueda resistencia

La tipificación de cepas se llevó adelante mediante la búsqueda de polimorfismos de nucleótido simple para clasificar linajes. Estos SNPs se encuentran reportados en la bibliografía y permiten discriminar cepas circulantes conocidas (Coll et al., 2014). En la tabla 4 se describen los polimorfismos asociados a los principales linajes. El diseño de cebadores se realizó de modo de obtener amplicones de entre 200 y 700 nucleótidos que incluyen en la zona central el SNP que se desea identificar.

Tabla 4. Polimorfismos asociados a genes para discriminar linaje. En la tabla se indica el linaje, el nombre o denominación que recibe, el gen y la posición de la mutación respecto al genoma de referencia. Además, se indica el cambio de base, el número de codón donde ocurre. Todas son sustituciones silenciosas por lo que no hay cambio de aminoácido.

<i>Linaje</i>	<i>Nombre sublinaje</i>	<i>Gen</i>	<i>Posición*</i>	<i>Cambio alelo</i>	<i>Número de codón</i>	<i>Cambio de codón</i>
Linaje 1	Linaje 1	hemL	615938	G/A	368	GAG/GAA
Linaje 2	Linaje 2	glnH	497491	G/A	270	GAC/GAT
Linaje 3	Linaje 3	drrA	3273107	C/A	298	GCC/GCA
Linaje 4	Linaje 4	lpqQ	931123	T/C	57	TAT/TAC
Linaje 4.1.1	X	ctpH	514245	C/T	359	GTG/GTA
Linaje 4.1.2.1	Haarlem	fcoT	107794	C/T	65	GCC/GCT
Linaje 4.3	LAM	rpoC	764995	C/G	542	GCC/GCG
Linaje 4.4.1.1	S	mycP3	355181	G/A	228	AAG/AAA
Linaje 5	Linaje 5	hisD	1799921	C/A	113	GGC/GGA
Linaje 6	Linaje 6	pykA	1816587	C/G	133	GTC/GTG
Linaje 7	Linaje 7	glmU	1137518	G/A	181	AAC/AAT

*De acuerdo al genoma de referencia NC_000962.3

Para conocer la presencia de mutaciones asociadas a fármacos antituberculosos, se buscaron regiones genéticas ya reportadas, tanto de genes asociados a fármacos de primera como de segunda línea. En la tabla 5, se especifican nueve genes de interés asociados a resistencia. El diseño de cebadores se realizó de manera tal de amplificar toda la región codificante de estos nueve genes.

Tabla 5. Genes asociados a resistencia a fármacos anti-tuberculosos. Se describen los genes más asociados a resistencia fenotípica contra los principales fármacos antituberculosos y la posición de cada gen respecto al genoma de referencia de *M. tuberculosis*.

<i>Resistencia asociada</i>	<i>Gen</i>	<i>Posición*</i>
Kanamicina	eis	2714124-2715332
Ethambutol	embB	4246514-4249810
Fluoroquinolonas	gyrA	7302-9818
Fluoroquinolonas	gyrB	5240-7267
Isoniazida	inhA	1674202-1675011
Isoniazida	katG	2153889-2156111
Pirazinamida	pncA	2288681-2289241
Rifampicina	rpoB	759807-763325
Estreptomina	rrS	1471846-1473382

*De acuerdo al genoma de referencia NC_000962.3

Anteriormente nuestro grupo clasificó mediante MIRU-VNTR y Spoligotyping una población de pacientes con TB internados en UCI, así como casos comunitarios que no requirieron internación. Algunas de estas cepas ya caracterizadas, junto con cepas de referencia proporcionadas por la CHLA-EP fueron utilizadas como pruebas de concepto.

3.1.3. Amplificación en múltiplex

Para el diseño de los cebadores se utilizaron herramientas informáticas disponibles en la web como PrimerPooler (Brown et al., 2017). Posteriormente se verificó la formación de dímeros y heterodímeros tomando en cuenta la posibilidad de realizar varias amplificaciones en una misma reacción, con herramientas online como las que se encuentran en “Integrated DNA Technologies” (www.idt-dna.com). En la tabla 6 se describen las secuencias de los cebadores utilizados para la amplificación de las regiones de caracterización de linaje.

Tabla 6. Cebadores diseñados para pool de linaje. En la tabla se indica el nombre del gen o región, a qué linajes se asocian, la orientación del cebador y su secuencia. También se indican el nombre de cada cebador y el tamaño esperado de la amplificación.

<i>Gen</i>	<i>Linaje</i>	<i>Fwd/ Rev</i>	<i>Secuencia 5'-3'</i>	<i>Nombre cebador</i>	<i>Tamaño amplificado</i>
ctpH	X (4.1.1)	Fwd	GGAAAAGTGCCTCGACTGC	ctpH_fwd_BC	374
		Rev	GTTTCGCGCTGATCCAAG	ctpH_rev_BC	
mycP3	S (4.4.1.1)	Fwd	TGACGGCTTCTCTGGTGTG	mycP3_fwd_BC	400
		Rev	GCCACCGACAACACGTAG	mycP3_rev_BC	
fcoT	Haarlem (4.1.2.1)	Fwd	GGTCCGATCGCAGAGGA	fcoT_fwd_BC	320
		Rev	TTGATCGCCGGGAGTTGTAA	fcoT_rev_BC	
rpoC	LAM (4.3)	Fwd	CGCCGGAAATCATCGTCAAC	rpoC_fwd_BC	648
		Rev	TCGGTGGTCAGGTAGTACAG	rpoC_rev_BC	
hemL	Linaje 1	Fwd	GAAGTGGTTGGTACGGAATC	hemL_fwd_BC	372
		Rev	ATGGCATGAAAGAACGCTGG	hemL_rev_BC	
glnH	Linaje 2	Fwd	GATACAACGTGTTCCAGGTG	glnH_fwd_BC	356
		Rev	GTCTACCTCGACGCCAAC	glnH_rev_BC	
drrA	Linaje 3	Fwd	TGAAGGATCTGGACGCTATC	drrA_fwd_BC	201
		Rev	TCAGAGACTCGGTGGGATC	drrA_rev_BC	
lpqQ	Linaje 4	Fwd	GATCAGTCGTGTGTGCTCC	lpqQ_fwd_BC	269
		Rev	TACGGAACCTGACCAGTCG	lpqQ_rev_BC	
hisD	Linaje 5	Fwd	TCTGGGAGAATGGTCTACGTG	hisD_fwd_BC	435
		Rev	CACCACGTTTCATCACCAC	hisD_rev_BC	
pykA	Linaje 6	Fwd	ACAAGGTCGCCTATGAGC	pykA_fwd_BC	364
		Rev	GAGATCCTCGATGTCTTC	pykA_rev_BC	

glmU	Linaje 7	Fwd	CTGGCATGTACGGTCTGG	glmU_fwd_BC	212
		Rev	GATGATCCCTTCGGCTAC	glmU_rev_BC	
alkA	Bovis	Fwd	GGCCTTCGTCGATACCTG	alkA_fwd_BC	732
		Rev	AATGGAATGTGCGTAGTGAC	alkA_rev_BC	

En la siguiente tabla (Tabla 7) se describe la información para el caso de las regiones amplificadas asociadas con resistencia a fármacos. En este caso, para lograr la amplificación completa de algunos genes, se diseñaron más de un par de cebadores solapantes (por ejemplo, para cubrir la totalidad de la región codificante del gen embB).

Tabla 7. Cebadores diseñados para pool de resistencia. En la tabla se indica el nombre del gen, resistencia asociada a fármacos, la orientación del cebador y su secuencia. También se indican el nombre de cada cebador y el tamaño esperado de la amplificación.

<i>Gen</i>	<i>Resistencia asociada</i>	<i>Fwd/Rev</i>	<i>Secuencia 5'-3'</i>	<i>Nombre cebador</i>	<i>Tamaño amplificado</i>
eis	Kanamicina	Fwd	AAAGCCCCGTCAGCCTAGC	eis_fwd_BC	1280
		Rev	ATCGCGTGATCCTTTGCC	eis_rev_BC	
embB	Ethambutol	Fwd1	GGATCGGTGGAGCAGTACC	embB_fwd1_BC	1281
		Rev1	GTTGTTGAACGGCATCCAC	embB_rev1_BC	
		Fwd2	TACCGCCGCATTCACACT	embB_fwd2_BC	1080
		Rev2	GGCTGGTTGGGTTTCATCACG	embB_rev2_BC	
		Fwd3	GAGTCCTGGCATCAATGGTT	embB_fwd3_BC	978
		Rev3	CCAACACCCCTGCGCCGAC	embB_rev3_BC	
gyrA	Fluoroquinolonas	Fwd	AAACGAGGAATAGATGACAG	gyrA_fwd_BC	2588
		Rev	CACCCCGACTCCTAACACTC	gyrA_rev_BC	
gyrB	Fluoroquinolonas	Fwd	GGGTAAAAACGAGGCCAGA	gyrB_fwd_BC	2114
		Rev	CAGGGTTGCGTTAGACATCC	gyrB_rev_BC	
inhA	Isoniazida	Fwd	GGCCACTGACACAACACAAG	inhA_fwd_BC	825
		Rev	CAGCAGGACGGCATCAAATTG	inhA_rev_BC	
katG	Isoniazida	Fwd1	CGATCAACCCGAATCAGC	katG_fwd1_BC	1141
		Rev1	GGTGGATCCGATCTATGAGC	katG_rev1_BC	
		Fwd2	GTCAGTGGCCAGCATCGTCG	katG_fwd2_BC	1044
		Rev2	GATCACAGCCCGATAACACC	katG_rev2_BC	
pncA	Pirazinamida	Fwd	AACAGTTCATCCCGTTTCG	pncA_fwd_BC	575
		Rev	GTCGCCCCGAACGTATGGTG	pncA_rev_BC	
rpoB	Rifampicina	Fwd1	GATTCCCGCCAGAGCAAAC	rpoB_fwd1_BC	1281
		Rev1	TAGTCCACCTCAGACGAGGG	Z2R	
		Fwd2	CCCTCGTCTGAGGTGGACTA	Z3F	906
		Rev2	GTCGAGCACGTAACCTCCCTT	rpoB_rev2_BC	

rrS	Estreptomicina	Fwd	TCCAAAGGGAGTGTGTTGGGT	rrs_fwd_BC	1594
		Rev	CAGTTGGGGCGTTTTTCGTG	rrs_rev_BC	

Se realizó el chequeo de todos los amplicones de forma independiente y posteriormente se realizó la puesta a punto por PCR en gradiente, utilizando diferentes enzimas para determinar la mejor mezcla de reacción. Se determinaron todos los parámetros necesarios para la amplificación en pool: enzima, concentración de cebadores, temperatura de fusión y extensión, agregado de cofactores. Finalmente, se puso a punto la concentración de cada cebador para conformar el pool de linaje (pool-L) y el pool de resistencia (pool-R) (Tabla 8). En la tabla 9, se detalla la mezcla de reacción final para cada pool, y las condiciones de ciclado para el pool-L y pool-R en la tabla 10 y 11 respectivamente.

Tabla 8. Generación de stock 10x de los pooles (100 uL finales). A la izquierda se indica el volumen utilizado para cada par de cebadores en la mezcla del pool-L y a la derecha pool-R. El stock de cada cebador es 100 mM.

<i>Genes</i>	<i>Fwd/Rev</i>	<i>Volúmen *</i>	<i>Genes</i>	<i>Fwd/Rev</i>	<i>Volúmen*</i>
<i>Linaje</i>			<i>Resistencia</i>		
ctpH	Fwd	2 uL	gyrA	Fwd	2 uL
	Rev	2 uL		Rev	2 uL
mycP3	Fwd	2 uL	gyrB	Fwd	2 uL
	Rev	2 uL		Rev	2 uL
rpoC	Fwd	1,5 uL	inhA	Fwd	1,5 uL
	Rev	1,5 uL		Rev	1,5 uL
hemL	Fwd	2 uL	embB	Fwd1	2 uL
	Rev	2 uL		Rev1	2 uL
glnH	Fwd	1 uL		Fwd2	2 uL
	Rev	1 uL		Rev2	2 uL
drrA	Fwd	1,5 uL	Fwd3	2 uL	
	Rev	1,5 uL	Rev3	2 uL	
lpqQ	Fwd	1,5 uL	katG	Fwd1	1,5 uL
	Rev	1,5 uL		Rev1	1,5 uL
hisD	Fwd	2 uL		Fwd2	1,5 uL
	Rev	2 uL		Rev2	1,5 uL
pykA	Fwd	1,5 uL	rrS	Fwd	2 uL
	Rev	1,5 uL		Rev	2 uL
glmU	Fwd	2 uL	Eis	Fwd	1,5 uL
	Rev	2 uL		Rev	1,5 uL
alkA	Fwd	1,5 uL	rpoB	Fwd1	1,5 uL
	Rev	1,5 uL		Rev1	1,5 uL
fcoT	Fwd	2 uL		Fwd2	1,5 uL
	Rev	2 uL		Rev2	1,5 uL

pncA	Fwd	1,5 uL			Rev	1,5 uL
------	-----	--------	--	--	-----	--------

*Se agrega H₂O c.s.p 100 uL

Tabla 9. Mezcla de reacción final para pooles de linaje y resistencia.

<i>Mix de reacción pool L y R (20 uL finales)</i>	x1
Phusion U Hot Start Polimerasa 2x (ThermoScientific)	9 uL
Pool de cebadores 10x	2 uL
H ₂ O	c.s.p 20 uL
ADN genómico (dilución 1/10)	2 uL

Tabla 10. Condiciones de ciclado pool-L (30 ciclos de reacción).

Temp. (°C)	98	98	59	72	72
Tiempo (min)	10	10s	30s	50s	10

Tabla 11. Condiciones de ciclado pool-R (30 ciclos de reacción).

Temp. (°C)	98	98	64	72	72
Tiempo (min)	10	10s	30s	1	10

3.1.4. Preparación de bibliotecas para secuenciación.

A partir de los amplicones generados con la PCR, se construyeron las bibliotecas de secuenciación compatibles con la plataforma MiSeq (Illumina, USA). Para cada aislado de Mtb, se realiza una primera amplificación con los cebadores del pool-L por un lado y del pool-R en otro tubo (Figura 13). A partir de la primera ronda de PCR, se realiza un paso de purificación con perlas magnéticas (AMPure XP, Beckman Coulter) con un protocolo modificado adicionando 1,8 volúmenes por cada pool. Posteriormente, se utilizó el kit “Nextera XT DNA library Prep” (Illumina) para la generación de bibliotecas siguiendo las instrucciones del fabricante.

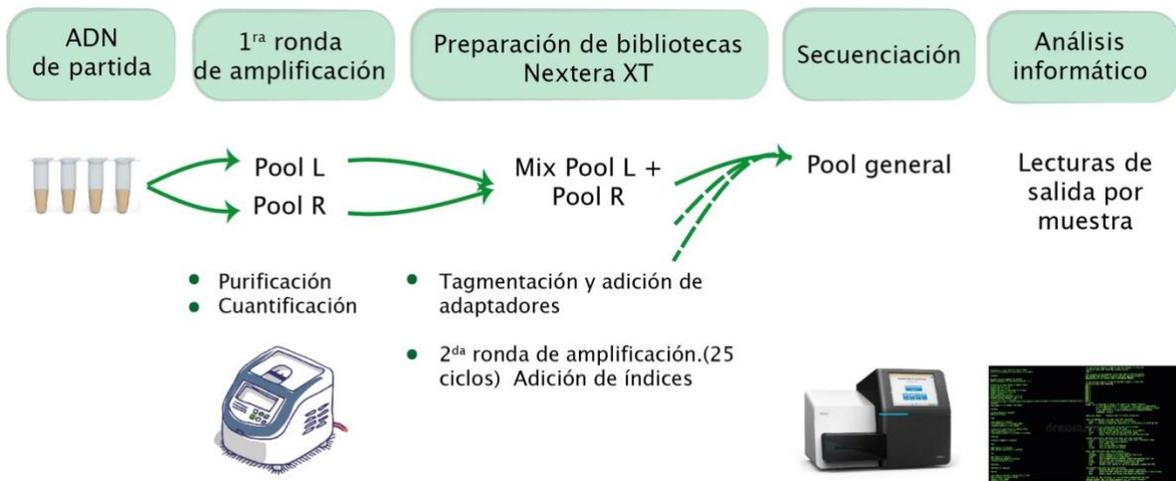


Figura 13. Esquema del flujo de trabajo general. Para cada ADNg de una muestra se amplifican los pools L y R, se purifican y cuantifican. Luego se mezclan los pools de una muestra en proporción 2:1 (L:R). Para la secuenciación en el MiSeq se realiza un pool general de todas las muestras, cada una identificada con una secuencia índice única. Las lecturas que realiza el equipo son procesadas posteriormente.

La preparación de bibliotecas se realizó con el kit Nextera XT DNA (Illumina, USA), partiendo de 1 ng de la mezcla de los pools L y R, donde se adiciona una enzima que fragmenta el ADN y liga las secuencias adaptadoras. Luego mediante una amplificación se agregan secuencias identificadoras únicas para cada muestra (secuencias índices) (Figura 14). Por último, se realiza una purificación y cuantificación y se genera una mezcla final de bibliotecas en cantidades equimolares a secuenciar.

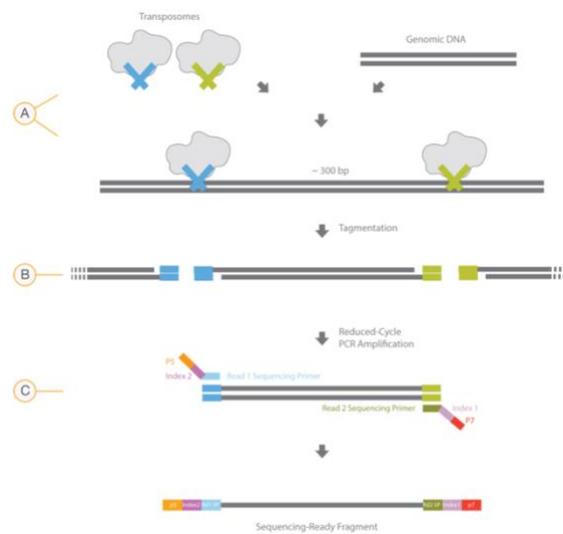


Figura 14. Esquema de generación de bibliotecas con kit Nextera XT DNA library Prep (Illumina). (A) transposasa que corta el ADN y adiciona adaptadores. (B) fragmentación del material de partida (C)

adiciones de secuencias índices únicas por muestra. Imagen tomada del protocolo Nextera XT DNA library Prep, Illumina.

Se utilizó un kit de 2x75 ciclos compatible con el equipo MiSeq de Illumina, disponible en el Institut Pasteur de Montevideo. Se realizó la secuenciación por ambos extremos (Paired-End).

3.1.5. Análisis de los datos

Desarrollamos un esquema de análisis propio, concatenando programas en línea de comandos con el objetivo de identificar SNPs para clasificar linaje y resistencia (Figura 15). En primer lugar, se realizó un chequeo de la calidad de los archivos de salida o lecturas en formato fastQ¹ y se recortaron los extremos con calidades menores a 30 y las secuencias de adaptadores con el programa Sickle (Joshi, 2011).



Figura 15. Esquema de trabajo para el análisis informático de las secuencias obtenidas. Se realiza un control de calidad y recorte de adaptadores y cebadores. Luego se realiza el alineamiento contra todo el genoma. Las lecturas que mapean se procesan con herramientas del paquete samtools para obtener una tabla que compila los SNPs detectados en cada gen.

En segundo lugar, se realizaron dos estrategias para alinear las lecturas contra las bases de datos: se intentó realizar un alineamiento global contra todo el genoma y un alineamiento contra el subset de genes que fueron amplificados. Se utilizaron varias herramientas para los alineamientos como bowtie2 (Langmead & Salzberg, 2012) y bwa-mem (Li, 2013). A partir de los archivos de salida de estos alineadores de secuencia, se utilizó bcftools mpileup para

¹ Archivo de texto conteniendo la secuencia nucleotídica y su correspondiente calidad (Phred score)

generar una tabla de datos con los polimorfismos detectados en cada gen o región comparando contra el genoma de referencia (H37Rv – NC_000962.3) (Figura 15). A modo de chequeo de este flujo de trabajo, también se comparó con programas online disponibles para realizar este análisis como PhyResSE (Feuerriegel et al., 2015), que brindan informes de tipificación y de resistencia a partir de archivos de secuenciación genómica.

3.2. Búsqueda de marcadores moleculares en suero

3.2.1. Población de estudio

Para la búsqueda de marcadores moleculares en el hospedero, se obtuvo suero de una población de pacientes ingresados en la UCI del Hospital Español-ASSE, que fueron diagnosticados en base a un cuadro clínico/radiológico compatible, y con confirmación bacteriológica, para TB (Tabla 12). Dichos sueros fueron recolectados entre 2016 y 2020 en el marco del proyecto ANII FSS_X_2016_1_127692 y cuenta con aval del Comité de Ética del Hospital Español. Se tomaron muestras de sangre en el primer día de ingreso a la UCI, previo a la administración de antibióticos. A partir de la centrifugación de la sangre, se obtuvo el suero. En total, se obtuvieron 32 sueros que fueron almacenados a -80 °C hasta su posterior uso.

Tabla 12. Datos generales y principales características de pacientes con TB grave. En la tabla se indica la edad del paciente, sexo, puntuación en el sistema de clasificación de severidad (APACHE II: Acute Physiology And Chronic Health Evaluation II) y desenlace (V: vive, F: fallece). También se indican los días de ventilación mecánica requeridos, fecha de ingreso en la UCI, total de días dentro de la Unidad y si era positivo para VIH.

<i>ID</i>	<i>Edad</i>	<i>Sexo</i>	<i>APACHE II</i>	<i>Desenlace</i>	<i>V. mec. (días)</i>	<i>Ingreso UCI</i>	<i>Días en UCI</i>	<i>VIH</i>
175-6	34	F	29	F	35	8/8/17	35	0
224-6	29	F	13	V	0	27/5/17	6	1
397-7	59	F	37	F	2	27/2/17	2	0
522-0	33	M	43	F	15	11/8/17	16	1
707-6	38	M	27	F	6	23/11/16	6	0
761-6	32	M	21	F	7	10/5/17	7	0
STB04	51	M	24	F	4	14/4/20	4	0
STB05	32	F	8	V	0	14/2/17	2	1
STB06	42	M	-	V	0	13/6/18	1	0
STB09	48	M	22	F	20	8/11/17	28	0
STB10	44	M	4	V	0	17/4/18	8	0
STB11	37	F	12	V	2	6/12/17	4	0
STB12	56	M	12	V	-	19/1/18	5	1
STB14	27	F	11	V	0	8/10/18	10	0

STB15	78	M	19	F	2	16/9/18	2	0
STB16	48	M	27	F	25	19/5/18	25	0
STB18	47	M	42	F	3	19/2/19	3	1
STB20	57	M	24	V	30	9/3/18	48	0
STB21	38	F	31	F	23	10/1/19	28	1
STB22	62	M	35	F	25	13/3/19	29	0
STB23	34	M	25	F	5	28/5/19	5	0
STB24	29	F	20	F	2	23/12/18	7	0
STB25	33	F	19	V	8	18/7/19	11	0
STB03	26	F	11	F	1	12/7/17	1	1
STB32	71	M	-	F	1	24/11/18	1	1
STB33	23	M	26	V	19	8/9/17	22	0
STB34	41	M	9	V	0	6/10/18	9	0
STB35	24	M	18	V	8	11/9/19	14	0
STB36	40	M	8	F	1	8/6/20	2	0
STB37	29	M	12	V	-	26/7/20	10	0
STB38	49	M	10	V	-	3/7/20	6	0
STB39	27	M	22	F	5	13/2/20	5	0

De los sueros recolectados, hubo 4 casos en los que el cultivo bacteriano resultó negativo, por lo que se descartaron en este análisis. Se destacan varias comorbilidades en común a todos ellos como policonsumo y situación de calle (44%), abandonos de tratamiento (31%), fumadores o ex fumadores intensos (28%), VIH positivo (22%) y privados de libertad o ex privados de libertad (12%).

Como población control se utilizaron sueros normales de donantes anónimos que tenían las siguientes características: entre 18 a 65 años, más de 50 kg de peso, buen estado de salud, sin historial de enfermedades infecciosas y con pruebas negativas para VIH, virus de la hepatitis (tipo B o C), enfermedad de Chagas, así como pruebas bacteriológicas negativas para hanseniasis (lepra), brucelosis, sífilis y tuberculosis.

3.2.2. Extracción de ARN

En primer lugar, se realizó la extracción de ARN total (incluidos los miARNs) a partir de suero, con el kit “Quick-cfRNA Serum and Plasma” (ZYMO Research, USA) siguiendo las instrucciones del fabricante. Para todos los casos se partió de 1 mL de suero. En el caso que la muestra se encontrara con signos de hemólisis, se realizó una centrifugación previa de 15 minutos a 10.000 g a 4°C.

3.2.3. Secuenciación de ARNs pequeños circulantes en suero

Se fabricaron bibliotecas para secuenciar los ARNs circulantes en suero con la plataforma MiSeq (Illumina, USA) utilizando el kit “NEBNext small RNA Library Prep Set for Illumina” (New England BioLabs, USA). El protocolo de trabajo implica en primer lugar, la ligación de adaptadores al extremo 3´-OH de los fragmentos de ARN. En un segundo paso se hibrida un cebador complementario a este adaptador y se coloca en exceso. De esta manera, se forman dímeros de adaptadores de ADN de doble hebra que no son sustrato de la ligasa 1 de ARN utilizada más adelante. El tercer paso se realiza la ligación del adaptador en el extremo 5´-P. A continuación, se sintetiza la primera hebra de ADN copia a partir de los híbridos (ARN-RT Primer). Por último, se realiza una amplificación por PCR con el agregado de secuencias únicas identificadoras para poder incluir diferentes aislados en una misma corrida de secuenciación (Figura 16).

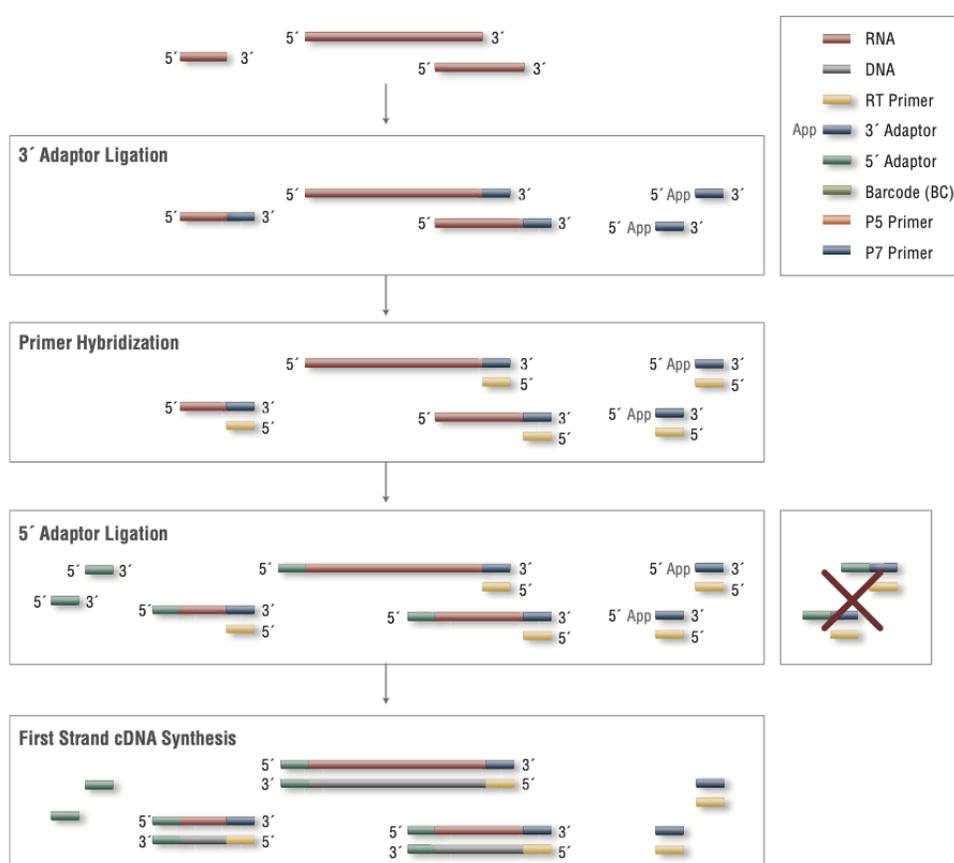


Figura 16. Esquema de generación de bibliotecas de ARNs pequeños. Imagen obtenida del protocolo NEBNext small RNA Library Prep, Illumina (NEB).

Las bibliotecas obtenidas se purificaron con el kit PureLink (Invitrogen, USA) y para seleccionar el tamaño correspondiente a los miARNs se realizó un gel de poliacrilamida al 6% (condiciones de corrida: 120 V, 1 hora). Las bandas correspondientes a ~140 nucleótidos

(tamaño esperado de los miARNs que incorporaron las secuencias adaptadoras) se recortan y purifican (Figura 17). Finalmente se realiza un control de calidad, del tamaño y de la concentración de cada biblioteca utilizando un chip de ADN de alta sensibilidad, en un Bioanalyzer 2100 (Agilent, USA).

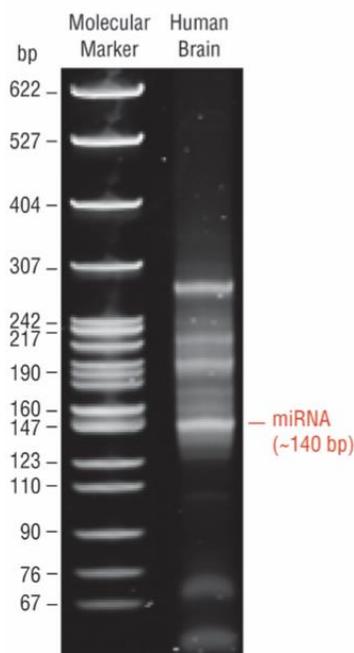


Figura 17. Gel de poliacrilamida indicando un resultado típico de la fabricación de bibliotecas. Las bandas de 140 y 150 pb corresponden a los miARN (21 nt). Imagen obtenida del protocolo NEBNext small RNA Library Prep, Illumina (NEB).

Las bibliotecas se cuantificaron con el kit Qubit “High Sensitivity dsDNA Assay”. En base a la concentración de las bibliotecas obtenidas, se realizó un pool general de 12 nM y se realizó una corrida de 50 ciclos con el cartucho “MiSeq Reagent Kit v3”, en el secuenciador MiSeq disponible en el Instituto.

3.2.4. Análisis de datos

Las secuencias obtenidas se analizaron con el paquete de programas disponible online sRNAtoolbox (<http://bioinfo5.ugr.es/srnatoolbox>). Estas herramientas, permiten analizar los perfiles de expresión o abundancia diferencial de los miARN, visualizar el contexto genómico en función del largo y mapear lecturas sin asignar a partir de los archivos fastq generados durante la secuenciación. Se utilizaron principalmente dos programas de este paquete. En primer lugar, el paquete sRNAbench (Aparicio-Puerta et al., 2019) se utilizó para observar el perfil de abundancia de ARNs pequeños de cada muestra. Este programa reconoce y elimina los adaptadores utilizados en la construcción de bibliotecas y permite

establecer el punto de corte para admitir o no una secuencia en función de su calidad de secuenciado (phred score >30). Luego, se mapean y anotan las secuencias alineándolas contra el genoma de referencia (en este caso genoma humano GRCh38). En segundo lugar, se utilizó el paquete sRNAde que permite la detección de ARNs pequeños diferencialmente abundantes entre grupo de muestras. Toma como punto de partida la información provista por sRNAbench y calcula un mapa de calor indicando el agrupamiento de las muestras según los niveles de abundancia de los ARNs pequeños. Además, se utilizan otros paquetes como DESeq2 (Love et al., 2014) para la normalización y el análisis diferencial, calculando el fold change y la dispersión de estos datos, lo que permite analizar expresión o abundancia diferencial entre grupos. Las tablas finales de resultados indican el p-valor ajustado. En nuestro análisis, un p-valor ajustado <0,05 fue considerado un cambio significativo.

3.2.5. Validación de miARNs

La validación de los miARNs se realizó mediante la amplificación por la técnica de stem-loop RT-qPCR (Chen et al., 2005) y el análisis de las abundancias relativas por el método del $\Delta\Delta C_t$ (Schmittgen & Livak, 2008). La retrotranscripción de los miARN implica la construcción de un cebador con forma de horquilla que contiene en su extremo 3' entre cinco y seis nucleótidos complementarios al miARN de interés (Figura 18).

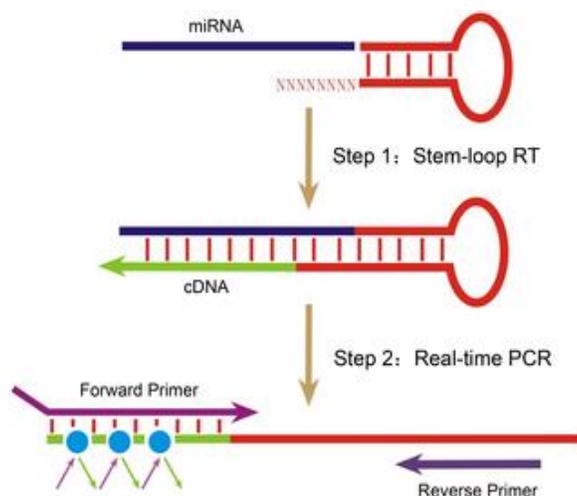


Figura 18. Esquema de amplificación de un miARN por stem-loop. El primer paso implica la retrotranscripción del miARN a través de la hibridación del cebador en forma de horquilla en el extremo 3'. En el segundo paso se realiza la PCR en tiempo real utilizando un cebador forward específico y un cebador reverso universal que hibrida en todos los cebadores SL_RT. Imagen modificad de Chen C. et al. Nucleic Acids Research, 2005.

En base a la información obtenida en el análisis de abundancia diferencial, los miARNs con un p-valor ajustado significativo, que presentaran un número de lecturas alta (o sea de mayor abundancia) se propusieron como candidatos a marcadores de enfermedad o como pronóstico de gravedad (Tabla 13).

Tabla 13. Lista de miARN candidatos y cebadores para retrotranscripción. Se indican el nombre de cada miARNs y la secuencia de cada cebador. En negrita se marcan los nucleótidos complementarios al extremo 3' de cada miARN en los cebadores stem-loop (SL-RT).

<i>miARN</i>	<i>Cebador</i>	<i>Secuencia</i>
hsa-miR-378a-3p		ACUGGACUUGGAGUCAGAAGGC
	SL-RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATA CGACGCCTTC
	Fwd	CCCCACTGGGACTTGGAGTCA
hsa-miR-126-3p		UCGUACCGUGAGUAAUAAUGCG
	SL-RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATA CGACTTACGC
	Fwd	CCCCTCGTACCGTGAGTAAT
hsa-miR-423-5p		UGAGGGGCAGAGAGCGAGACUUU
	SL-RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATA CGACACTGAG
	Fwd	CAAGCTCGGTCTGAGGCC
hsa-let-7b-5p		UGAGGUAGUAGGUUGUGUGGUU
	SL-RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATA CGACAACCAC
	Fwd	GACCTGAGGTAGTAGGTTGTGT
Reverso universal	Rev	GTGCAGGGTCCGAGGT

El primer paso, previo a la retrotranscripción es realizar el plegamiento de la horquilla de cada cebador según las condiciones de la tabla 14.

Tabla 14. Condiciones de temperatura para el plegamiento de cebadores stem-loop.

Temp. (°C)	95	75	68	65	62	60
Tiempo (min)	10	15	15	15	15	30

Las condiciones para la reacción de retrotranscripción se detallan en la tabla 15. Los ADNc obtenidos son almacenados en -20°C hasta su utilización.

Tabla 15. Mix de retrotranscripción, por cada cebador utilizado.

<i>Mix retrotranscripción</i>	x1
First Strand Buffer (5x)	1,5 uL
dNTPs (10 uM)	0,75 uL
DTT (0,1 M)	0,75 uL
RNase out 40 U/uL	0,095 uL
H ₂ O	2,81 uL
SuperScript II Reverse Transcriptase (Invitrogen) 10,000 U	0,09 uL
Cebador SL-RT (200 nm)	0,5 uL
ARN	0,5 uL

Por cada muestra, se realizó una PCR en tiempo real para los ADNc obtenidos en la retrotranscripción. Se eligió como gen normalizador (“housekeeping”) o referencia al microARN let-7b-5p, dado que resultó ser abundante y ser homogéneo entre todas las muestras. Todas las corridas de qPCR se realizaron en duplicado y de acuerdo a las condiciones de la tabla 16 en un QuantStudio 3 (Applied Biosystems, USA). Se puso a punto las concentraciones de cebadores stem-loop, así como los cebadores *forward* y reverso universal, y la cantidad de ADNc utilizado.

Tabla 16. Cálculo del mix de qPCR para una reacción.

<i>Mix qPCR</i>	x1
SensiFast SYBR lo-Rox 2x	6 uL
Cebador fwd (10 uM)	0,4 uL
Cebador reverso universal (10 uM)	0,2 uL
H ₂ O	3,4 uL
ADNc	1 uL

Para comparar la abundancia relativa de los miARNs entre dos muestras diferentes se utilizó la Ecuación 1, donde cada muestra está referida a un control interno en cada grupo.

$$2^{-\Delta\Delta Ct} = \frac{(Ct \text{ miARN de interés} - Ct \text{ miARN referencia})_{\text{grupo TB}}}{(Ct \text{ miARN de interés} - Ct \text{ miARN referencia})_{\text{grupo control}}} \quad (1)$$

4. RESULTADOS

PARTE A - Caracterización de aislados en población de personas privadas de libertad

4.1. Amplificación de pools

En la figura 19 se observan, a modo de ejemplo, los productos de amplificación correspondientes al pool Linaje (pool-L) y pool Resistencia (pool-R) de dos aislados de Mtb. Se detectó la presencia de dímeros de cebadores y se procedió a realizar una purificación para eliminar fragmentos menores a 200 nucleótidos previo al comienzo del protocolo de las bibliotecas.

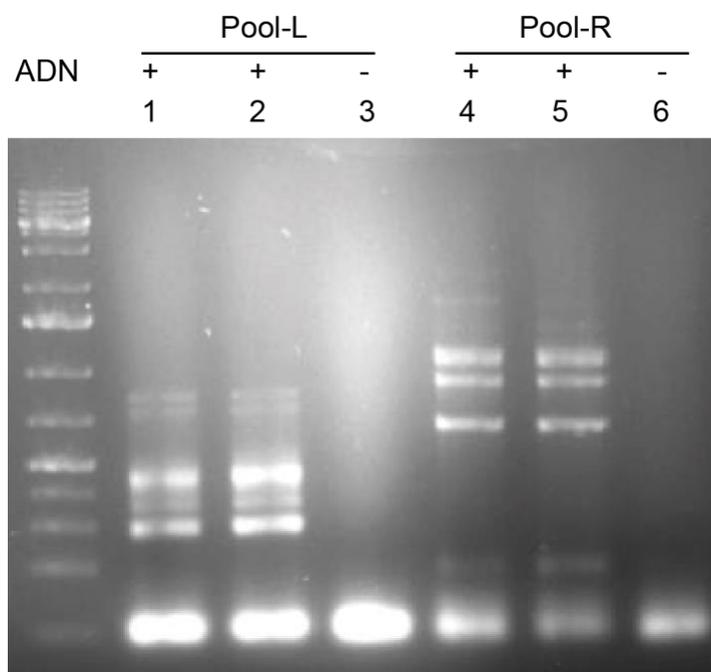


Figura 19. Gel de agarosa para el chequeo de amplificación de pools. A la izquierda se visualiza el marcador de peso molecular de 1 kb plus (Thermo). Los carriles 1 y 2 corresponden a los productos obtenidos de la amplificación del pool-L. Los carriles 4 y 6 corresponden a los productos de amplificación del pool-R. Los controles negativos de amplificación se encuentran en los carriles 3 y 6.

Luego de la construcción de las bibliotecas se realizó un control de calidad a través de un chip de ADN de alta sensibilidad en un Bioanalyzer 2100, según las especificaciones del fabricante. En la figura 20 se pueden distinguir en cada carril, las diferentes bandas correspondientes a las bibliotecas. Se observa una distribución de largos de entre 300 a 2500 pares de bases. La cuantificación de las mismas promedió 120 ng/ μ l.

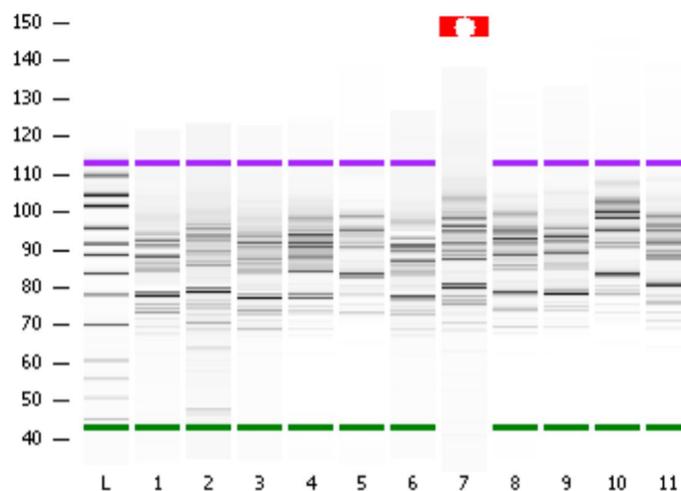


Figura 20. Control de calidad de bibliotecas. Electroferograma de chip de ADN de alta sensibilidad en Bioanalyzer 2100. En el carril L se encuentra el marcador de peso molecular y en los carriles 1 a 11 un subset de bibliotecas. Las bandas observadas se distribuyen entre 300 y 2500 pares de bases. Las bandas magenta y verde corresponden al marcador interno superior (10380 pb) e inferior (35 pb) respectivamente.

Una vez chequeadas las bibliotecas se realizó la secuenciación en el MiSeq en una corrida paired-end de 2x75 ciclos y se obtuvieron las siguientes lecturas (secuencias) (Tabla 17)

Tabla 17. Cantidad de secuencias obtenidas para cada muestra. Se indica además la cobertura calculada como número de secuencias multiplicado por el largo de las lecturas (150 bases) dividido la sumatoria de bases de todas las regiones amplificadas (21500 bases)

ID	Lecturas totales	Cobertura
1	154.744	1.079,6
2	354.744	2.475,0
3	863.052	6.021,3
4	845.600	5.899,5
5	491.620	3.429,9
6	99.596	694,9
7	1.692.386	11.807,3
8	0	0,0
9	936.634	6.534,7
10	779.822	5.440,6
11	353.196	2.464,2
12	228.974	1.597,5
13	328.622	2.292,7
14	1.658.846	11.573,3

15	273.142	1.905,6
16	375.578	2.620,3
17	554.432	3.868,1
18	398.174	2.778,0
19	628.306	4.383,5
20	377.300	2.632,3
21	236.970	1.653,3
22	525.948	3.669,4
23	708.584	4.943,6
24	714.020	4.981,5
25	346.180	2.415,2
26	444.864	3.103,7
27	1.983.186	13.836,2
28	6.455.688	45.039,7
29	448.650	3.130,1
30	316.844	2.210,5

31	822.074	5.735,4
32	866.910	6.048,2
33	335.198	2.338,6
34	325.378	2.270,1
35	264.204	1.843,3
36	357.186	2.492,0
37	219.414	1.530,8
38	3.543.140	24.719,6
39	373.936	2.608,9

40	893.082	6.230,8
41	135.764	947,2
42	546.196	3.810,7
43	309.286	2.157,8
44	350.926	2.448,3
45	555.398	3.874,9
46	201.562	1.406,2
47	271.766	1.896,0

4.2. Análisis primario de secuencias obtenidas

En un análisis inicial, mediante el programa FastQC, se identificaron cada una de las muestras y se controló la calidad de las secuencias obtenidas, el largo esperado (150 nucleótidos), la cantidad de lecturas y otros parámetros como contenido GC, sobrerrepresentación de secuencias o cantidad de secuencias repetidas. En la figura 21 se resumen la calidad de la identificación de cada base, siendo que la mayoría (83%) tiene un nivel de calidad de Phred mayor o igual a 30 (99,9% de precisión). Solamente 8 tuvieron un valor de entre 20 y 30, lo que equivale a una probabilidad de error de 1 en 100.

Cuando se analizó la cantidad de lecturas se observó que un 40% de las muestras obtuvo más de 500.000 secuencias, 51% de ellas estuvo entre 100.000 y 400.000 lecturas y en un caso no se pudo obtener secuencias. Se observa, además, un alto porcentaje de repetición en todas ellas, lo cual es esperable para este set de bibliotecas obtenidas desde productos de amplificación por PCR (Figura 22).

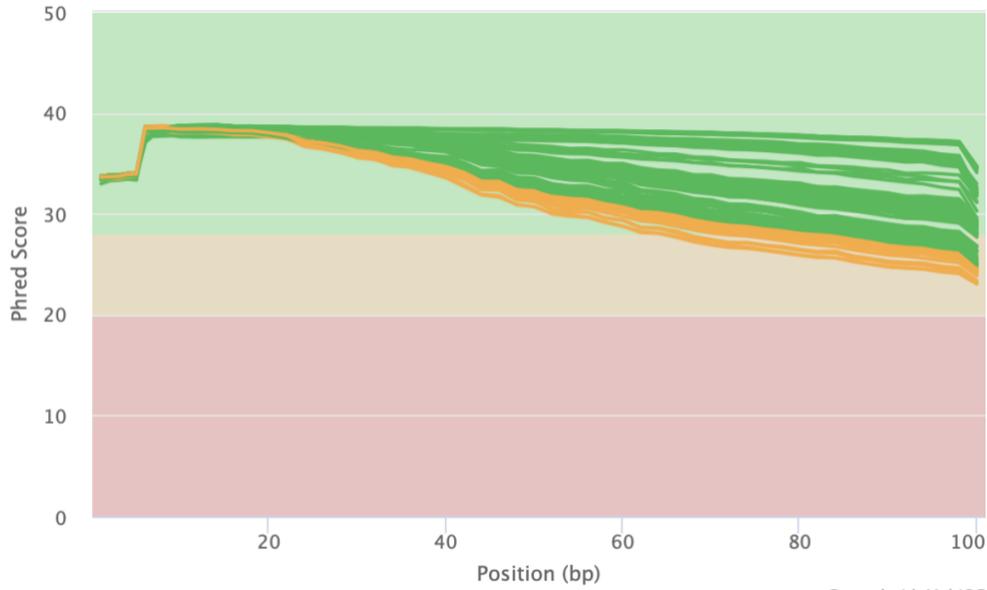


Figura 21. Chequeo de la calidad de las secuencias. Se grafica la calidad de las corridas según la puntuación de Phred que tiene cada posición a lo largo de la lectura. En naranja se señalan set de datos con secuencias que tienen una puntuación de Phred entre 20 y 30.

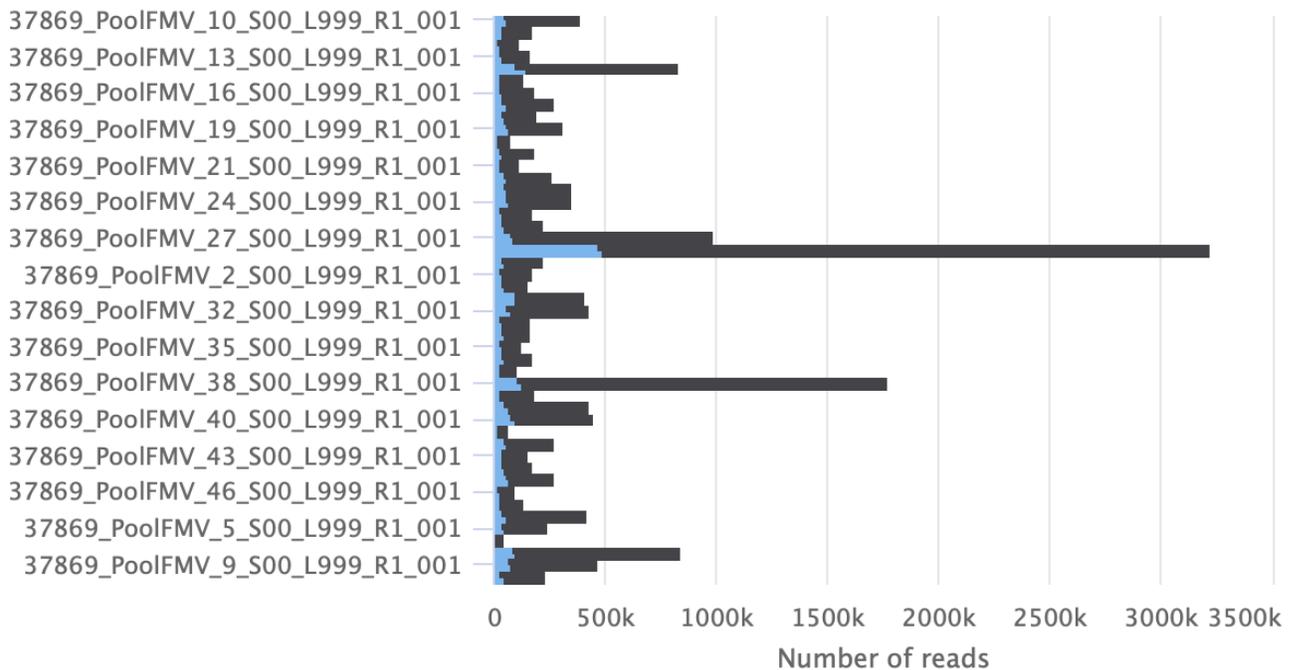


Figura 22. Cantidad de secuencias por muestra. En celeste se identifican las lecturas únicas, mientras que en negro las lecturas duplicadas o repetidas.

4.3. Mapeo de secuencias

Las secuencias obtenidas fueron alineadas mediante los programas bowtie2 y bwa-mem contra el genoma de referencia de *Mycobacterium tuberculosis* H37Rv (NC_000962.3). Para este set de muestras, bwa-mem asignó una mayor proporción de alineamientos y se decidió continuar con este programa.

En la tabla 18 se resume la cantidad de secuencias obtenidas y el porcentaje de mapeo utilizando bwa-mem contra el genoma de referencia. Como puede observarse en la tabla, el número de secuencias que mapean contra el genoma es relativamente bajo: cerca de 2/3 de las lecturas tuvieron porcentajes de mapeo menor al 40%. Esto puede deberse a la baja calidad o degradación del ADN inicial o contaminación en la preparación de las bibliotecas.

Tabla 18. Número de secuencias obtenidas por muestra y porcentaje de alineamiento contra el genoma de referencia *Mycobacterium tuberculosis* H37Rv. En la tabla se indica la cantidad de secuencias obtenidas y cuántas de esas secuencias pudieron ser alineadas con el alineador bwa-mem. También se indica la cobertura corregida en base a las lecturas mapeadas.

<i>ID</i>	<i>Lecturas totales</i>	<i>Nº de lecturas mapeadas</i>	<i>Porcentaje de mapeo (%)</i>	<i>Cobertura corregida</i>
1	154.744	27.573	18	192,4
2	354.744	244.472	69	1.705,6
3	863.052	145.809	17	1.017,3
4	845.600	98.566	12	687,7
5	491.620	94.008	19	655,9
6	99.596	18.103	18	126,3
7	1.692.386	997.928	59	6.962,3
8	0	-	-	-
9	936.634	918.895	98	6.410,9
10	779.822	759.350	97	5.297,8
11	353.196	331.144	94	2.310,3
12	228.974	38.153	17	266,2
13	328.622	69.770	21	486,8
14	1.658.846	188.486	11	1.315,0
15	273.142	170.391	62	1.188,8
16	375.578	55.634	15	388,1
17	554.432	95.050	17	663,1
18	398.174	70.952	18	495,0
19	628.306	225.551	36	1.573,6
20	377.300	70.113	19	489,2
21	236.970	45.743	19	319,1
22	525.948	97.201	18	678,1
23	708.584	661.350	93	4.614,1

24	714.020	161.229	23	1.124,9
25	346.180	43.167	12	301,2
26	444.864	106.131	24	740,4
27	1.983.186	1.935.272	98	13.501,9
28	6.455.688	6.305.933	98	43.994,9
29	448.650	73.687	16	514,1
30	316.844	112.173	35	782,6
31	822.074	708.113	86	4.940,3
32	866.910	130.621	15	911,3
33	335.198	45.176	13	315,2
34	325.378	91.536	28	638,6
35	264.204	50.324	19	351,1
36	357.186	68.926	19	480,9
37	219.414	35.876	16	250,3
38	3.543.140	3.455.085	98	24.105,2
39	373.936	354.176	95	2.471,0
40	893.082	162.082	18	1.130,8
41	135.764	20.716	15	144,5
42	546.196	83.341	15	581,4
43	309.286	52.665	17	367,4
44	350.926	70.462	20	491,6
45	555.398	93.768	17	654,2
46	201.562	24.439	12	170,5
47	271.766	153.368	56	1.070,0

Se analizaron la cantidad de secuencias que fueron mapeadas contra el genoma y se identificaron cuáles de ellas correspondían a los genes de linaje y de resistencia. En general, se obtuvo más cantidad de alineamientos para el pool-R, pese a haber puesto el doble de proporción del pool-L en la generación de las bibliotecas (Figura 23).

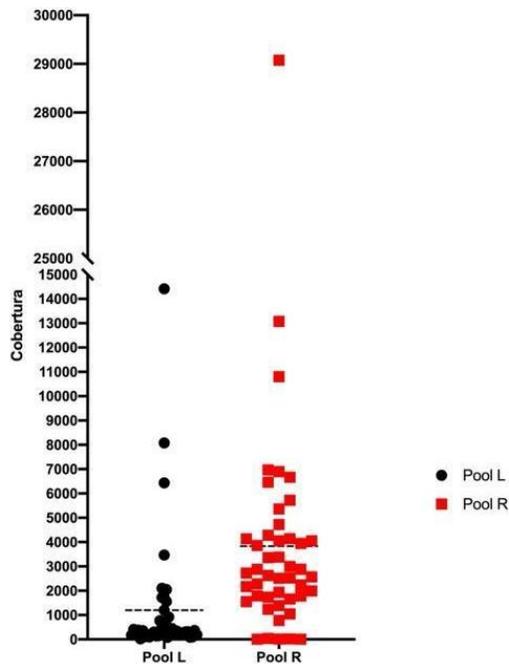


Figura 23. Cobertura de cada gen en el pool-L y en el pool-R. La media se indica con línea punteada.

Las secuencias de cada una de las muestras que mapean contra los genes de linaje y de resistencia se encuentran graficadas en la figura 24 (los datos crudos se encuentran en la Tabla Suplementaria 1 y 2 respectivamente - Anexo I). Como se desprende de los gráficos, existe una distribución heterogénea entre los distintos genes; para *hemL*, *hisD*, *ctpH*, *fcoT*, *gyrB* y *embB*, se obtuvieron muy pocas lecturas. Además, la cobertura fue muy baja para cuatro aislados donde se obtuvieron muy pocas lecturas en el pool-R.

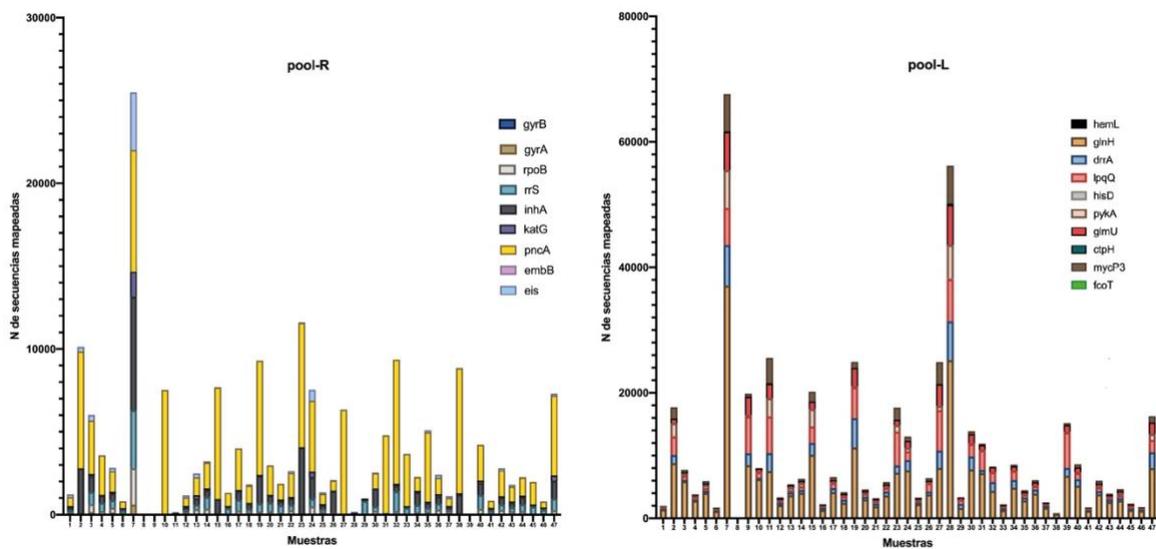


Figura 24. Mapeo de secuencias contra el genoma de referencia. A la izquierda se grafica la cantidad de secuencias por cada muestra que alinearon contra las regiones del pool-R. A la derecha aquellas secuencias que alinearon contra los genes del pool-L.

4.4. Clasificación de linajes

A partir de los archivos de salida del alineamiento (en formato SAM: Sequence Aligment Map), se realizó la visualización del mapeo de las secuencias con el programa IGV (Integrative Genomic Viewer), y se identificaron las variantes en las coordenadas reportadas en la bibliografía (Figura 25). Además, con bcftools se identificaron los polimorfismos respecto al genoma de referencia, y se consideró un SNP si existía un cambio reportado en las posiciones mencionadas y tenía una cobertura de al menos 5x (Tabla 19).

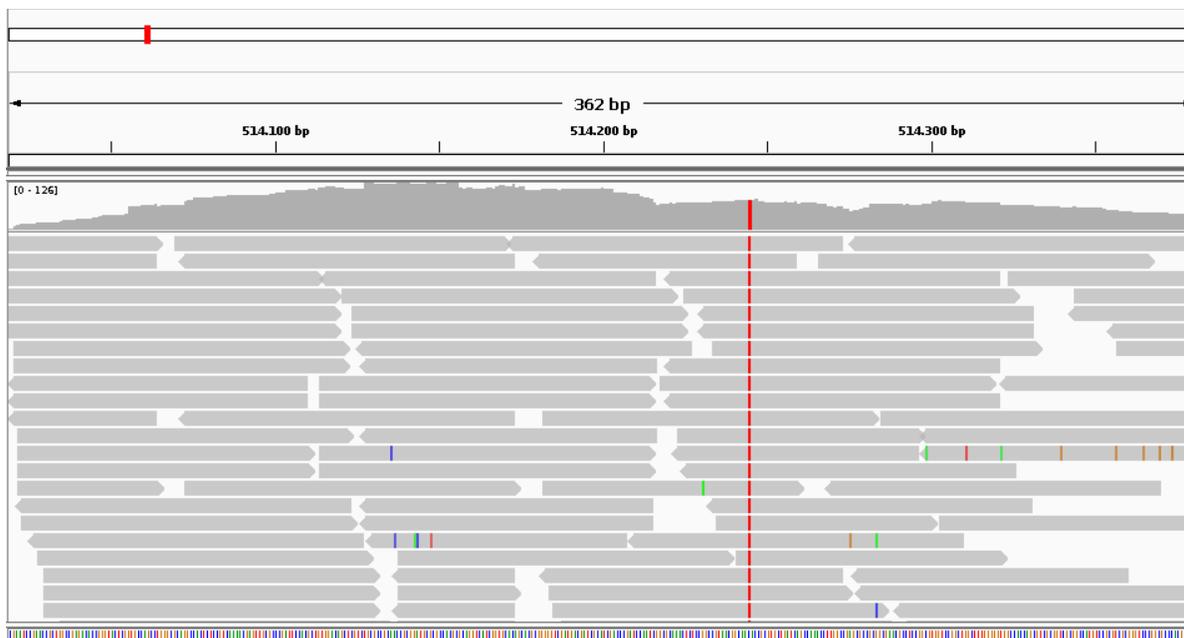


Figura 25. Visualización de los alineamientos con el programa IGV para la identificación del linaje X. En el ejemplo se observa la región correspondiente al gen *ctpH* para una muestra. Se identifican las secuencias alineadas al genoma en gris, la cobertura debajo de la línea de referencia y el cambio de nucleótido en rojo (C/T) identificado en la posición 514245 del genoma de referencia.

Tabla 19. Resumen de los polimorfismos detectados por bcftools (en formato VCF). En este caso se indican los resultados de la muestra 17 y 18 a modo de ejemplo, donde se indica la posición del polimorfismo, la base de la referencia y la alteración detectada, seguido de la puntuación de calidad de Phred-escalada (cuanto mayor es, mayor es la confianza para esa alteración).

	<i>Posición</i>	<i>Ref</i>	<i>Alt</i>	<i>Calidad</i>
17_sorted_aln_genoma.bam				
M. tuberculosis NC_000962.3	7362	G	C	159
M. tuberculosis NC_000962.3	7585	G	C	171
M. tuberculosis NC_000962.3	9304	G	A	201
M. tuberculosis NC_000962.3	760115	C	T	225

18_sorted_aln_genoma.bam				
M. tuberculosis NC_000962.3	7362	G	C	148
M. tuberculosis NC_000962.3	7585	G	C	225
M. tuberculosis NC_000962.3	9304	G	A	183
M. tuberculosis NC_000962.3	2289017	A	G	225

En base a los resultados anteriores, se resumen en la tabla 20 la asignación de linaje para cada una de las muestras. Para los aislados 12, 14, 33 y 34 en los que no hubo secuencias o la cobertura fue muy baja para evidenciar polimorfismos, se clasificaron los linajes mediante MIRU-VNTR.

Tabla 20. Clasificación de las cepas aisladas del sistema penitenciario. Se indican las secuencias en todo el pool-L, la cobertura que tuvo el mismo y la clasificación que surge como resultado de observar la presencia de los polimorfismos en las posiciones de los genes seleccionados.

<i>ID</i>	<i>Secuencias en pool L</i>	<i>Cobertura en pool L</i>	<i>Clasificación de linajes</i>
1	2860	104	LAM
2	47075	1713	LAM
3	8703	317	X
4	5175	188	LAM
5	9578	348	Haarlem
6	2630	96	X
7	222113	8081	X
8	-	-	-
9	57814	2103	Haarlem
10	5148	187	X
11	95368	3470	LAM
12	5575	203	LAM
13	8321	303	Haarlem
14	10959	399	LAM
15	56365	2051	X
16	3944	143	Haarlem
17	11343	413	Haarlem
18	7713	281	X
19	54529	1984	Haarlem
20	7721	281	X
21	5143	187	Haarlem
22	9895	360	S
23	47414	1725	LAM
24	21198	771	X
25	4837	176	LAM
26	10097	367	X
27	176934	6437	-
28	396169	14413	Haarlem
29	4899	178	X
30	25152	915	X
31	15299	557	LAM
32	12192	444	LAM
33	4211	153	LAM
34	14470	526	LAM
35	7458	271	LAM
36	9505	346	Haarlem
37	2261	82	X
38	642	23	X
39	42855	1559	LAM
40	15257	555	X
41	2431	88	-
42	10493	382	Haarlem
43	6420	234	LAM
44	8517	310	LAM
45	4586	167	LAM
46	2722	99	X
47	32983	1200	X

Los resultados obtenidos indican que dentro del sistema penitenciario (SP) se detectaron 39% de cepas del linaje LAM (4.3), 34% del linaje X (4.1.1), 25% Haarlem (4.1.2.1) y 2% S (4.4.1.1) (Figura 26). La distribución de los linajes aislados del SP contrasta con la observada en los aislamientos de la comunidad. En particular, la cepa X se encuentra sobrerrepresentada en el SP (chi-cuadrado con p-valor = 0,0001). Este linaje no fue detectado en ninguno de los aislados de pacientes internados en cuidados intensivos, y es detectado en muy bajo porcentaje en aislados de la comunidad (entorno al 1%) (Hurtado et al., 2021).

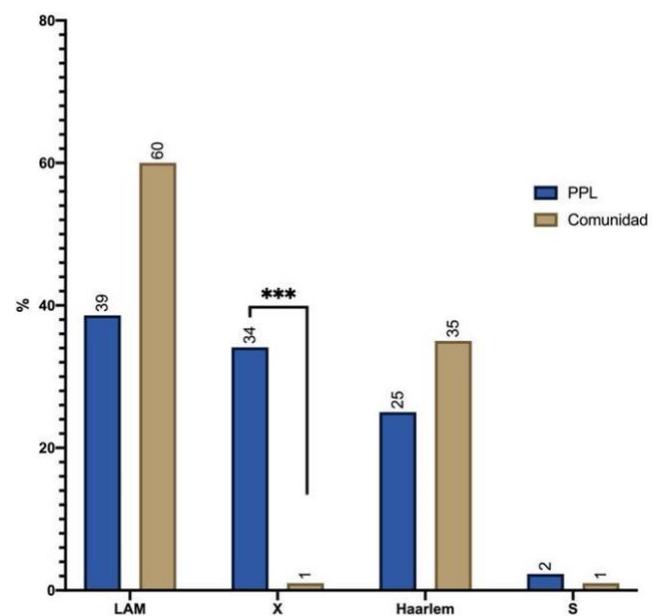


Figura 26. Porcentaje de linajes en población de personas privadas de libertad (azul) vs linajes circulantes en la comunidad (beige).

Cuando se observa el origen de la población de estudio, 28 corresponden a casos aislados de personas que fueron detectados al ingreso del SP en el área metropolitana, 25 de estas personas ya tenían ingresos previos al SP y solamente 3 fueron ingresos sin antecedentes. Para estos últimos, las cepas aisladas fueron identificadas como LAM o Haarlem, que son los linajes que más circulan en la comunidad. Por el contrario, los aislados tipificados como linaje X fueron obtenidos de personas dentro del SP o de personas que re-ingresaban al sistema. Los 19 aislados restantes corresponden a la identificación de casos en base a consultas médicas compatibles con cuadro clínico que ya se encontraban dentro del SP. También el linaje X fue hallado en más de un 30% de estos aislados.

A modo de comparación, se realizó un análisis de tipificación de cepas utilizando el software online PhyResSE (<https://bioinf.fz-borstel.de/mchips/phyresse/>). En dicho programa se cargaron las corridas obtenidas del MiSeq y como resultados se pudieron clasificar 23 cepas.

Las restantes no tuvieron suficientes lecturas para ser procesadas por este programa, aunque fueron asignadas al linaje 4 (sin identificar el sublinaje). Solamente en un caso (muestra 26) la tipificación no coincidió con nuestra metodología propuesta (Tabla 21).

Tabla 21. Comparación de las clasificaciones obtenidas mediante la estrategia desarrollada en este trabajo y mediante el paquete PhyResSE. Se indica la clasificación asignada por cada metodología. En la columna de MIRU-VNTR solamente se tipificaron aquellas que no pudieron ser clasificadas. El recuadro en rojo indica la única discrepancia entre el análisis informático hecho por nuestro grupo y PhyResSE, la tipificación de este aislado por MIRU resultó del linaje Haarlem.

<i>ID</i>	<i>Análisis informático</i>	<i>PhyResse</i>	<i>MIRU-VNTR</i>	<i>ID</i>	<i>Análisis informático</i>	<i>PhyResse</i>	<i>MIRU-VNTR</i>
				24	X		
1	LAM	LAM		25	LAM	LAM	
2	LAM	LAM		26	Haarlem	X	Haarlem
3	X	X		27	-		
4	LAM	LAM		28	Haarlem	Haarlem	
5	Haarlem			29	X		
6	X	X		30	X		
7	X	X		31	LAM	LAM	
8	-			32	LAM	LAM	
9	Haarlem			33			LAM
10	X	X		34			LAM
11	LAM	LAM		35	LAM	LAM	
12			LAM	36	Haarlem		
13	Haarlem			37	X		
14			LAM	38	X	X	
15	X	X		39	LAM	LAM	
16	Haarlem			40	X		
17	Haarlem			41	-		
18	X			42	Haarlem		
19	Haarlem			43	LAM	LAM	
20	X			44	LAM	LAM	
21	Haarlem			45	LAM	LAM	
22	S	S		46	X		
23	LAM	LAM		47	X	X	

4.5. Resistencia a fármacos antituberculosis

Para la búsqueda de mutaciones asociadas a resistencia a fármacos, se utilizó el mismo esquema de trabajo que la parte anterior, acotando la búsqueda a los genes asociados a resistencia a fármacos seleccionados. Las tablas obtenidas para observar las variantes indican la posición del cambio de base y la cantidad de lecturas (Tabla suplementaria 3 – Anexo I). En la tabla 22, se resumen todos los polimorfismos detectados para cada una de las muestras.

Solamente en 4 aislados no se pudo obtener lecturas y no se pudo realizar el análisis para la búsqueda de polimorfismos de resistencia.

Tabla 22. Búsqueda de SNPs en genes de resistencia. Se indica el gen, la posición del cambio de SNP y cuantos aislados presentaron esa sustitución. Además, se indica el cambio de codón (en negrita la base que se modifica), el cambio de aminoácido y si el mismo se encuentra reportado en la literatura asociado a resistencia.

<i>Gen</i>	<i>Posición 1</i>	<i>N.º de muestras</i>	<i>Cambio de SNP</i>	<i>Cambio codón</i>	<i>Cambio de AA²</i>	<i>Bibliografía</i>
gyrA	7362	30	G/C	GAG/CAG	Glu62Gln	No reportado
gyrA	7585	28	G/C	AGC/ACC	Ser95Thr	No reportado
gyrA	8040	4	G/A	GGC/AGC	Gly247Ser	No reportado
gyrA	9304	27	G/A	GGC/GAC	Gly668Asp	No reportado
gyrB	6438	1	C/G	CCC/CGC	Pro400Arg	No reportado
gyrB	6034	1	C/T	AAC/AAT	Asn265Asn	No reportado
katG	2154279	2	G/A	CTC/CTT	Leu611Leu	No reportado
katG	2154015	1	C/G	GGG/GGC	Gly699Gly	No reportado
pncA	2289017	17	A/G	GGT/GGC	Gly75Gly	No reportado
rpoB	760106	2	G/A	TCG/TCA	Ser100Ser	No reportado
rpoB	760115	10	C/T	GAC/GAT	Asp103Asp	No reportado
rpoB	760629	1	C/G	CTG/GTG	Leu275Val	No reportado
rpoB	760852	1	T/C	ATG/ACG	Met349Thr	No reportado
rrS	1472337	3	C/T	ribosomal	c492t	Estreptomina ³
rrS	1473314	4	A/C	ribosomal	a1469c	No reportado

1 Respecto al genoma de referencia *Mycobacterium tuberculosis* (NC_000962.3)

2 Respecto a la posición en el gen

3 (Cabibbe et al., 2020)

La mayoría de las mutaciones producen cambios sinónimos de aminoácidos (57%). En tres aislados se detectó la mutación c1472337t en el gen *rrS* que codifica para la subunidad ribosomal 16S. La mutación está asociada a resistencia a estreptomina, aunque no está caracterizada como SNP de alta confianza. Para el resto de los aislados, en donde hubo un cambio aminoacídico, no se encontró evidencia bibliográfica que indique resistencia fenotípica y es coincidente con la información de estudios de sensibilidad realizados previamente por la CHLA-EP y con la respuesta clínico bacteriológica que tuvieron los PPL.

A modo de corroborar estos resultados, se realizó un análisis con PhyResSE donde se obtuvo una tabla que indica todas las mutaciones observadas por muestra, y si esas mutaciones se encuentran reportadas en la bibliografía. Las secuencias ingresadas al programa contenían una alta tasa de repetición que fueron identificadas como contaminaciones, y en algunos

casos no se pudieron alinear contra la referencia. Del total, se pudo obtener información de 38 aislados. En la tabla 23 se resumen solamente aquellas que tuvieran una mayor cobertura.

Tabla 23. Mutaciones asociadas a resistencia encontradas por el programa PhyresSE. Para un subset de muestras se indica la posición y la mutación respecto a la referencia. También se indica un score de calidad, la región donde ocurre y el cambio de aminoácido y si la variante está asociada a resistencia fenotípica. Todas las mutaciones fueron encontradas en nuestro análisis anterior.

<i>ID</i>	<i>Posición</i>	<i>Ref</i>	<i>Muestra</i>	<i>Calidad</i>	<i>Región</i>	<i>Cambio AA</i>	<i>Variante conocida</i>
32	1472337	C	T	7567,77	Rvnr01	rRNA	estreptomicina
45	1472337	C	T	6624,77	Rvnr01	rRNA	estreptomicina
1	1472337	C	T	2969,77	Rvnr01	rRNA	estreptomicina
1	7585	G	C	8770,77	Rv0006 (<i>gyrA</i>)	Ser95Thr	Genotípica
40	7585	G	C	3550,77	Rv0006 (<i>gyrA</i>)	Ser95Thr	Genotípica
47	7585	G	C	3138,77	Rv0006 (<i>gyrA</i>)	Ser95Thr	Genotípica
28	7585	G	C	2912,77	Rv0006 (<i>gyrA</i>)	Ser95Thr	Genotípica
38	7585	G	C	2660,77	Rv0006 (<i>gyrA</i>)	Ser95Thr	Genotípica
22	7585	G	C	2157,77	Rv0006 (<i>gyrA</i>)	Ser95Thr	Genotípica
13	7585	G	C	2083,77	Rv0006 (<i>gyrA</i>)	Ser95Thr	Genotípica

PhyResse también encontró las mismas mutaciones dentro de las muestras analizadas. Tres muestras presentaron la mutación en el gen *rrS* (posición 1472337) y se les asignó resistencia fenotípica a estreptomicina (Coll et al., 2014). En el resto de los casos se encontró la variante g7585c en el gen *gyrA*, que no está asociado hasta ahora, a resistencia fenotípica.

PARTE B - Búsqueda de marcadores con valor pronóstico de la enfermedad

4.6. ARNs pequeños en suero (miARN)

Se recolectaron un total de 32 sueros de pacientes graves internados en la UCI del Hospital Español durante los años 2018 y 2020. Estos pacientes fueron diagnosticados con tuberculosis pulmonar, y 28 fueron confirmados con cultivo positivo para *Mycobacterium tuberculosis*. Se extrajo ARN total de 26 de ellos, y 13 fueron seleccionados para secuenciación de ARNs pequeños: 175-6, 224-6, 397-7, 522-0, 707-6, 761-1, STB04, STB05, STB06, STB09, STB10, STB11, STB16 y sueros normales 1 al 5 (Tabla 12 en Materiales y Métodos).

4.7. Secuenciación de bibliotecas para miARNs

A modo de control y cuantificación de la fabricación de las bibliotecas se realizó un control de calidad en chip de ARN de alta sensibilidad en un Bioanalyzer 2100, según las especificaciones del fabricante. Se obtuvieron bandas de acuerdo a lo esperado de alrededor de ~140 pares de bases con una concentración promedio de 46 ng/ μ L (Figura 27).

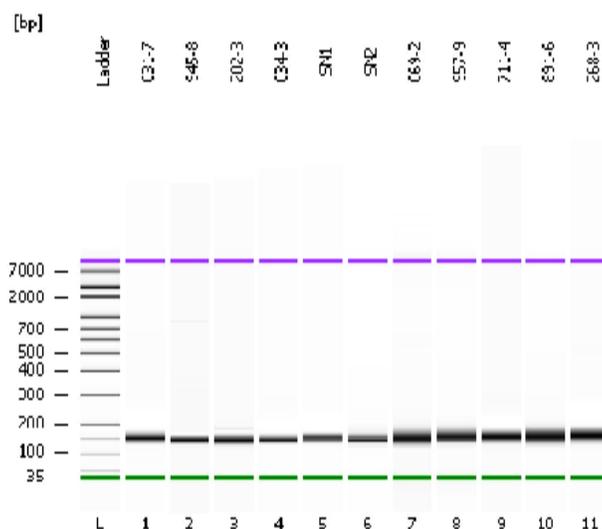


Figura 27. Control de calidad de bibliotecas. Electroferograma de chip de ARN de alta sensibilidad en Bioanalyzer 2100. En el carril L se encuentra el marcador de peso molecular y en los carriles 1 a 11 un subset de bibliotecas para secuenciar. Las bandas observadas corresponden a 150 pb que es lo esperado para las bibliotecas de miARNs. Las bandas magenta y verde corresponden al marcador interno superior (10380 pb) e inferior (35 pb) respectivamente.

Las secuencias obtenidas superan el millón de lecturas por muestra, mapeando en general más del 90% contra el genoma humano (Tabla 24).

Tabla 24. Datos primarios de las corridas de secuenciación de miARNs. Se indican el número de lecturas iniciales y las lecturas utilizadas para el análisis luego de recortar los adaptadores y filtrar por calidad. Además, se detallan la cantidad de secuencias y el porcentaje de mapeo contra el genoma humano.

<i>ID</i>	<i>Nº lecturas</i>	<i>Lecturas en análisis</i>	<i>Nº lecturas que mapean contra el genoma humano</i>
175-6	776331	546420 (70%)	507900 (93%)
224-6	160208	151430 (94%)	150393 (99%)
397-7	1184486	969957 (88%)	874181 (90%)
552-0	724835	585161 (81%)	557177 (95%)
707-6	521069	312034 (60%)	295657 (95%)
761-6	564956	337404 (60%)	308601 (91%)
STB10	3080508	2632878 (85%)	2521328 (96%)
STB11	1431326	1268368 (89%)	1181796 (93%)
STB16	907365	826034 (91%)	791199 (96%)
STB4	426722	372597 (87%)	345096 (93%)
STB5	241425	168223 (70%)	151535 (90%)
STB6	454667	410839 (90%)	388649 (95%)
STB9	1689659	1262336 (74%)	1067285 (84%)

Dentro de la diversidad de ARNs pequeños que fueron secuenciados, las lecturas con 21 a 23 nucleótidos corresponden en su mayoría a secuencias anotadas como microARNs en la base de datos miRBASE. Además, se identificaron otras lecturas anotadas como ARN ribosomal y ARN de transferencia y un porcentaje alto de secuencias mayores a 40 nt que no pudieron ser asignadas por el programa (Figura 28). Se continuaron los análisis con las secuencias anotadas de microARNs para evaluar su abundancia entre las muestras clínicas.

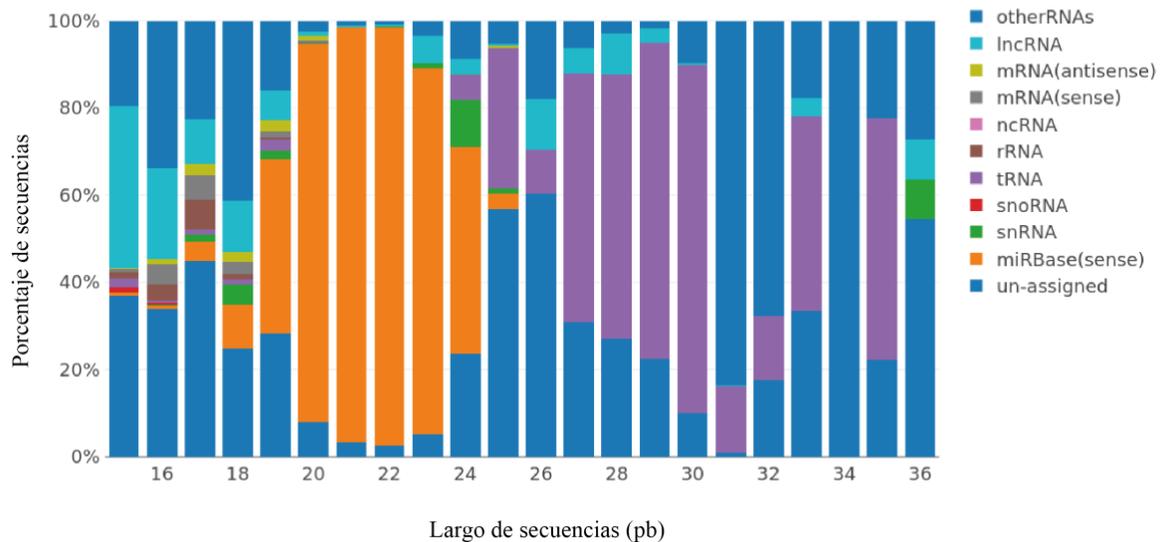


Figura 28. Composición de secuencias mapeadas de ARNs pequeños. Se indican los porcentajes de las lecturas que pudieron ser alineadas contra el genoma humano. La región de 20 a 23 nucleótidos corresponde al largo esperado para un miARN maduro; en este ejemplo mapean en su mayoría contra miARNs de la base de datos miRBase. Se destaca la presencia de ARNs ribosomales y de transferencia que no se abordaron en este estudio.

El programa sRNAbench ofrece la visualización de los miARNs más representados en una muestra dada. En la mayoría de los casos el miR-451a encabeza la lista siendo el más abundante en 12 de 13 muestras. Otros miARN que también son abundantes en la mayoría de las muestras son miR-148a-3p y miR-122.

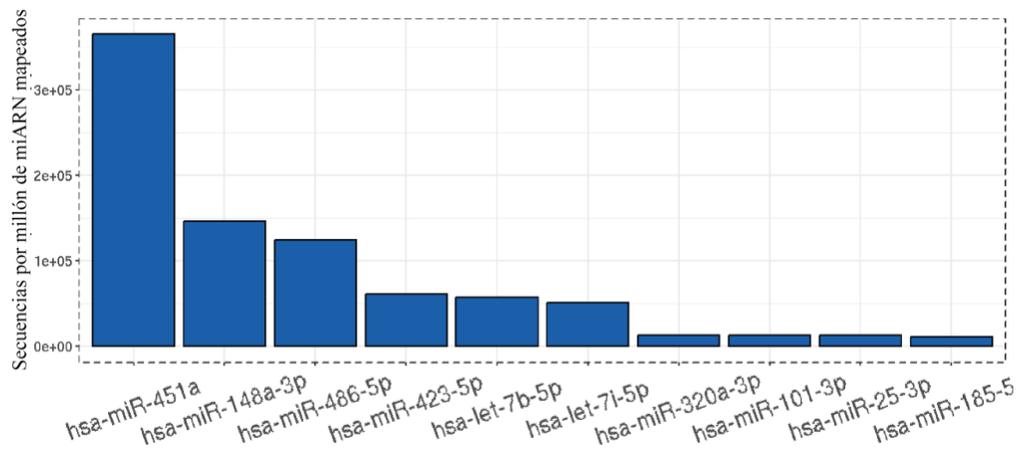


Figura 29. Top 10 de miARNs más abundantes. Ejemplo de los miARNs más abundantes para una muestra dada. El miR-451a es el más abundante en la mayoría de las muestras.

4.8. Abundancia diferencial de microARNs

Se comparó la abundancia diferencial de los miARNs obtenidos comparando los TB y los controles. En primer lugar, se realizó un *heatmap* que permite identificar datos atípicos o por fuera de la media (Figura 30). El agrupamiento de las muestras, según el perfil de abundancia diferencial de miARNs, logra clasificar las muestras según el grupo al que pertenecen.

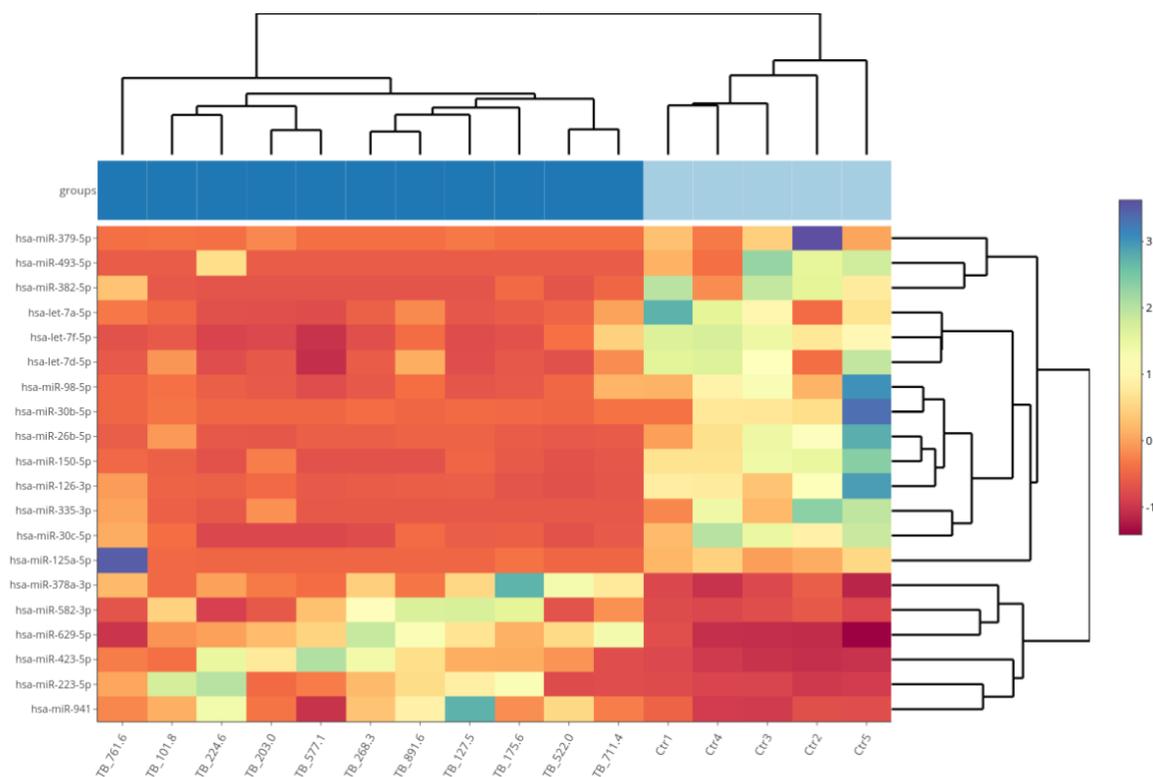


Figura 30. Heatmap de abundancia diferencial de miARNs en suero. Se grafican los 20 miARNs con el p-valor más bajo diferencial entre pacientes con TB (indicado con azul oscuro en la barra superior) vs muestras del grupo control (indicados en celeste). La barra de color indica la abundancia diferencial de cada miARN por muestra.

En segundo lugar, la abundancia diferencial de miARNs es calculada por medio de DESeq2. La normalización realizada por DESeq2 es utilizada para eliminar el efecto de diferente número de secuencias por muestras, o sea, que diferencias entre muestras no se deban a la profundidad de secuenciado. Los valores de conteo para cada miARN se dividen por un factor de escala que toma en cuenta el logaritmo en base e del total de lecturas para cada muestra. A partir de este análisis se seleccionaron 3 miARNs que se encontraron más abundantes en las muestras, con poca dispersión intra-grupo y que mostraron diferencias estadísticamente significativas entre ambos grupos (Tabla 25).

Tabla 25. Abundancia diferencial de miARNs entre el grupo TB y el grupo control. Se indican el promedio de lecturas de cada grupo, la diferencia de cambio (fold change) entre ellos, el logaritmo en base 2 de dicho cambio (Log2FoldChange) y el p-valor ajustado.

	<i>Prom. Controles</i>	<i>Prom. TB</i>	<i>Fold Change</i>	<i>Log2Fold Change</i>	<i>padj</i>
hsa-miR-378a-3p	79.78	1716.57	0.05	-3.05	0,00
hsa-miR-126-3p	3020.71	343.02	8.78	3.12	0,00
hsa-miR-423-5p	1398.25	15113.19	0.09	-3.43	0,00

En la figura 31, se representan las lecturas sobre millón de lecturas mapeadas que tuvieron los tres miARNs mencionados anteriormente, para cada una de las muestras analizadas.

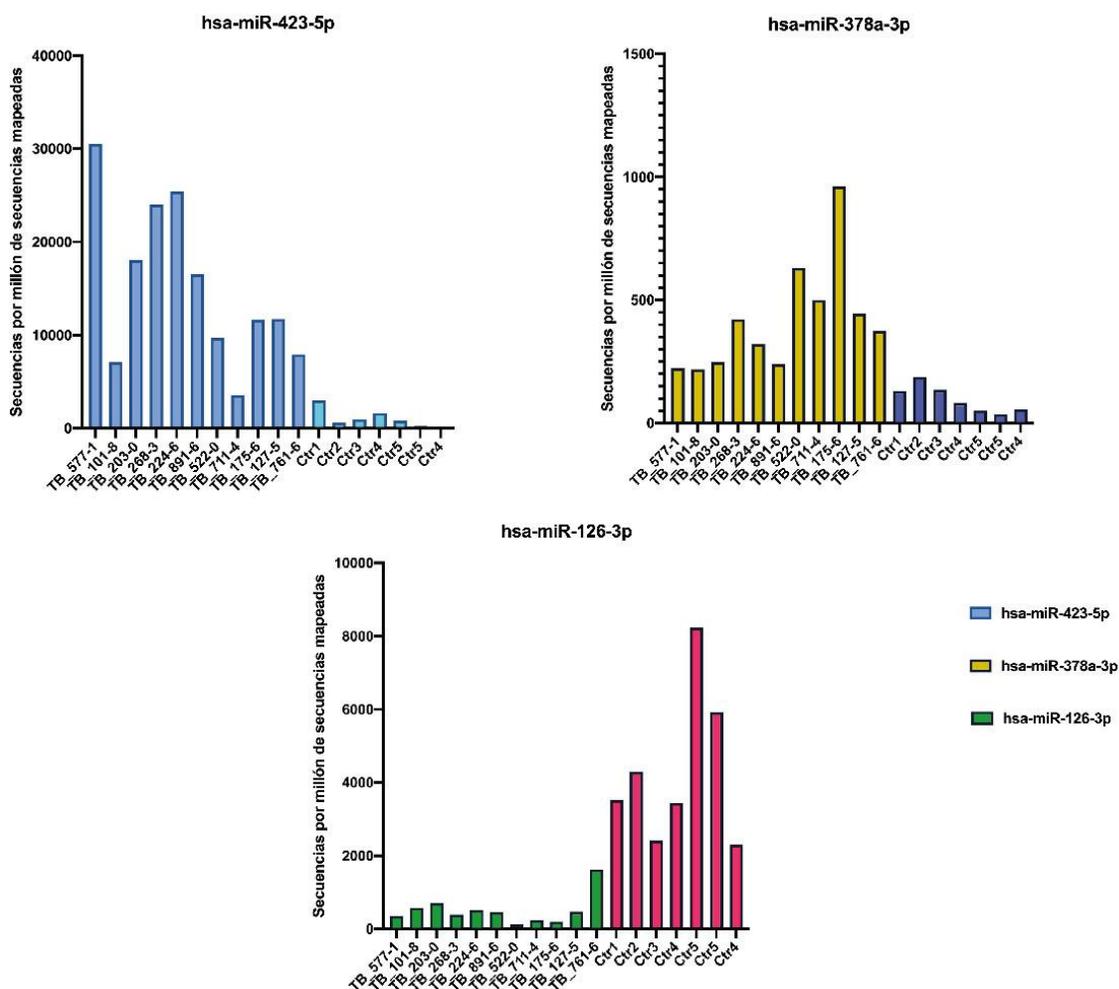


Figura 31. Representación gráfica de la cantidad de lecturas de miARNs por millón de secuencias mapeadas. miR-423-5p (azul), miR-378a-3p (amarillo) y miR-126-3p (verde). Los controles se identifican con otros colores para mayor claridad.

También se representaron los mismos datos en función de los tres miARNs diferencialmente abundantes (Figura 32). Se realizó un t-test múltiple, sobre las medias de todas las muestras por cada miARN. La diferencia entre las medias es estadísticamente significativa para estos miARNs.

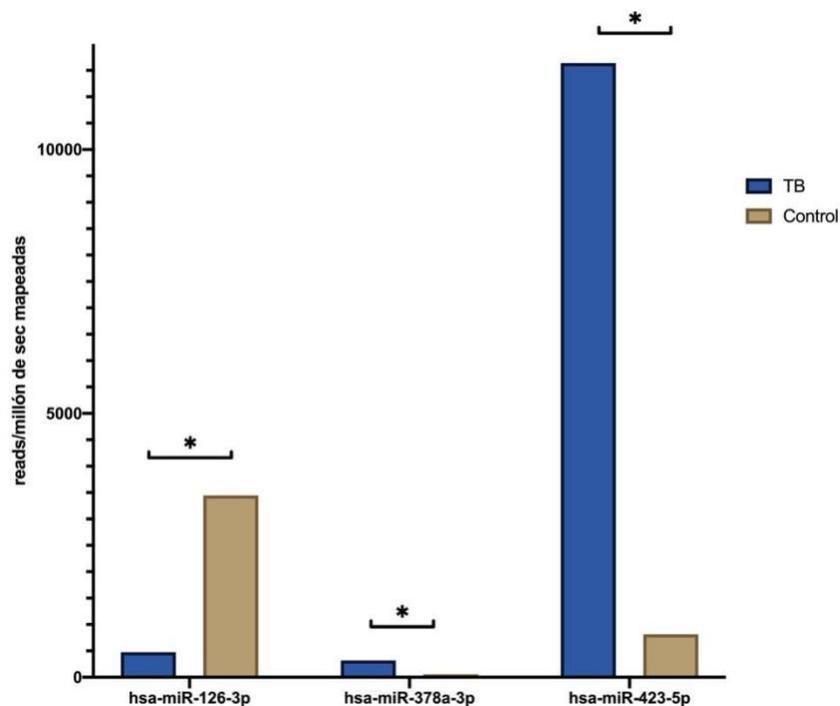


Figura 32. Comparación de miARNs diferenciales entre TB y controles. Se grafican el número de secuencias de cada miARN sobre el total de secuencias mapeadas. El asterisco indica diferencia significativa entre los grupos para miR-126-3p (p-valor= 0,0017), miR-378a-3p (p-valor = 0,009) y miR-423-5p (p-valor = 0,011), lo cuál coincide con el p-valor ajustado calculado por DESeq2.

Se desprende de este primer análisis que se encontraron miARNs diferencialmente abundantes entre ambos grupos y que esa diferencia es significativa. El miR-126-3p se observa en mayor proporción en el grupo control que en el grupo TB. Por el contrario, el miR-378a-3p y el miR-423-5p son más abundantes en el grupo TB respecto a los controles.

Se repitió el análisis de expresión diferencial, pero sin las muestras 761-6 donde el miR-126-3p parecería estar más abundante que en el resto de los TB y sin la muestra 522-0 ya que el cultivo resultó negativo en el transcurso del análisis. En este nuevo estudio se evidencia de igual forma, una clara separación de grupos y un perfil de abundancia de miARN similar (Figura 33). Nuevamente se observan diferencias significativas entre los grupos para los miARN antes mencionados, según la tabla 26.

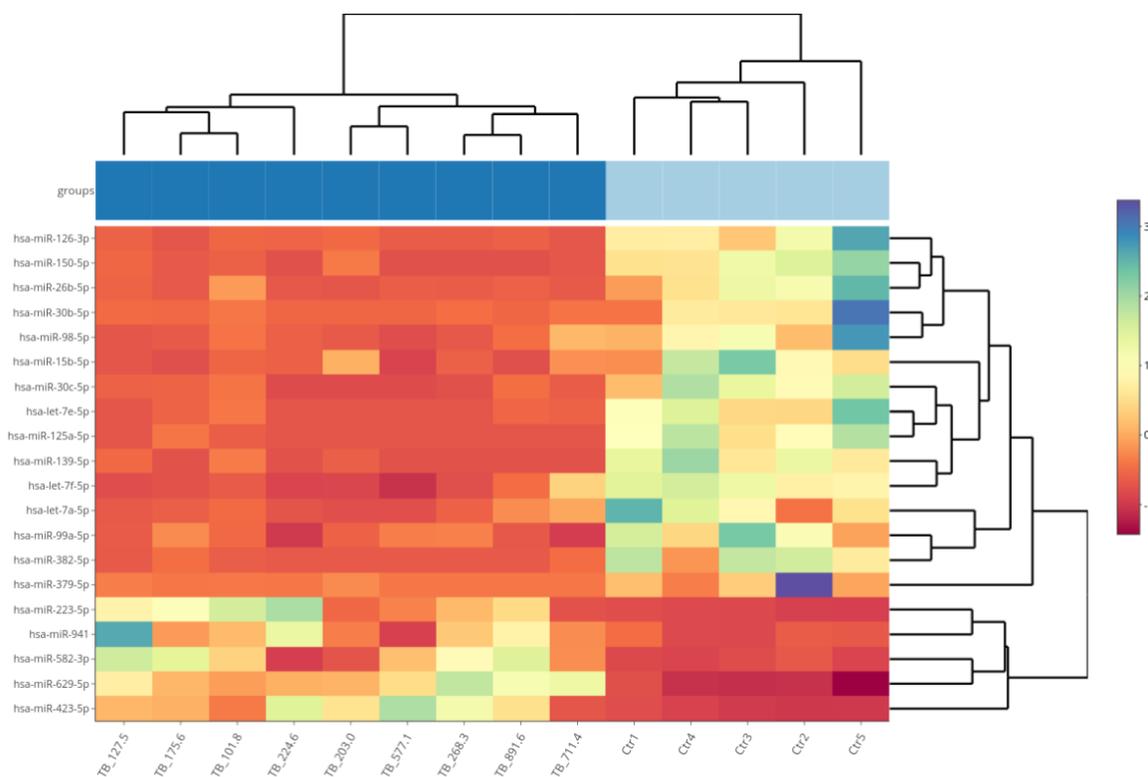


Figura 33. Heatmap de abundancia diferencial de miARNs en suero. Se grafican los 20 miARNs con el p-valor más bajo, circulantes en sueros de pacientes con TB vs control. Se omitieron las muestras 707-6, 397-7 y 577-1. La barra de color indica el factor de escala de cada miARN por muestra.

Tabla 26. miARNs diferencialmente expresados - TB vs Control sin las muestras 707-6 y 397-7. Se indican los promedios de cada grupo, el fold change entre ellos y el p-valor ajustado.

	<i>Promedio</i> <i>TB</i>	<i>Promedio</i> <i>Controles</i>	<i>Fold Change</i>	<i>Log2FoldChange</i>	<i>padj</i>
hsa-miR-378a-3p	429.42	127.28	0.3	-1.77	0,00
hsa-miR-126-3p	468.28	4677.52	9.97	3.32	0,00
hsa-miR-423-5p	17811.99	1487.22	0.08	-3.58	0,00

Dentro del grupo diagnosticado con TB, también se encuentran pacientes que ingresan a la UCI con diferentes estadios de la enfermedad y comorbilidades asociadas. La puntuación de APACHE II, cuyo valor elevado se asocia con peor pronóstico, es más alta en aquellos pacientes que fallecen. Se realizó un análisis de abundancia diferencial entre pacientes que fallecen (n=6) y sobreviven (n=5).

En primer lugar, el agrupamiento de las muestras no logra separar entre vivos y fallecidos. Al observar las lecturas de cada miARN por millón de lecturas mapeadas, tampoco se observan diferencias a simple vista entre los grupos (Figura 34), pero cabe destacar que las muestras 707-6 y 397-7 contribuyen de buena manera a la abundancia del miR-378a-3p en el grupo de TB que fallece.

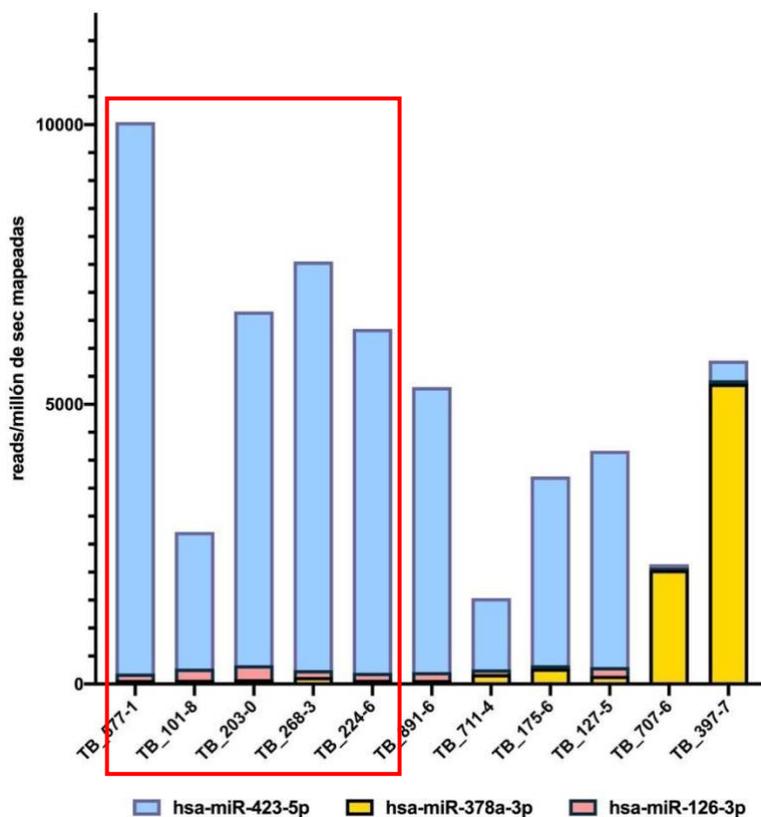


Figura 34. Representación gráfica de la cantidad de lecturas de miARNs por millón de lecturas mapeadas en función de cada muestra. El recuadro en rojo identifica aquellos pacientes que sobrevivieron.

Al analizar el total de lecturas en función de los tres miARNs seleccionados, se observa una abundancia en el miR-378a-3p en aquellos pacientes que fallecieron sin embargo esta diferencia no logra ser suficiente para ser estadísticamente significativa (t-test multivariado) (Figura 35).

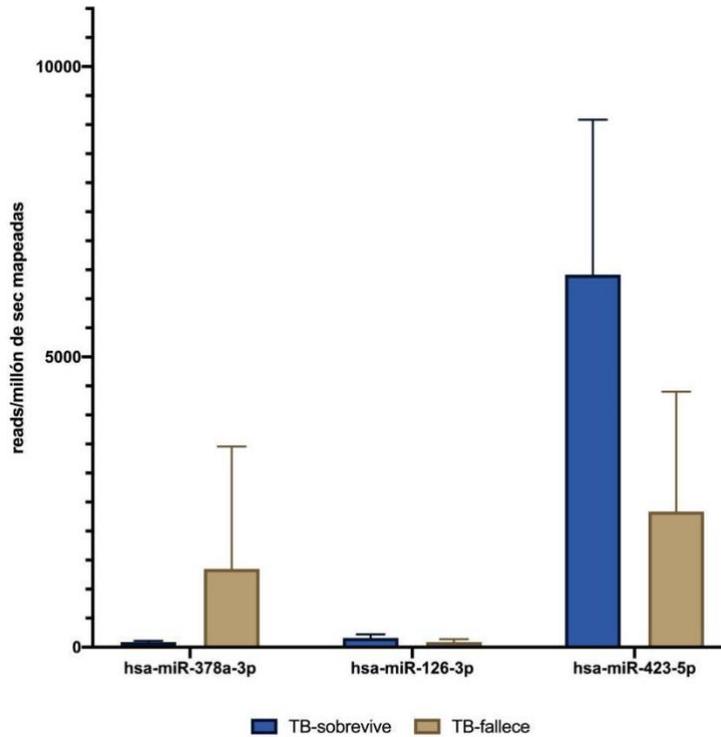


Figura 35. Cantidad de lecturas por millón de lecturas mapeadas en función de los miARN 378a-3p, 126-3p y 423-5p. Se distinguen los grupos de TB fallece y TB sobrevive

4.9. Validación de miARNs candidatos para identificación de TB

En base a lo observado anteriormente, se procedió a validar el miR-378a-3p, el miR-423-5p y el miR-126-3p a través de una cuantificación relativa por RT-qPCR y del análisis por el método $\Delta\Delta Ct$. Como miARN control o “housekeeping” se optó por el hsa-let-7b-5p ya que se encuentra presente tanto en los sueros controles como en los TB con niveles similares de abundancia, sin diferencias significativas entre los TB y el grupo control (Tabla 27).

Tabla 27. miARN de referencia entre TB y el grupo control. Se indican los promedios de cada grupo, el fold change entre ellos y el p-valor ajustado.

	<i>Promedio</i> <i>TB</i>	<i>Promedio</i> <i>Controles</i>	<i>Fold</i> <i>Change</i>	<i>Log2FoldChange</i>	<i>p</i> _{adj}
hsa-let-7b-5p	5988.49	3239.1	0.54	-0.89	0.13

Se efectuó el mismo protocolo de extracción de ARNs pequeños detallado anteriormente y se cuantificó la cantidad de ARN por fluorimetría utilizando el kit de detección de alta sensibilidad de ARN (Qubit, Invitrogen). Para la reacción de retrotranscripción se partió de cantidades equimolares de ARN. Se realizó el stem-loop RT-qPCR para las comparaciones de acuerdo al protocolo detallado en materiales y métodos.

Se logró amplificar el miARN de referencia let-7b-5p para los sueros utilizados en estos análisis (Figura 36). Para todas las corridas se realizaron controles negativos que incluyeron: control sin molde de ADNc (NTC), mix de reacción que incluye únicamente los cebadores del mix de qPCR (Mix), y control de retrotranscripción sin la enzima respectiva (S/E). Este último control resulta de importancia para además corroborar la amplificación de dímeros entre el cebador *forward* y el cebador stem-loop utilizado en la retrotranscripción.

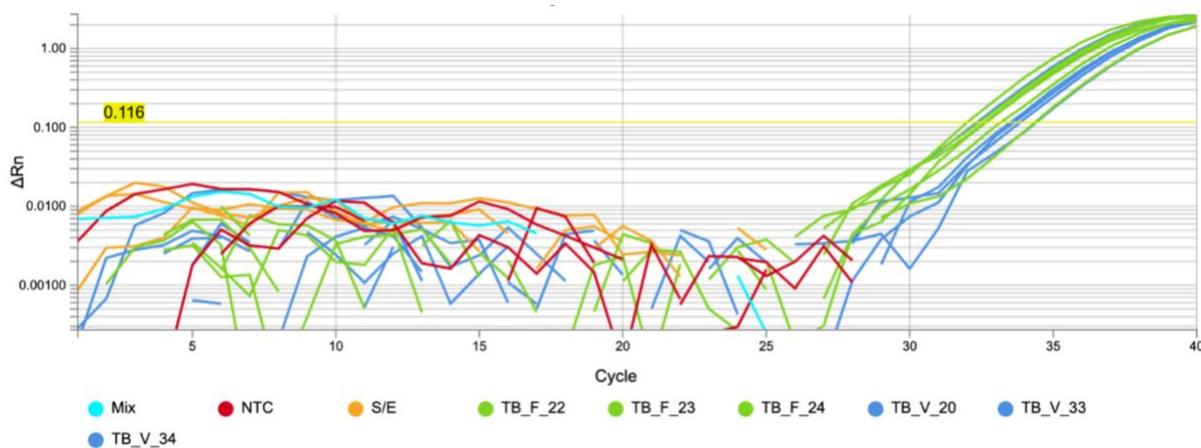


Figura 36. Curvas de amplificación para let-7b-5p. Las muestras se corrieron en duplicado y para este caso se incluyeron controles negativos sin molde (NTC), sin retrotranscriptasa (S/E) y se incluyó mix de reacción (Mix). Se seleccionó el umbral de corte y los Cts obtenidos se encuentran entre 31 y 34.

El análisis de los resultados de las curvas de PCR mostró diferencia significativa entre los controles y las muestras de TB cuando se cuantificó la abundancia relativa del miR-378a-3p (Figura 37-a) utilizando el miR-let7b como normalizador. No se pudo obtener valor estadístico significativo para sostener la diferencia observada en el miR-126-3p, aunque existe una tendencia a estar más abundante en los controles que en los sueros de pacientes con tuberculosis (Figura 37-b). Por otra parte, no se pudo obtener resultados para la validación del miR-423-5p ya que las curvas de amplificación de los controles negativos invalidaron la comparación de las curvas.

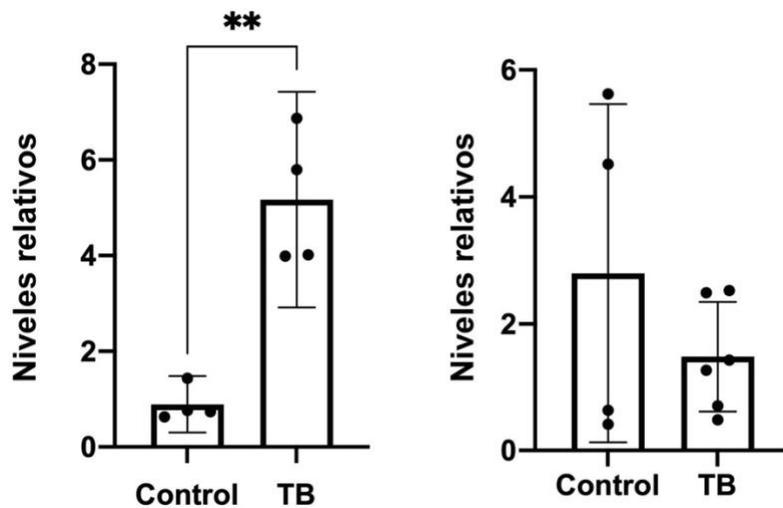


Figura 37. Cuantificación relativa de miARNs candidatos circulantes en suero. La cuantificación relativa de las curvas de qPCR se realizaron por el método de $\Delta\Delta C_t$. (a) cuantificación del miR-378a-3p relativa al let-7b-5p. (b) cuantificación del miR-126-3p relativa al let-7b-5p.

5. DISCUSIÓN Y CONCLUSIONES

PARTE A- Tipificación de aislados y búsqueda de resistencia

5.1. Tipificación de cepas de una población de personas privadas de libertad

La prevalencia de TB en la población de privados de libertad (PPL) en el mundo es elevada. En general esta alta tasa está relacionada con características de esta población como malas condiciones socioeconómicas, HIV, alcoholismo, adicción, poco apego al tratamiento, etc. (Walter et al., 2021). Además, la convivencia con personas con TB en condiciones de sobrepoblación, hacinamiento, poca ventilación y poco acceso al sistema de salud, aumentan las chances de transmisión de la TB dentro de las cárceles. En países con alta tasa de TB y condiciones de hacinamiento extremo en cárceles, resulta en una probabilidad de transmisión anual del 90% (Johnstone-Robertson et al., 2017). La prevalencia de TB en cárceles en Uruguay es elevada y en 2020 se ubicó en 639 casos cada 100.000 habitantes según la CHLA-EP. La situación del sistema penitenciario resulta crítica y hasta el año pasado se estimó que hay algo más de 130 presos por cada 100 plazas disponibles, siendo la densidad más alta registrada desde 2013 (Comisionado Parlamentario Penitenciario, 2020). Hasta marzo de 2021 la población carcelaria era de 13.402 internos, y la tendencia al hacinamiento crítico no parecería revertirse en los próximos años. La situación de la TB en el sistema penitenciario no es solamente crítica para los PPL, sino que además estas instituciones actúan como amplificadores de la tuberculosis comunitaria, manteniendo la propagación de la enfermedad desde los entornos penitenciarios hacia la población exterior (Walter et al., 2021).

En este trabajo, se puso a punto una técnica de tipificación y caracterización de cepas a través del secuenciado profundo y análisis de polimorfismos en regiones de interés. Este abordaje es cada vez más utilizado en el mundo y hasta la fecha no se había realizado en el país. Se realizó la caracterización de 47 aislados del sistema penitenciario (SP) obtenidos durante todo el año 2019. Se obtuvieron en total 47 ADN de cepas aisladas del SP, de los cuales 28 corresponden a aislados que se detectaron en el triage al ingreso del SP. De estos últimos, 25 ya tenían ingresos previos en el sistema mientras que 3 correspondían a personas que ingresaban por primera vez. Por otra parte, 19 aislados fueron detectados dentro de las cárceles. De los 47 aislados secuenciados, hubo uno donde no se logró la amplificación probablemente debido a la baja calidad del ADN utilizado, mientras que para 6 casos no se pudo detectar polimorfismos para clasificar el genotipo. Para estos últimos se realizó la tipificación por MIRU-VNTR, que dio como resultado la asignación al sublinaje LAM para 4 de ellos, mientras que en los 2 restantes no se logró determinar el linaje.

Los resultados observados para los 44 aislados tipificados distingue la presencia de 17 cepas del sublinaje LAM - 4.3 - (39%), 15 aislados correspondientes al linaje X - 4.1.1 - (34%), 11 cepas del linaje Haarlem - 4.1.2.1 - (25%), y 1 cepa del linaje S - 4.4.1.1 - (2%). En primer

lugar, cabe destacar que las cepas circulantes son todas pertenecientes al linaje 4 de *M. tuberculosis*, el cual es el linaje más extendido en todo el mundo. En un trabajo previo de nuestro grupo, se pudieron tipificar mediante MIRU-VNTR 60 aislados de pacientes graves en la UCI y 90 aislados comunitarios, que no requirieron internación y que fueron obtenidos al azar del banco de ADN de aislados de la CHLA-EP entre los años 2016 hasta 2019. Los aislados circulantes tanto en UCI como en la comunidad también pertenecen al linaje 4, siendo LAM y Haarlem los sublinajes más abundantes en todos ellos (Hurtado, et al. 2021). Se desprende de estos resultados que en el SP circula en mayor proporción una cepa en particular (sublinaje X) la cual no circula en la comunidad. Además, todas las personas cuyos aislados fueron tipificados como X, eran reincidentes o se encontraban dentro del SP.

Existe una asociación positiva entre algunos linajes de *Mycobacterium tuberculosis* y resistencia a fármacos anti-tuberculosis (como la familia Beijín), mientras que otros como LAM y X son más frecuentes en personas con VIH positivo (Couvin et al., 2019). La familia X se encuentra en baja proporción en el continente americano, siendo el 1,2% de los casos reportados en la base de datos SITVIT2 (<http://www.pasteur-guadeloupe.fr:8081/SITVIT2/index.jsp>). Estados Unidos concentra el 76% de los casos asignados como X, reportados en esta base de datos.

Si bien la cepa encontrada está en muy baja proporción en los aislados de la comunidad, restaría profundizar esta información haciendo un relevamiento de cepas circulantes más amplio. Por ejemplo, realizar un seguimiento epidemiológico de la transmisión extra-cárcel a personal de los centros de reclusión y al entorno más próximo a los PPL en la comunidad. De esta forma se podría obtener información sobre su transmisión, así como el origen y procedencia de esta familia de *Mtb* en nuestro país.

Para finalizar, se compararon los resultados propuestos por nuestra metodología local con los resultados arrojados por el programa PhyResSE disponible online. La clasificación resultó concordante para 23 de 47 de los aislados y en un caso la asignación fue ambigua. Este programa, permite realizar un árbol de máxima probabilidad y sobre él se anotaron los linajes de cada aislado (Figura 38). En la figura se señala con flecha aquellos aislados que fueron concordantes por ambas aproximaciones. Para el caso de la muestra 26, cuya clasificación era ambigua, se decidió tomar como referencia el árbol y se asignó esta muestra como linaje X. También, hubo dos casos en que la clasificación por nuestra metodología era ambigua si se comparaba en el contexto del árbol generado por PhyreSE (**), en estos dos casos se continuó con la asignación realizada por nuestra metodología.

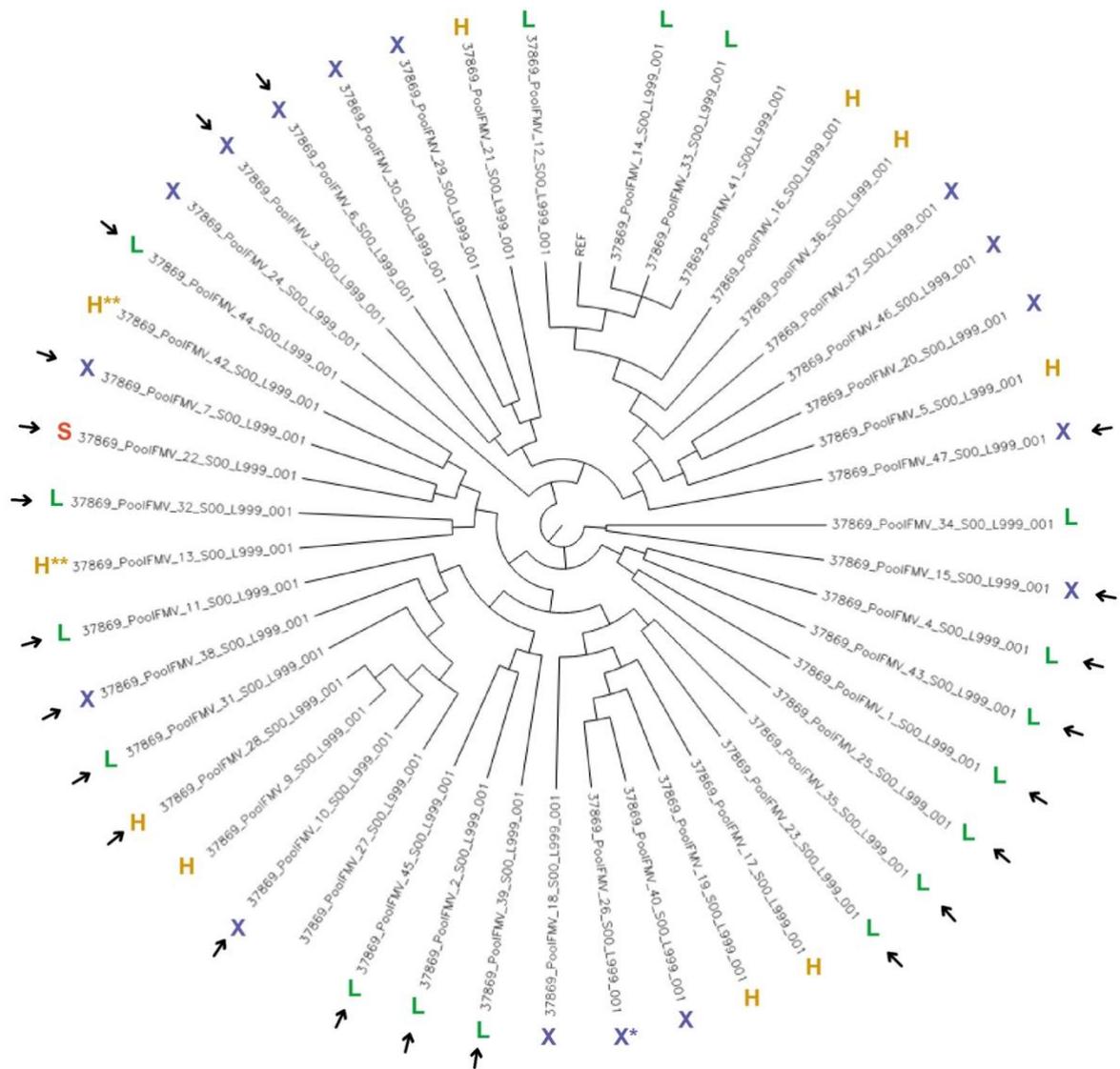


Figura 38. Árbol filogenético de máxima probabilidad generado con PhyResSE. H: Haarlem, L: LAM X:X y S:S. Los aislados señalados con flecha identifican aquellas muestras que fueron tipificadas por ambos análisis y fueron coincidentes. Para el caso de la muestra 26, la misma fue identificada como familia X por PhyResSE y como familia Haarlem por nuestra metodología (*). Para la muestra 42 y 13, la tipificación realizada no es concordante con lo observado en el árbol (**). Las muestras 27 y 41 no pudieron ser identificadas por ningún método.

Si bien hace falta una mayor vigilancia epidemiológica y hacer un estudio profundo de la transmisión intrafamiliar, o incluso de la transmisión dentro de los funcionarios penitenciarios, todo hace indicar que esta cepa circula activamente dentro del SP, con una diseminación comunitaria baja.

5.2. Resistencia en aislados del sistema penitenciario

La resistencia a la TB surge principalmente como consecuencia de dos mecanismos: resistencia primaria o transmitida, es decir cuando una cepa resistente se transmite a un nuevo hospedero, o resistencia secundaria o adquirida, que ocurre por adquisición de mutaciones en genes “target” de una o más drogas. Estas mutaciones son por lo general sustituciones nucleotídicas, inserciones o deleciones pequeñas. A diferencia de otras bacterias, en *Mycobacterium tuberculosis* no hay transferencia horizontal de genes a través de elementos genéticos móviles como los plásmidos.

La secuenciación de genoma completo de *M. tuberculosis* proporciona información temporal y geográfica respecto a la adquisición de resistencia (Ektefaie et al., 2021). Se estima que la probabilidad de adquirir una mutación espontánea a rifampicina es de 1 en 10^8 bacilos y 1 en 10^6 bacilos para isoniazida, estreptomycin y etambutol. Sin embargo, los datos observados indican que el número de eventos independientes de aparición de resistencia es mucho menor que los aislados resistentes registrados en el mundo. También, la tasa de mutaciones que causan resistencia varía dependiendo del linaje (Dookie et al., 2018). En otro estudio realizado por Lanzas et al., 2013 se registró un brote de MDR-TB en Panamá que se estaba dando por la expansión clonal de varias cepas de la familia LAM, dentro de la cual una familia en particular, LAM9-c1, estaba muy relacionada a cepas XDR de origen africano.

Las instituciones como hospitales y centros de reclusión no solamente actúan como amplificadores de la TB en la comunidad, sino que además aumentan el riesgo de la ocurrencia de tuberculosis multi-drogoresistente (Warren et al., 2018). Para la población de estudio de este trabajo, se detectaron mutaciones en la mayoría de los aislados, no obstante, el 57% de esas mutaciones son sinónimas. Para aquellas mutaciones no sinónimas, se detectó que al menos 27 aislados presentaban mutaciones en el gen *gyrA*: Glu62Gln, Ser95Thr y Gly668Asp. Este gen está asociado a resistencia a las fluoroquinolonas y en él se describen algunas mutaciones de alta confianza para resistencia fenotípica. Por ejemplo, la mutación Asp94Gly ocurre frecuentemente en cepas XDR (Singh et al., 2020). En cualquier caso, las mutaciones detectadas por nuestro flujo de trabajo, que a su vez fueran corroboradas por los resultados observados de PhyResSE, no se encuentran reportadas en las tablas de datos para mutaciones asociadas a resistencia a fluoroquinolonas (Chaiyachat et al., 2021).

También se encontraron dos mutaciones puntuales para el gen *rpoB* en dos aislados diferentes, Leu275Val en el 100% de las secuencias (163x de cobertura) y la mutación Met349Thr en el 99% de las secuencias (169x de cobertura) respectivamente. Sin embargo, estas mutaciones tampoco se encuentran reportadas asociadas a resistencia fenotípica. El gen *rpoB* codifica para la subunidad beta de la ARN polimerasa y en el 96% de los aislados clínicos que exhiben resistencia fenotípica, el mismo presenta alguna alteración en la región 507-533, como la mutación Ser531Leu (Dreyer et al., 2020). Ninguno de los SNPs

observados por nuestro flujo de trabajo, ni tampoco aquellos detectados con PhyResSE coincidieron con esta región o se encontraron reportados en la bibliografía.

Por otra parte, se encontró la mutación c1472337t (c492t) en el gen *rrS* para tres aislados y dicha observación fue corroborada por el análisis de PhyResSE. Esta mutación se encuentra reportada por (Coll et al., 2014), aunque no se encontró ninguna otra evidencia que indicara que la asociación con la resistencia fenotípica está fuertemente relacionada. Sin embargo, mutaciones en el gen *rrS* se encuentran asociadas a resistencia a estreptomicina, particularmente las regiones 530 y 912, que son los sitios más frecuentes. En particular las mutaciones a514c, c517t y a908g fueron encontradas en cepas Beijin resistentes (Smittipat et al., 2016). También existe evidencia de que el tipo y la frecuencia de mutaciones en *rrS* dependen del sitio geográfico y que algunos linajes, como Beijin o LAM están más frecuentemente asociados con la ocurrencia de resistencia a estreptomicina (Lipin et al., 2007). Los aislados con la mutación c492t corresponden a los aislados 1, 32 y 45 que fueron tipificados como LAM.

El análisis por PhyResSE detectó 10 aislados con mutaciones en genes de resistencia. En todos los casos, las mutaciones encontradas fueron concordantes con lo observado en nuestro flujo de trabajo, por lo que es una buena aproximación para validar el análisis informático realizado. De todo lo anterior, surge que todos los aislados estudiados son sensibles a los fármacos de primera línea utilizados para tratar la TB. Sin embargo, es necesario controlar los aislados con resistencia a estreptomicina detectados, ya que, de requerir el tratamiento con este fármaco, podría conducir a resistencia al mismo.

De esta forma, por medio del análisis informático, se logró filtrar y examinar la información contenida en las secuencias de ADN de aislados de TB. Se logró discriminar los linajes de la mayoría de cepas, evidenciar cambios en los genes asociados a resistencia a fármacos anti-tuberculosis y determinar si estaban asociados a resistencia fenotípica. El flujo de trabajo propuesto tiene la ventaja de que puede ser ajustado para la búsqueda de cepas y de resistencia más frecuentes en la realidad local. A través del secuenciado profundo se obtiene información crítica que puede mejorar los tratamientos de TB. Se pueden conocer las cepas circulantes, si están asociadas con peor pronóstico y tener conocimiento de la predisposición genotípica de un aislado a desarrollar resistencia a un fármaco para poder hacer un mejor uso de los antibióticos. La existencia de programas online que realizan las mismas funciones haría posible la aplicación del análisis informático sin tener conocimiento previo de línea de comandos ni herramientas informáticas. Por otro lado, el sistema utilizado en este trabajo también podría mejorarse, por ejemplo, modificando el pool-L para incluir más polimorfismos que permitan una mejor resolución de las cepas del linaje 4 (las únicas observadas en nuestro país) y quitar los cebadores de los linajes menos frecuentes. También podrían ser incorporados, en el pool-R, nuevos genes asociados a resistencia en la medida que surjan.

5.3. Puesta a punto y consideraciones finales

Se logró generar dos PCR en múltiplex que permiten amplificar regiones específicas de *Mycobacterium tuberculosis*. Por un lado, se puso a punto una múltiplex para detectar polimorfismos característicos en 11 regiones para los diferentes linajes y sublinajes y por otro, en 9 regiones para detectar resistencia a los principales fármacos antituberculosis. Para obtener datos de un aislado particular, se debe realizar la amplificación de ambos pools por separado, para luego hacer una única mezcla en la preparación de bibliotecas para secuenciación.

Las bibliotecas fabricadas con el kit Nextera XT DNA Library Kit (Illumina), fueron secuenciadas en un MiSeq en corrida paired-end de 2x75 bases. Utilizando el cartucho de secuenciación MiSeq Reagent Kit v3, se podrían secuenciar hasta 384 muestras a la vez y si se obtuvieran 10 millones de secuencias, se lograría una cobertura teórica de 181x por muestra (cantidad de lecturas por el largo -150 pb- dividido el tamaño de las regiones a secuenciar -21500 pb-). La secuenciación de amplicones es diferente a la secuenciación de genoma completo, en el sentido de que lo esperable para secuenciar todo el genoma son bibliotecas con mediana cobertura, con secuencias únicas, distribuidas homogéneamente y con una distribución de bases equitativa. No obstante, en la secuenciación de amplicones, se obtienen muchas secuencias idénticas que mapean exactamente en una misma región del genoma, por lo que es esperable que exista una alta cobertura y también un número elevado de secuencias duplicadas. A través de esta estrategia y realizando secuencias que permitan una cobertura alta, es posible visualizar variantes poco frecuentes. Sin embargo, esto también lleva a una baja diversidad de secuencias, por lo que pueden aumentar los errores en la asignación de bases debido a que es más difícil para el equipo distinguir “clusters” entre bibliotecas homogéneas, es decir, que tome como una sola señal algunos “clusters” diferentes que se encontraran próximos físicamente en la celda de secuenciación. Este problema, puede disminuirse mediante una matriz de corrección que utiliza el propio MiSeq y también adicionando un porcentaje más alto de la biblioteca estándar Phi X (genoma viral), que aumenta la diversidad y mejora la asignación de bases.

Respecto a los resultados obtenidos, se obtuvo un alto porcentaje de cobertura para la mayoría de las regiones que se analizaron, sin embargo, se observó una gran diversidad en la cobertura de las diferentes regiones (tanto de linaje como de resistencia), indicando la necesidad de realizar cambios en las proporciones de cebadores hasta lograr tener una cobertura homogénea para todas las regiones. Como la PCR en múltiplex es un compromiso y un balance entre las diferentes temperaturas de fusión de los cebadores, algunos amplicones pueden haber sido favorecidos. Para ajustar el desbalance se intentó realizar la estrategia de una primera ronda de PCR de 5 ciclos y también ajustar las concentraciones de cebadores. Sin embargo, en algunos genes la cobertura fue baja, aunque la calidad de las corridas fue en su gran mayoría muy buena (Phred score > 30).

Para el análisis informático, se compararon dos estrategias de alineamiento: una contra todo el genoma y otra únicamente contra las regiones buscadas. Para esta última, se confeccionó una base de datos concatenando todos los genes o regiones. Sin embargo, el mapeo contra esta base de datos tuvo un rendimiento más bajo. Esto puede deberse a que existen amplicones que también amplifican regiones por fuera del gen o de las regiones analizadas, y la referencia de genes concatenados no permite el alineamiento de amplicones en el límite. Por lo tanto, se estudiaron las lecturas de las bibliotecas contra el genoma de referencia de *Mycobacterium tuberculosis* y se analizaron por un lado las regiones para linaje y por otro los genes para resistencia.

PARTE B- Búsqueda de marcadores moleculares de tuberculosis

5.4. miARNs como marcadores moleculares de enfermedad

El avance tecnológico en el área de la biología molecular y en particular en el área de secuenciación, resulta un insumo necesario para la epidemiología molecular de la TB. Cada vez es más urgente aunar todos los esfuerzos posibles para alcanzar el objetivo de erradicación de la TB en el mundo. Hoy en día, se puede contar con información precisa de cuáles son las cepas circulantes en una comunidad dada, y poner el foco en variantes de preocupación (como el caso de Beijing, o variantes LAM asociadas a casos MDR) y cómo fue la relación epidemiológica entre ellas. Toda esta información al servicio de la clínica de la TB podría mejorar el rastreo y el muestreo de los casos activos, a pesar de ello, también se precisan biomarcadores que puedan revelar la progresión desde la enfermedad latente a una enfermedad avanzada o que pronostiquen la probabilidad de reactivación luego de curarse. La existencia de biomarcadores robustos de TB podría prevenir o advertir casos de enfermedad grave o brotes de TB en la comunidad.

Los ácidos nucleicos extracelulares, tanto dentro de vesículas como asociados a proteínas de unión a ARN o ADN, se pueden encontrar circulando en sangre. Las variaciones en su concentración y su composición son reflejo de estados metabólicos particulares o procesos de enfermedad y es por ello que conocer las secuencias y las abundancias de estos ácidos nucleicos puede ser información útil para la búsqueda de biomarcadores. Se ha visto que *Mycobacterium tuberculosis* tiene la habilidad de evadir la respuesta del sistema inmune y sobrevivir de forma intracelular en el hospedero. Los mecanismos de evasión también incluyen la supresión de ciertos miARNs del hospedero involucrados en la regulación del sistema inmune y adaptativo (Sinigaglia et al., 2020). Algunos miARN que se han visto más abundantes dentro de pacientes con TB son miR-146a-5p, miR-21-5p, miR-99b-5p y miR-132-5p, que están asociados negativamente con la respuesta inflamatoria del hospedero,

participando en la disminución de la señalización de los receptores de tipo toll (TLRs), favoreciendo el crecimiento de la bacteria (Kim et al., 2017). Otros miARNs como miR-29a-3p y miR-125-5p que se encuentran más abundantes en macrófagos infectados, directamente actúan sobre IFN-gamma y TNF-alfa suprimiendo la respuesta inmune (Rajaram et al., 2011). También se destacaron a hsa-miR-196b y hsa-miR-376c como potenciales marcadores de diagnóstico, luego de analizar suero de 15 pacientes con TB y 82 controles. Estos fueron validados por RT-PCR (Zhang et al., 2014). Lo cierto es que existen numerosos artículos que proponen miARNs como marcadores de TB, pero los resultados no son muy consistentes. Si bien es sabido que las bacterias, a través de la expresión de miARNs, pueden manipular diferentes vías moleculares del hospedero (Maudet et al., 2014) y que los ARNs pequeños en vesículas extracelulares pueden actuar en la comunicación a distancia entre células, no se ha logrado comprender en profundidad cuáles son los mecanismos celulares detrás de ello. Los perfiles de expresión de miARN también dependen de características de la población de estudio, como edad, sexo, comorbilidades, aptitud del sistema inmune pero también del tipo de cepa de *Mtb* o del estado o avance de la enfermedad (Ruiz-Tagle et al., 2020). También es posible que ocurran ciertos sesgos en cuanto a los diferentes métodos de aislación, así como degradación de ARN en los biofluidos, lo cual hace que la cuantificación e identificación de la abundancia de los ARNs extracelulares esté sujeto a artefactos producidos durante el aislamiento y/o procedimiento (Max et al., 2018).

Para tratar de mitigar estas variaciones en la cuantificación de muestras, en este trabajo todos los sueros de pacientes graves y controles, partieron de la misma cantidad de suero y fueron extraídos de la misma manera con kit comercial diseñado para la extracción de miARNs. Además, los sueros fueron inmediatamente congelados a -80° hasta su posterior uso. También se prepararon las bibliotecas para secuenciar utilizando el mismo protocolo y partiendo de la misma cantidad de ARN cuantificado por métodos fluorométricos. Otra estrategia que se podría realizar es la de incluir en el proceso de extracción, un miARN sintético que no mapeara contra el genoma humano (spike de miARN). De esta forma se podrían analizar la eficiencia de recuperación y analizar si ocurrieron cambios significativos entre los grupos, durante todo el proceso.

5.5. Abundancia diferencial de miARNs

De la población de pacientes graves internados en UCI utilizados para este análisis, se desprende de sus datos clínicos que se trata de una cohorte con predominancia masculina (62%) en torno a los 40 años de edad, donde el 62% falleció en su estadía en la UCI con un puntaje promedio de APACHE II de 29. Se trata entonces de pacientes jóvenes, con un *score* de severidad muy alto, que se corresponde con una mortalidad que triplica la mortalidad promedio observada en otros casos. A partir de la extracción de ARN, se realizaron las bibliotecas para secuenciación de ARNs pequeños y se obtuvieron secuencias de muy buena calidad que mapean más del 90% contra el genoma humano.

Cuando se observó la composición de miARNs de las muestras, los miARNs 451a y 486-5p se encontraron dentro de los más abundantes en todos los sueros analizados. Sin embargo, estos miARNs se encuentran normalmente enriquecidos en glóbulos rojos de la sangre (Max et al., 2018). El análisis de abundancia diferencial, no mostró diferencias entre los grupos para estos miARNs. La presencia de ellos puede deberse a la incorrecta separación de fases en el proceso de recolección de muestras.

Del análisis de expresión diferencial, surgieron tres candidatos como marcadores de enfermedad. El hsa-miR-378a-3p y el hsa-miR-423-5p se hallaron más abundantes en el suero de los pacientes graves con TB mientras que el hsa-miR-126-3p estaba disminuido en este grupo. El análisis realizado por DESeq2 encontró diferencias significativas. El miR-378a-3p está asociado a la regulación negativa de la proliferación celular y está involucrado en el silenciamiento de la respuesta inmune (Dubois-Camacho et al., 2019). Asimismo, el miR-126-3p se encuentra asociado a la regulación negativa de la migración celular y supresión de tumores (Felli et al., 2013). Sin embargo, estos dos miARNs no se encuentran reportados como marcadores de TB. En cambio, el miR-423-5p está al menos reportado en un artículo científico como más abundante y de forma estadísticamente significativa, cuando se comparan suero de pacientes con tuberculosis pulmonar activa y controles sanos (Tu et al., 2019). Los autores señalan, además, que se logró establecer una asociación entre la abundancia de este microARN y la inhibición en la maduración del fago-lisosoma en macrófagos.

También se realizó un análisis de abundancia diferencial de miARNs circulantes en suero entre pacientes con TB que sobrevivieron y que fallecieron en la UCI. En este caso, el bajo número de muestras y la dispersión de los datos no permitió observar miARNs con diferencias estadísticamente significativas entre los grupos para ser considerados candidatos con valor pronóstico. Restaría aumentar el número de muestras analizadas, para poder determinar si el miR-378a-3p o el 423-5p son además biomarcadores pronósticos de gravedad.

En una revisión publicada recientemente por Ruiz-Tagle et al., se resumen algunos trabajos de expresión diferencial de miARNs en tuberculosis, que incluyen análisis de suero. Los autores señalan que encontraron algunos candidatos que constantemente se asocian a formas activas de TB: hsa-miR-144-3p, hsa-miR-144-5p, hsa-miR-146a, hsa-miR-155, hsa-miR-21 y hsa-miR-29a. En el análisis de abundancia diferencial realizado en este trabajo, no se encontraron diferencias en la abundancia para estos miARNs. Cabe destacar que el miR-144 también se ha reportado como un miARN abundante en biofluidos (Max et al., 2018).

En suma, partimos de una población de estudio muy heterogénea que ingresaron a UCI por cuadro clínico compatible con tuberculosis pulmonar, que resultaron tener cultivo positivo,

pero que también tenían comorbilidades asociadas como tabaquismo, consumo de drogas, situación de calle y donde 1 de cada 5 era VIH positivo. Existen a su vez, diferentes estadios en la progresión de la enfermedad en el ingreso a la UCI, que pueden verse reflejados en la puntuación de APACHE II. La estadía en UCI (que en general es más corta en pacientes que fallecen), podría ser una aproximación también para definir diferentes estadios, indicando una enfermedad muy avanzada. Cuando se analizó la cohorte en función de la mortalidad no se encontraron miARNs circulantes en suero significativamente abundantes en un grupo respecto a otro, pero se observaron que algunas muestras presentaron un perfil de abundancias distinto al resto. Dado que se trata de una población muy heterogénea y que la cantidad de pacientes analizados es pequeña, no pudimos observar marcadores con valor pronóstico ni excluir la asociación de las diferentes comorbilidades en los resultados observados. Sin embargo, el miR-423-5p se encontró abundante y de forma significativa en los pacientes con TB y también se encontró reportado en la bibliografía, lo que lo convierte en un posible marcador de tuberculosis activa.

5.6. Validación de los miARNs

Los sueros elegidos para la validación no fueron los mismos que los utilizados en el análisis de abundancia diferencial ya que en muchos casos se trataba de muestras únicas. No obstante, encontrar estos marcadores en otro set de muestras podría indicar mayor robustez a los mismos.

La amplificación del ADNc en el mix de qPCR tuvo algunas dificultades a la hora de analizar las curvas en relación a los controles negativos. Para la mayoría de los casos, el mix de PCR (sin agregado de ADN molde) amplificó a Cts altos. Por otra parte, el propio diseño de los cebadores puede dar lugar a complementariedad entre el cebador SL-RT y el cebador *forward* para algunos miARNs. Tomando en cuenta que solamente tienen 21 nt de largo, puede ocurrir que existan cebadores solapantes y no sea posible realizar la validación. Se observaron los perfiles de las curvas de *melting* para asegurarnos que el perfil de los controles negativos fuera diferente al de las muestras. Cuando se realizó la validación del miR-423-5p se obtuvo amplificación tanto en los controles negativos como en el mix de qPCR. Los Cts de las curvas pertenecientes a la amplificación del ADNc y los controles negativos eran prácticamente los mismos y al realizar un gel de agarosa se pudo corroborar la amplificación de bandas de muy bajo peso molecular en todas las réplicas. Esto puede ocurrir debido a la complementariedad entre el extremo 3' del cebador *forward* y la región del SL-RT que hibrida contra el miARN, por lo que no fue posible su validación. Se intentó confeccionar un cebador modificado que incluía dos bases extra en el extremo 3' para disminuir los heterodímeros, pero la amplificación inespecífica seguía siendo muy alta.

Cuando se realizó la qPCR del miR-126-3p se encontraron mayores niveles relativos de este en comparación al control interno, en el grupo control. Esta tendencia es consistente con la

observada en el análisis de abundancia diferencial, aunque no logró ser suficientemente alta para ser estadísticamente significativa.

Por último, se logró la validación del miR- 378a-3p dado que los valores relativos al control interno let-7b-5p dentro del grupo de TB fue significativamente más alto que en los controles. Cabe decir que se trata de un set pequeño de muestras y sería adecuado validarlo en una cohorte mayor. Sin embargo, este miARN podría ser utilizado como marcador de tuberculosis activa en suero.

6. PERSPECTIVAS

Se cuenta con una metodología de tipificación de cepas y de identificación de resistencia basado en secuenciación profunda que puede ser fácilmente transferible y aplicable en los laboratorios de biología molecular que normalmente realizan el seguimiento de los casos de TB en el país. De hecho, en el marco del proyecto FMV_3_2018_1_148367, se realizó la transferencia de esta metodología a la CHLA-EP. Las herramientas para el análisis informático pueden ser ajustadas para hilar más fino entre las cepas que circulan en la población (linaje 4), así como también seleccionar las mutaciones asociadas a resistencias más frecuentes. Los programas de análisis online son una muy buena opción para realizar informes de tipificación y de resistencia en base a la secuenciación profunda de los genes o de los genomas de interés, y no requiere entrenamiento informático. También el costo de las bibliotecas y de los kits para secuenciación han disminuido en las últimas décadas y actualmente existen nuevas plataformas como MinIon de Oxford Nanopore que requieren una infraestructura menor y disminuyen los costos por secuencia. La calidad de la asignación de bases ha mejorado notablemente y también permite secuenciar fragmentos más grandes comparado con Illumina, por lo que sería provechoso diseñar una metodología para MinIon que pueda ser aplicado como rutina en la epidemiología de la TB.

También sería deseable poder continuar con el rastreo de los linajes X en los PPL, su inserción en la comunidad y realizar un monitoreo de los casos observados de resistencia a estreptomycin. En este sentido también, se podrían secuenciar los genomas de estas cepas resistentes para intentar establecer su enlace epidemiológico.

Respecto a los marcadores moleculares de TB en suero, sería importante reproducir los resultados obtenidos en un set de muestras mayor para poder disminuir la incertidumbre y también comparar con otros grupos de pacientes graves con otras patologías (por ejemplo, neumonías comunitarias, o COVID-19) para observar si estos marcadores son específicos de TB o más bien se deben a otras características inespecíficas de la gravedad del cuadro clínico. Asimismo, podrían clasificarse a los pacientes según otros criterios, como avance o gravedad de la enfermedad, o estadía corta y estadía larga, para observar si los marcadores se mantienen entre los grupos o se distinguen nuevos candidatos. También, se podría intentar realizar una cuantificación relativa del miR-423-5p diseñando otros cebadores para disminuir la formación de heterodímeros.

Otro aspecto interesante para abordar podría ser estudiar la expresión diferencial de otros ARNs pequeños no codificantes como ARN transferencia, ARN ribosomal, ARNs largos, que fueron secuenciados en este estudio, pero que no fueron analizados.

Respecto a la respuesta del hospedero y a la búsqueda de marcadores de TB que puedan discriminar esta enfermedad entre otras e incluso entre TB activa y TB latente, se podrían

plantear otros abordajes como estudiar el metabolismo en suero de estos pacientes. La metabolómica basada en resonancia magnética nuclear permite describir perfiles de una condición dada en relación al estado metabólico y compararla con otras condiciones para discriminar qué metabolitos se encuentran en diferente proporción. Por ejemplo, es sabido que concentraciones altas de lactato es utilizado como un marcador asociado al peor pronóstico de un paciente. La metabolómica podría ayudar a descubrir nuevos marcadores moleculares y es una herramienta de muy alta sensibilidad que también se encuentra disponible en nuestro país.

La TB continúa siendo un grave problema de salud en el mundo y en nuestro país. El doble abordaje desarrollado en este trabajo, aporta a un mejor conocimiento de la enfermedad. Por un lado, desde el punto de vista de la epidemiología molecular: identificando por primera vez en nuestro país, las cepas circulantes en distintos subgrupos de enfermos con TB y estudiando, al mismo tiempo, la presencia de multiresistencia a drogas antituberculosas. Por otra parte, se desarrolló la caracterización de la respuesta del hospedero frente a la infección. El análisis de ARNs pequeños en suero, eventuales biomarcadores específicos de la enfermedad, podría aportar herramientas con valor diagnóstico y pronóstico.

7. BIBLIOGRAFÍA

- Aparicio-Puerta, E., Lebrón, R., Rueda, A., Gómez-Martín, C., Giannoukakos, S., Jaspez, D., Medina, J. M., Zubkovic, A., Jurak, I., Fromm, B., Marchal, J. A., Oliver, J., & Hackenberg, M. (2019). SRNAbench and sRNAtoolbox 2019: intuitive fast small RNA profiling and differential expression. *Nucleic Acids Research*, *47*(W1), W530–W535. <https://doi.org/10.1093/nar/gkz415>
- Bañuls, A. L., Sanou, A., van Anh, N. T., & Godreuil, S. (2015). Mycobacterium tuberculosis: Ecology and evolution of a human bacterium. In *Journal of Medical Microbiology* (Vol. 64, Issue 11, pp. 1261–1269). Microbiology Society. <https://doi.org/10.1099/jmm.0.000171>
- Barbier, M., & Wirth, T. (2016). The Evolutionary History, Demography, and Spread of the Mycobacterium tuberculosis Complex. *Microbiology Spectrum*, *4*(4). <https://doi.org/10.1128/microbiolspec.tb2-0008-2016>
- Berrada, Z. L., Lin, S. Y. G., Rodwell, T. C., Nguyen, D., Schechter, G. F., Pham, L., Janda, J. M., Elmaraachli, W., Catanzaro, A., & Desmond, E. (2016). Rifabutin and rifampin resistance levels and associated rpoB mutations in clinical isolates of Mycobacterium tuberculosis complex. *Diagnostic Microbiology and Infectious Disease*, *85*(2), 177–181. <https://doi.org/10.1016/j.diagmicrobio.2016.01.019>
- Boritsch, E. C., Khanna, V., Pawlik, A., Honoré, N., Navas, V. H., Ma, L., Bouchier, C., Seemann, T., Supply, P., Stinear, T. P., & Brosch, R. (2016). Key experimental evidence of chromosomal DNA transfer among selected tuberculosis-causing mycobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(35), 9876–9881. <https://doi.org/10.1073/pnas.1604921113>
- Brown, S. S., Chen, Y. W., Wang, M., Clipson, A., Ochoa, E., & Du, M. Q. (2017). PrimerPooler: Automated primer pooling to prepare library for targeted sequencing. *Biology Methods and Protocols*, *2*(1). <https://doi.org/10.1093/biomethods/bpx006>
- Bussi, C., & Gutierrez, M. G. (2019). Mycobacterium tuberculosis infection of host cells in space and time. In *FEMS Microbiology Reviews* (Vol. 43, Issue 4, pp. 341–361). Oxford University Press. <https://doi.org/10.1093/femsre/fuz006>
- Cabibbe, A. M., Spitaleri, A., Battaglia, S., Colman, R. E., Suresh, A., Uplekar, S., Rodwell, T. C., & Cirillo, D. M. (2020). Application of targeted next-generation sequencing assay on a portable sequencing platform for culture-free detection of drug-resistant tuberculosis from clinical samples. *Journal of Clinical Microbiology*, *58*(10). <https://doi.org/10.1128/JCM.00632-20>
- Chaiyachat, P., Chaiprasert, A., Nonghanphithak, D., Smithtikarn, S., Kamolwat, P., Punggrassami, P., Reechaipichitkul, W., Ong, R. T. H., Teo, Y. Y., & Faksri, K. (2021). Whole-genome analysis of drug-resistant Mycobacterium tuberculosis reveals novel mutations associated with fluoroquinolone resistance. *International Journal of Antimicrobial Agents*, *58*(3). <https://doi.org/10.1016/j.ijantimicag.2021.106385>
- Chen, C., Ridzon, D. A., Broomer, A. J., Zhou, Z., Lee, D. H., Nguyen, J. T., Barbisin, M., Xu, N. L., Mahuvakar, V. R., Andersen, M. R., Lao, K. Q., Livak, K. J., & Guegler, K. J. (2005). Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Research*, *33*(20). <https://doi.org/10.1093/nar/gni178>
- CHLA-EP. (2016). *Guía Nacional para el manejo de la Tuberculosis*.
- Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., 3rd, Tekaiia, F., Badcock, K., Basham, D., Brown, D.,

- Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., ... Barrell, B. G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 393(6685), 537–544. <https://doi.org/10.1038/31159>
- Coll, F., McNerney, R., Guerra-Assunção, J. A., Glynn, J. R., Perdigão, J., Viveiros, M., Portugal, I., Pain, A., Martin, N., & Clark, T. G. (2014). A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nature Communications*, 5. <https://doi.org/10.1038/ncomms5812>
- Comisionado Parlamentario Penitenciario. (2020). *Informe anual 2020*. https://parlamento.gub.uy/sites/default/files/DocumentosCPP/informe_2020_para_web.pdf
- Couvin, D., David, A., Zozio, T., & Rastogi, N. (2019). Macro-geographical specificities of the prevailing tuberculosis epidemic as seen through SITVIT2, an updated version of the *Mycobacterium tuberculosis* genotyping database. *Infection, Genetics and Evolution*, 72, 31–43. <https://doi.org/10.1016/j.meegid.2018.12.030>
- Dookie, N., Rambaran, S., Padayatchi, N., Mahomed, S., & Naidoo, K. (2018). Evolution of drug resistance in *Mycobacterium tuberculosis*: A review on the molecular determinants of resistance and implications for personalized care. *Journal of Antimicrobial Chemotherapy*, 73(5), 1138–1151. <https://doi.org/10.1093/jac/dkx506>
- Dreyer, V., Utpatel, C., Kohl, T. A., Barilar, I., Gröschel, M. I., Feuerriegel, S., & Niemann, S. (2020). Detection of low-frequency resistance-mediating SNPs in next-generation sequencing data of *Mycobacterium tuberculosis* complex strains with binoSNP. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-64708-8>
- Dubois-Camacho, K., Diaz-Jimenez, D., de la Fuente, M., Quera, R., Simian, D., Martínez, M., Landskron, G., Olivares-Morales, M., Cidlowski, J. A., Xu, X., Gao, G., Xie, J., Chnaiderman, J., Soto-Rifo, R., González, M. J., Calixto, A., & Hermoso, M. A. (2019). Inhibition of miR-378a-3p by Inflammation Enhances IL-33 Levels: A Novel Mechanism of Alarmin Modulation in Ulcerative Colitis. *Frontiers in Immunology*, 10. <https://doi.org/10.3389/fimmu.2019.02449>
- Ektefaie, Y., Dixit, A., Freschi, L., & Farhat, M. R. (2021). Globally diverse *Mycobacterium tuberculosis* resistance acquisition: a retrospective geographical and temporal analysis of whole genome sequences. *The Lancet Microbe*, 2(3), e96–e104. [https://doi.org/10.1016/S2666-5247\(20\)30195-6](https://doi.org/10.1016/S2666-5247(20)30195-6)
- Felli, N., Felicetti, F., Lustrì, A. M., Errico, M. C., Bottero, L., Cannistraci, A., de Feo, A., Petrini, M., Pedini, F., Biffoni, M., Alvino, E., Negrini, M., Ferracin, M., Mattia, G., & Carè, A. (2013). miR-126&126* Restored Expressions Play a Tumor Suppressor Role by Directly Regulating ADAM9 and MMP7 in Melanoma. *PLoS ONE*, 8(2). <https://doi.org/10.1371/journal.pone.0056824>
- Feuerriegel, S., Schleusener, V., Beckert, P., Kohl, T. A., Miotto, P., Cirillo, D. M., Cabibbe, A. M., Niemann, S., & Fellenberg, K. (2015). PhyResSE: A web tool delineating *Mycobacterium tuberculosis* antibiotic resistance and lineage from whole-genome sequencing data. *Journal of Clinical Microbiology*, 53(6), 1908–1914. <https://doi.org/10.1128/JCM.00025-15>
- Fontalvo Rivera, D., & Gómez Camargo, D. (2014). *Genes del Mycobacterium tuberculosis involucrados en la patogenicidad y resistencia a antibióticos durante la tuberculosis pulmonar y extrapulmonar*.
- Ford, C. B., Shah, R. R., Maeda, M. K., Gagneux, S., Murray, M. B., Cohen, T., Johnston, J. C., Gardy, J., Lipsitch, M., & Fortune, S. M. (2013). *Mycobacterium tuberculosis* mutation rate

- estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nature Genetics*, 45(7), 784–790. <https://doi.org/10.1038/ng.2656>
- Gagneux, S. (2018). Ecology and evolution of Mycobacterium tuberculosis. In *Nature Reviews Microbiology* (Vol. 16, Issue 4, pp. 202–213). Nature Publishing Group. <https://doi.org/10.1038/nrmicro.2018.8>
- Geneva: World Health Organization. (2021). *Global Tuberculosis Report 2021*. <http://apps.who.int/bookorders>.
- Hernández-Pando, R., López, B., Aguilar, D., Orozco, H., Burger, M., Espitia, C., Ritacco, V., Barrera, L., Kremer, K., Hernandez-Pando, R., Huygen, K., & van Soolingen, D. (2003). A marked difference in pathogenesis and immune response induced by different Mycobacterium tuberculosis genotypes. In *Clin Exp Immunol* (Vol. 133).
- Huang, W. L., Chi, T. L., Wu, M. H., & Jou, R. (2011). Performance assessment of the GenoType MTBDRsl test and DNA sequencing for detection of second-line and ethambutol drug resistance among patients infected with multidrug-resistant mycobacterium tuberculosis. *Journal of Clinical Microbiology*, 49(7), 2502–2508. <https://doi.org/10.1128/JCM.00197-11>
- Hurtado, J., Coitinho, C., Nin, N., Buroni, M., Hurtado, F. J., Robello, C., & Greif, G. (2021). Clinical and epidemiological features of tuberculosis isolated from critically ill patients. *Revista Argentina de Microbiologia*. <https://doi.org/10.1016/j.ram.2021.02.011>
- Jena, L., Waghmare, P., Kashikar, S., Kumar, S., & Harinath, B. C. (2014). Computational approach to understanding the mechanism of action of isoniazid, an anti-TB drug. *International Journal of Mycobacteriology*, 3(4), 276–282. <https://doi.org/10.1016/j.ijmyco.2014.08.003>
- Johnstone-Robertson, S., Welte, A., Bekker, L.-G., Tutu, D., & Wood, R. (2017). *Tuberculosis in a South African prison-a transmission modelling analysis*.
- Joshi NA, Fass JN. (2011). SickLe: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at <https://github.com/najoshi/sickle>.
- Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S., Bunschoten, A., Molhuizen, H., Shaw, R., Goyal, M., & van Embden, J. (1997). Simultaneous Detection and Strain Differentiation of Mycobacterium tuberculosis for Diagnosis and Epidemiology. In *JOURNAL OF CLINICAL MICROBIOLOGY* (Vol. 35, Issue 4).
- Kendall, E. A., Fofana, M. O., & Dowdy, D. W. (2015). Burden of transmitted multidrug resistance in epidemics of tuberculosis: A transmission modelling analysis. *The Lancet Respiratory Medicine*, 3(12), 963–972. [https://doi.org/10.1016/S2213-2600\(15\)00458-0](https://doi.org/10.1016/S2213-2600(15)00458-0)
- Kim, J. K., Kim, T. S., Basu, J., & Jo, E. K. (2017). MicroRNA in innate immunity and autophagy during mycobacterial infection. In *Cellular Microbiology* (Vol. 19, Issue 1). Blackwell Publishing Ltd. <https://doi.org/10.1111/cmi.12687>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Lanoix, J. P., Gaudry, S., Flicoteaux, R., Ruimy, R., & Wolff, M. (2014). Tuberculosis in the intensive care unit: A descriptive analysis in a low-burden country. *International Journal of Tuberculosis and Lung Disease*, 18(5). <https://doi.org/10.5588/ijtld.13.0901>
- Lanzas, F., Karakousis, P. C., Sacchetti, J. C., & Ioerger, T. R. (2013). Multidrug-resistant tuberculosis in Panama is driven by clonal expansion of a multidrug-resistant mycobacterium tuberculosis strain related to the KZN extensively drug-resistant m.

- tuberculosis strain from South Africa. *Journal of Clinical Microbiology*, 51(10), 3277–3285. <https://doi.org/10.1128/JCM.01122-13>
- Lempens, P., Meehan, C. J., Vandellannoote, K., Fissette, K., de Rijk, P., van Deun, A., Rigouts, L., & de Jong, B. C. (2018). Isoniazid resistance levels of Mycobacterium tuberculosis can largely be predicted by high-confidence resistance-conferring mutations. *Scientific Reports*, 8(1). <https://doi.org/10.1038/s41598-018-21378-x>
- Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. <http://arxiv.org/abs/1303.3997>
- Lipin, M. Y., Stepanshina, V. N., Shemyakin, I. G., & Shinnick, T. M. (2007). Association of specific mutations in katG, rpoB, rps L and rrs genes with spoligotypes of multidrug-resistant Mycobacterium tuberculosis isolates in Russia. *Clinical Microbiology and Infection*, 13(6), 620–626. <https://doi.org/10.1111/j.1469-0691.2007.01711.x>
- Liu, Q., Zuo, T., Xu, P., Jiang, Q., Wu, J., Gan, M., Yang, C., Prakash, R., Zhu, G., Takiff, H. E., & Gao, Q. (2018). Have compensatory mutations facilitated the current epidemic of multidrug-resistant tuberculosis? *Emerging Microbes and Infections*, 7(1). <https://doi.org/10.1038/s41426-018-0101-6>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12). <https://doi.org/10.1186/s13059-014-0550-8>
- Maudet, C., Mano, M., & Eulalio, A. (2014). MicroRNAs in the interaction between host and bacterial pathogens. In *FEBS Letters* (Vol. 588, Issue 22, pp. 4140–4147). Elsevier B.V. <https://doi.org/10.1016/j.febslet.2014.08.002>
- Max, K. E. A., Bertram, K., Akat, K. M., Bogardus, K. A., Li, J., Morozov, P., Ben-Dov, I. Z., Li, X., Weiss, Z. R., Azizian, A., Sopeyin, A., Diacovo, T. G., Adamidi, C., Williams, Z., & Tuschl, T. (2018). Human plasma and serum extracellular small RNA reference profiles and their clinical utility. *Proceedings of the National Academy of Sciences of the United States of America*, 115(23), E5334–E5343. <https://doi.org/10.1073/pnas.1714397115>
- Meehan, C. J., Moris, P., Kohl, T. A., Pečerska, J., Akter, S., Merker, M., Utpatel, C., Beckert, P., Gehre, F., Lempens, P., Stadler, T., Kaswa, M. K., Kühnert, D., Niemann, S., & de Jong, B. C. (2018). The relationship between transmission time and clustering methods in Mycobacterium tuberculosis epidemiology. *EBioMedicine*, 37, 410–416. <https://doi.org/10.1016/j.ebiom.2018.10.013>
- Muthu, V., Agarwal, R., Dhooria, S., Aggarwal, A. N., Behera, D., & Sehgal, I. S. (2018). Outcome of critically ill subjects with tuberculosis: Systematic review and meta-analysis. *Respiratory Care*, 63(12), 1541–1554. <https://doi.org/10.4187/respcare.06190>
- Orgeur, M., & Brosch, R. (2018). Evolution of virulence in the Mycobacterium tuberculosis complex. In *Current Opinion in Microbiology* (Vol. 41, pp. 68–75). Elsevier Ltd. <https://doi.org/10.1016/j.mib.2017.11.021>
- Palomino, J. C., & Martin, A. (2014). Drug resistance mechanisms in Mycobacterium tuberculosis. In *Antibiotics* (Vol. 3, Issue 3, pp. 317–340). MDPI AG. <https://doi.org/10.3390/antibiotics3030317>
- Rajaram, M. V. S., Ni, B., Morris, J. D., Brooks, M. N., Carlson, T. K., Bakthavachalu, B., Schoenberg, D. R., Torrelles, J. B., & Schlesinger, L. S. (2011). Mycobacterium tuberculosis lipomannan blocks TNF biosynthesis by regulating macrophage MAPK-activated protein kinase 2 (MK2) and microRNA miR-125b. *Proceedings of the National Academy of Sciences of the United States of America*, 108(42), 17408–17413. <https://doi.org/10.1073/pnas.1112660108>

- Rajendran, M. R., Zaki, R. A., Bakar, Z. A., & Aghamohammadi, N. (2020). *Prevalence of multidrug resistant tuberculosis among TB patients in Malaysia*. <https://doi.org/10.21203/rs.3.rs-48963/v1>
- Ruiz-Tagle, C., Naves, R., & Balcells, M. E. (2020). *Unraveling the Role of MicroRNAs in Mycobacterium tuberculosis Infection and Disease: Advances and Pitfalls*. <https://doi.org/10.1128/IAI>
- Sakamoto, K. (2012). The Pathology of Mycobacterium tuberculosis Infection. In *Veterinary Pathology* (Vol. 49, Issue 3, pp. 423–439). <https://doi.org/10.1177/0300985811429313>
- Schmittgen, T. D., & Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*, 3(6), 1101–1108. <https://doi.org/10.1038/nprot.2008.73>
- Singh, R., Dwivedi, S. P., Gaharwar, U. S., Meena, R., Rajamani, P., & Prasad, T. (2020). Recent updates on drug resistance in Mycobacterium tuberculosis. In *Journal of Applied Microbiology* (Vol. 128, Issue 6, pp. 1547–1567). Blackwell Publishing Ltd. <https://doi.org/10.1111/jam.14478>
- Sinigaglia, A., Peta, E., Riccetti, S., Venkateswaran, S., Manganelli, R., & Barzon, L. (2020). Tuberculosis-Associated MicroRNAs: From Pathogenesis to Disease Biomarkers. In *Cells* (Vol. 9, Issue 10). NLM (Medline). <https://doi.org/10.3390/cells9102160>
- Smittipat, N., Juthayothin, T., Billamas, P., Jaitrong, S., Rukseree, K., Dokladda, K., Chaiyasirinroje, B., Disratthakit, A., Chairprasert, A., Mahasirimongkol, S., Yanai, H., Yamada, N., Tokunaga, K., & Palittapongarnpim, P. (2016). Mutations in rrs, rpsL and gidB in streptomycin-resistant Mycobacterium tuberculosis isolates from Thailand. *Journal of Global Antimicrobial Resistance*, 4, 5–10. <https://doi.org/10.1016/j.jgar.2015.11.009>
- Supply, P., Magdalena, J., Himpens, S., & Locht, C. (1997). *Identification of novel intergenic repetitive units in a mycobacterial two-component system operon*.
- Syeda, Z. A., Langden, S. S. S., Munkhzul, C., Lee, M., & Song, S. J. (2020). Regulatory mechanism of microrna expression in cancer. In *International Journal of Molecular Sciences* (Vol. 21, Issue 5). MDPI AG. <https://doi.org/10.3390/ijms21051723>
- Tiberi, S., du Plessis, N., Walzl, G., Vjecha, M. J., Rao, M., Ntoumi, F., Mfinanga, S., Kapata, N., Mwaba, P., McHugh, T. D., Ippolito, G., Migliori, G. B., Maeurer, M. J., & Zumla, A. (2018). Tuberculosis: progress and advances in development of new drugs, treatment regimens, and host-directed therapies. In *The Lancet Infectious Diseases* (Vol. 18, Issue 7, pp. e183–e198). Lancet Publishing Group. [https://doi.org/10.1016/S1473-3099\(18\)30110-5](https://doi.org/10.1016/S1473-3099(18)30110-5)
- Tosar, J. P., Gámbaro, F., Sanguinetti, J., Bonilla, B., Witwer, K. W., & Cayota, A. (2015). Assessment of small RNA sorting into different extracellular fractions revealed by high-throughput sequencing of breast cell lines. *Nucleic Acids Research*, 43(11), 5601–5616. <https://doi.org/10.1093/nar/gkv432>
- Tu, H., Yang, S., Jiang, T., Wei, L., Shi, L., Liu, C., Wang, C., Huang, H., Hu, Y., Chen, Z., Chen, J., Li, Z., & Li, J. (2019). Elevated pulmonary tuberculosis biomarker miR-423-5p plays critical role in the occurrence of active TB by inhibiting autophagosome-lysosome fusion. *Emerging Microbes and Infections*, 8(1), 448–460. <https://doi.org/10.1080/22221751.2019.1590129>
- van Embden, J. D. A., Cave, M. D., Crawford, J. T., Dale, J. W., Eisenach, K. D., Gicquel, B., Hermans, P., Martin, C., Mcadam, R., Shinnick, T. M., & Small, P. M. (1993). Strain Identification of Mycobacterium tuberculosis by DNA Fingerprinting: Recommendations for a Standardized Methodology. In *JOURNAL OF CLINICAL MICROBIOLOGY*.
- Walter, K. S., Martinez, L., Arakaki-Sanchez, D., Sequera, V. G., Estigarribia Sanabria, G., Cohen, T., Ko, A. I., García-Basteiro, A. L., Rueda, Z. V., López-Olarte, R. A., Espinal, M.

- A., Croda, J., & Andrews, J. R. (2021). The escalating tuberculosis crisis in central and South American prisons. In *The Lancet* (Vol. 397, Issue 10284, pp. 1591–1596). Elsevier B.V. [https://doi.org/10.1016/S0140-6736\(20\)32578-2](https://doi.org/10.1016/S0140-6736(20)32578-2)
- Walzl, G., McNerney, R., du Plessis, N., Bates, M., McHugh, T. D., Chegou, N. N., & Zumla, A. (2018). Tuberculosis: advances and challenges in development of new diagnostics and biomarkers. In *The Lancet Infectious Diseases* (Vol. 18, Issue 7, pp. e199–e210). Lancet Publishing Group. [https://doi.org/10.1016/S1473-3099\(18\)30111-7](https://doi.org/10.1016/S1473-3099(18)30111-7)
- Warren, J. L., Grandjean, L., Moore, D. A. J., Lithgow, A., Coronel, J., Sheen, P., Zelner, J. L., Andrews, J. R., & Cohen, T. (2018). Investigating spillover of multidrug-resistant tuberculosis from a prison: A spatial and molecular epidemiological analysis. *BMC Medicine*, 16(1). <https://doi.org/10.1186/s12916-018-1111-x>
- Weerasuriya, C. K., Clark, R. A., White, R. G., & Harris, R. C. (2020). New tuberculosis vaccines: advances in clinical development and modelling. In *Journal of Internal Medicine* (Vol. 288, Issue 6, pp. 661–681). Blackwell Publishing Ltd. <https://doi.org/10.1111/joim.13197>
- Zhang, H., Sun, Z., Wei, W., Liu, Z., Fleming, J., Zhang, S., Lin, N., Wang, M., Chen, M., Xu, Y., Zhou, J., Li, C., Bi, L., & Zhou, G. (2014). Identification of serum microRNA biomarkers for tuberculosis using RNA-seq. *PLoS ONE*, 9(2). <https://doi.org/10.1371/journal.pone.0088909>

ANEXO I

Tabla Suplementaria 1. Cobertura del pool-L contra el genoma de referencia.

	L1	L2	L3	L4	L5	L 6	L7	X	S	LAM	Haarlem
ID	hemL	glnH	drxA	lpqQ	hisD	pykA	glmU	ctpH	mycP3	rpoC	fcoT
1	0,0	1404,2	88,7	161,5	0,0	224,5	53,3	0,0	29,9	101,2	0,0
2	1,9	8778,1	1312,3	2967,6	2,0	2099,8	725,1	16,5	1791,8	1396,2	0,0
3	1,4	5906,7	240,8	627,3	0,0	387,4	164,6	18,7	380,8	150,6	0,0
4	0,0	2811,2	188,8	411,7	0,0	95,8	88,3	0,3	237,2	152,9	0,0
5	2,0	4045,7	367,0	437,0	0,7	437,1	265,7	4,1	358,4	317,7	0,0
6	0,5	1137,3	116,5	137,5	0,0	125,1	85,0	4,1	103,4	76,8	0,0
7	138,2	36950,3	6461,5	5922,2	17,8	6006,2	6092,5	63,6	5948,4	6041,6	0,3
8	-	-	-	-	-	-	-	-	-	-	-
9	10,0	8438,6	1921,1	5854,1	6,7	261,1	2863,8	58,4	440,9	64,3	0,0
10	7,2	6228,1	225,4	925,3	9,4	63,6	483,5	20,6	53,2	15,7	0,1
11	22,7	7504,4	2878,1	5834,1	48,6	2856,0	2288,6	112,5	4011,9	216,9	0,0
12	0,5	2103,7	349,1	310,0	0,0	136,7	243,4	0,0	203,3	199,0	0,0
13	0,1	3668,5	467,4	329,2	0,0	463,2	351,3	0,0	156,5	308,9	0,0
14	2,9	4058,4	458,9	569,2	0,0	528,9	354,5	0,7	329,3	355,2	0,0
15	6,7	10113,0	1898,7	2635,6	0,4	2782,8	1185,3	3,0	1559,0	2527,5	0,4
16	0,0	1332,0	324,8	313,9	0,0	16,2	201,2	0,0	46,7	188,2	0,0
17	0,2	4165,0	664,8	816,7	0,0	94,8	363,0	0,6	464,2	437,7	0,0
18	3,7	2450,8	531,8	537,9	0,4	74,5	306,8	0,5	257,1	295,6	0,0
19	46,4	11243,1	4673,5	4922,2	0,0	176,1	2912,7	0,0	951,5	2044,9	0,0
20	0,5	3010,2	427,5	520,4	0,0	140,9	236,1	1,2	268,5	297,4	0,0
21	1,5	1882,3	453,0	421,5	0,4	35,3	286,3	0,3	120,7	176,1	0,0
22	1,4	3625,6	638,6	609,4	0,0	121,6	421,3	0,0	366,2	383,8	0,0
23	5,1	7248,1	1163,2	5397,9	5,1	1050,2	834,3	14,7	1930,7	353,2	0,0
24	1,7	7636,9	1644,6	1405,6	0,0	560,5	1100,0	0,3	683,8	574,2	0,0
25	0,0	2245,4	223,9	257,6	0,0	232,3	209,9	0,0	141,9	162,5	0,0
26	0,9	3797,7	460,7	1331,7	0,0	121,3	289,4	3,0	406,6	105,5	0,4
27	21,8	8002,4	2776,3	6442,5	17,1	674,3	3415,5	52,1	3509,4	108,5	0,3
28	224,3	24956,1	6211,6	6721,3	109,0	5371,7	6368,0	196,2	6021,3	2152,9	14,6
29	2,8	1618,9	679,7	673,7	0,0	3,1	355,1	0,0	13,6	28,9	0,0
30	8,4	7780,6	2066,6	2015,4	0,0	94,1	1455,4	1,3	455,0	1066,9	0,0
31	7,3	7188,1	511,4	3135,3	5,9	98,3	715,7	35,0	210,3	10,9	0,3
32	3,5	4350,0	1438,9	1568,3	0,0	10,2	807,5	0,0	74,4	192,7	0,0
33	0,9	1296,1	297,5	279,9	0,0	32,8	187,5	0,0	128,0	185,2	0,0
34	0,2	4856,0	1256,3	1456,4	0,0	38,6	705,7	0,4	232,4	501,2	0,4
35	2,2	2833,7	487,2	485,3	0,0	53,3	353,9	0,0	273,9	297,8	0,0
36	5,7	3925,9	664,7	564,5	0,0	98,1	482,1	0,0	341,7	378,9	0,3
37	2,4	1681,5	317,4	293,0	0,0	0,8	223,7	0,0	8,2	16,3	0,0
38	7,5	689,9	18,5	23,9	5,4	15,3	43,3	13,2	13,8	7,2	0,3
39	2,1	6759,6	1285,1	5637,8	3,6	105,5	1069,8	9,7	305,0	12,9	0,0
40	11,5	5167,3	1067,9	989,4	0,4	145,2	742,0	0,0	488,9	624,8	0,4

41	1,6	1186,9	121,3	118,3	0,2	130,4	96,9	3,7	78,1	73,4	0,0
42	1,0	3808,6	520,5	602,2	0,0	247,0	357,3	2,6	458,5	371,8	0,0
43	2,6	2588,9	390,6	395,7	0,0	82,6	258,9	0,5	223,0	270,4	0,0
44	5,2	2819,0	396,7	453,4	1,2	328,7	338,4	3,2	310,0	296,6	0,0
45	0,0	1317,0	266,0	396,4	0,0	26,2	162,1	0,0	181,7	139,6	0,0
46	1,1	1216,6	166,5	136,1	0,0	46,6	108,6	0,4	93,9	125,6	0,4
47	0,3	7979,6	2540,5	1990,7	0,7	943,5	1820,6	4,0	987,5	1017,5	0,0
H37	1,8	4667,7	549,3	501,6	0,0	21,6	376,4	0,0	148,9	417,8	0,0

Tabla Suplementaria 2. Cobertura del pool-R contra el genoma de referencia.

	gyrB	gyrA	rpoB	rrS	inhA	katG	pncA	eis	embB
1	0,2	18,8	115,8	115,3	200,3	60,4	552,5	151,5	0,4
2	0,3	0,2	47,1	0,1	2763,8	1,6	7042,8	277,3	0,2
3	1,7	162,9	477,3	731,6	963,3	111,9	3237,9	339,9	0,1
4	0,5	30,6	194,7	497,6	446,4	17,9	2391,2	3,9	0,1
5	5,0	130,4	290,4	452,1	413,9	98,9	1236,9	206,8	0,5
6	1,9	16,0	84,7	151,5	132,6	15,3	395,4	8,2	0,0
7	43,8	568,3	2198,3	3502,5	6847,1	1505,8	7350,4	3484,1	3,6
8	-	-	-	-	-	-	-	-	-
9	1,7	4,2	5,6	2,1	10,1	2,2	13,6	10,3	0,9
10	1,0	2,4	5,9	1,3	14,2	4,5	7488,9	6,8	1,3
11	0,5	1,5	15,8	1,0	103,8	1,7	23,6	1,9	0,7
12	0,3	24,7	135,1	120,7	171,1	72,1	520,4	114,5	0,6
13	3,4	79,7	266,3	270,0	361,6	187,9	1088,0	224,2	3,2
14	5,6	78,2	283,9	715,2	438,4	67,6	1557,5	72,1	0,2
15	0,0	0,2	44,0	0,5	897,3	1,7	6734,5	10,8	0,1
16	4,9	3,1	59,2	326,7	91,5	37,7	791,9	0,2	0,0
17	6,1	36,5	209,2	647,6	543,3	54,7	2500,9	13,0	0,0
18	1,1	23,0	120,1	201,7	332,0	29,6	1054,2	64,8	0,0
19	0,4	0,7	187,3	501,8	1667,9	47,3	6877,2	24,1	0,1
20	5,6	33,5	171,7	529,8	412,0	38,1	1786,4	6,7	0,0
21	10,1	8,3	104,9	454,6	260,6	67,8	959,9	1,8	0,1
22	4,1	68,8	200,2	330,9	376,9	84,1	1457,2	122,9	0,1
23	0,4	0,9	16,4	1,8	4061,4	4,1	7490,4	26,9	0,8
24	1,3	52,1	441,9	475,3	1315,4	302,5	4292,2	657,2	0,0
25	0,6	24,6	169,5	160,3	225,6	54,8	609,9	115,3	0,0
26	0,2	0,5	81,3	10,9	1350,9	12,4	609,9	29,5	0,0
27	0,6	3,4	4,9	2,6	5,8	1,5	6312,4	7,2	1,3
28	10,1	18,8	27,7	14,0	33,0	14,2	10,6	31,5	12,5
29	18,8	2,5	47,4	769,5	25,6	89,8	24,9	0,0	0,0
30	0,9	1,0	241,7	259,8	1081,1	6,7	908,3	36,6	0,2
31	1,4	3,8	12,3	2,3	77,7	6,3	4679,0	15,7	2,7
32	18,4	26,4	167,7	1215,3	338,3	102,8	7485,9	4,0	0,0
33	7,2	0,8	46,4	365,0	30,6	61,2	3154,8	0,3	0,0

34	0,5	0,6	164,2	426,2	760,3	86,0	823,2	18,2	0,1
35	1,0	31,8	169,9	212,8	305,9	66,9	4184,6	118,3	0,1
36	2,9	58,1	255,0	355,0	439,3	112,3	1009,1	179,6	0,0
37	1,0	28,0	136,5	145,4	153,0	57,6	493,2	97,0	0,1
38	0,8	10,4	27,7	1,7	1174,2	86,5	7534,2	17,4	1,8
39	0,3	0,7	2,3	0,7	28,7	0,8	3,9	1,7	0,4
40	18,2	73,2	266,4	838,1	736,4	117,1	2168,2	8,9	0,0
41	2,6	9,3	91,1	153,1	137,5	26,4	413,3	3,7	0,0
42	1,4	61,7	228,9	337,1	423,0	51,5	1576,8	126,6	0,1
43	1,8	48,8	163,7	244,7	286,3	48,6	901,8	105,2	0,1
44	7,2	48,0	174,7	411,0	436,8	67,7	1092,4	27,8	0,0
45	5,7	1,2	51,9	435,4	53,8	66,4	1362,9	0,3	0,0
46	6,0	6,4	82,2	221,8	93,7	30,5	355,0	0,6	0,0
47	2,3	38,3	229,2	710,8	1076,8	301,8	4825,0	126,6	0,0
H37	2,4	95,2	339,6	496,2	442,1	104,4	1114,7	216,7	0,1

Tabla Suplementaria 3. SNPs detectados en cada una de las muestra, para los genes de resistencia

<i>ID</i>	<i>Gen</i>	<i>Posición SNP</i>	<i>Observaciones</i>	<i>Bibliografía reportada</i>
1	gyrA	7362	9 reads 100% G/C	No reportada
	gyrA	7585	27 reads 100% G/C	No reportada
	gyrA	9304	16 reads 100% G/A	No reportada
	rrS	1472337	95 reads 100% C/T	Coll, et al. 2014
2	katG	2154279	7 reads 100% G/A	No reportada
3	gyrA	7362	42 reads 100% G/C	No reportada
	gyrA	7585	109 reads 100% G/C	No reportada
	gyrA	9304	66 reads 100% G/A	No reportada
	pncA	2289017	4098 reads 100% A/G	No reportada
	rrS	1473314	4370 rads 64% A/C	No reportada
4	gyrA	7362	5 reads 100% G/C	No reportada
	gyrA	7585	15 reads 100% G/C	No reportada
	gyrA	8040	28 reads 97% G/A	No reportada
	gyrA	9304	22 reads 100% G/A	No reportada
5	gyrA	7362	69 reads 100% G/C	No reportada
	gyrA	7585	113 reads 100% G/C	No reportada
	gyrA	9304	66 reads 100% G/A	No reportada
	gyrB	6438	7 reads 100% C/G	No reportada
	katG	2154015	317 reads 99% C/G	No reportada
	rpoB	760115	169 reads 100% C/T	No reportada
6	gyrA	7362	6 reads 100% G/C	No reportada
	gyrA	7585	8 reads 100% G/C	No reportada
	gyrA	9304	11 rads 100% G/A	No reportada
	pncA	2289017	513 reads 100% A/G	No reportada
7	gyrA	7362	681 reads 100% G/C	No reportada
	gyrA	7585	985 reads 100% G/C	No reportada
	gyrA	9304	436 reads 100% G/A	No reportada
	pncA	2289017	34573 reads 100% A/G	No reportada
9	gyrA	7362	4 reads 100% G/C	No reportada
	gyrA	7585	7 reads 100% G/C	No reportada
	gyrA	9304	6 reads 100% G/A	No reportada
	rpoB	760115	9 reads 89% C/T	No reportada
	rrS	1473314	11 reads 73% A/C	No reportada
10	gyrA	7362	3 reads 100% G/C	No reportada
	pncA	2289017	141266 reads 100% A/G	No reportada
11	gyrA	7362	7 reads 100% G/C	No reportada
	gyrA	7585	12 reads 100% G/C	No reportada
12	gyrA	7362	22 reads 100% G/C	No reportada
13	gyrA	7362	55 reads 100% G/C	No reportada

	gyrA	7585	69 reads 100% G/C	No reportada
	gyrA	9304	55 reads 100% G/A	No reportada
	rpoB	760115	166 reads 100% C/T	No reportada
14	gyrA	7362	27 reads 100% G/C	No reportada
15	pncA	2289017	13614 reads 100% A/G	No reportada
16	rpoB	760115	25 reads 100% C/T	No reportada
17	gyrA	7362	16 reads 100% G/C	No reportada
	gyrA	7585	17 reads 100% G/C	No reportada
	gyrA	9304	18 reads 100% G/A	No reportada
	rpoB	760115	210 reads 100% C/T	No reportada
18	gyrA	7362	12 reads 100% G/C	No reportada
	gyrA	7585	27 reads 100% G/C	No reportada
	gyrA	9304	10 reads 100% G/A	No reportada
	pncA	2289017	1276 reads 100% A/G	No reportada
	rpoB	760106	172 reads 99% G/A	No reportada
	rpoB	760115	171 reads 99% C/T	No reportada
19	rpoB	760106	172 reads 99% G/A	No reportada
	rpoB	760115	171 reads 99% C/T	No reportada
20	gyrA	7362	9 reads 100% G/C	No reportada
	gyrA	7585	27 reads 100% G/C	No reportada
	gyrA	9304	12 reads 100% G/A	No reportada
	pncA	2289017	2184 reads 100% A/G	No reportada
21	gyrA	7585	12 reads 100% G/C	No reportada
	gyrA	9304	11 reads 100% G/A	No reportada
	pncA	2289017	1173 reads 100% A/G	No reportada
22	gyrA	7362	59 reads 100% G/C	No reportada
	gyrA	7585	71 reads 100% G/C	No reportada
	gyrA	9304	36 reads 97% G/A	No reportada
	gyrB	6034	5 reads 100% C/T	No reportada
23	gyrA	9304	6 reads 100% G/A	No reportada
24	gyrA	7362	48 reads 100% G/C	No reportada
	gyrA	7585	58 reads 100% G/C	No reportada
	gyrA	9304	34 reads 100% G/A	No reportada
	pncA	2289017	5586 reads 100% A/G	No reportada
25	gyrA	7362	18 reads 100% G/C	No reportada
	gyrA	7585	29 reads 100% G/C	No reportada
	gyrA	9304	22 reads 100% G/A	No reportada
	katG	2154279	255 reads 100% G/A	No reportada
26	pncA	2289017	9750 reads 100% A/G	No reportada
27	rrS	1473314	8 reads 88% A/C	No reportada
28	eis	2713795	19 reads 100% C/T	No reportada
	gyrA	7362	37 reads 100% G/C	No reportada
	gyrA	7585	24 reads 100% G/C	No reportada
	gyrA	9304	18 reads 100% G/A	No reportada
	rpoB	760115	35 reads 100% C/T	No reportada
	rrS	1471659	10 reads 100% C/T	No reportada
29	pncA	2289017	1090 reads 100% A/G	No reportada
30	pncA	2289017	5875 readas 100% A/G	No reportada

31	rrS	1473314	20 reads 75% A/C	No reportada
32	gyrA	7362	5 reads 100% G/C	No reportada
	gyrA	7585	14 reads 100% G/C	No reportada
	gyrA	9304	13 reads 100% G/A	No reportada
	rpoB	764995	153 reads 100% C/G	No reportada
	rrS	1472337	859 reads 100% C/T	Coll, et al. 2014
33	-			
34	-			
35	gyrA	7362	28 reads 100% G/C	No reportada
	gyrA	7585	39 reads 100% G/C	No reportada
	gyrA	8040	29 reads 100% G/A	No reportada
	gyrA	9304	16 reads 100% G/A	No reportada
	rpoB	760852	168 reads 99% T/C	No reportada
36	gyrA	7362	57 reads 100% G/C	No reportada
	gyrA	7585	60 reads 100% G/C	No reportada
	gyrA	9304	38 reads 100% G/A	No reportada
	rpoB	760115	185 reads 100% C/T	No reportada
37	gyrA	7362	23 reads 100% G/C	No reportada
	gyrA	7585	27 reads 100% G/C	No reportada
	gyrA	9304	12 reads 100% G/A	No reportada
	pncA	2289017	533 reads 100% A/G	No reportada
38	gyrA	7362	58 reads 100% G/C	No reportada
	gyrA	7585	88 reads 100% G/C	No reportada
	pncA	2289017	704834 reads 100% A/G	No reportada
39	-			
40	gyrA	7362	64 reads 100% G/C	No reportada
	gyrA	7585	95 reads 100% G/C	No reportada
	gyrA	9304	64 reads 100% G/A	No reportada
	pncA	2289017	2711 reads 100% A/G	No reportada
41	gyrA	7362	6 reads 100% G/C	No reportada
42	gyrA	7362	35 reads 100% G/C	No reportada
	gyrA	7585	45 reads 100% G/C	No reportada
	gyrA	9304	41 reads 100% G/A	No reportada
	rpoB	760115	181 reads 100% C/T	No reportada
43	gyrA	7362	38 reads 100% G/C	No reportada
	gyrA	7585	66 reads 100% G/C	No reportada
	gyrA	8040	44 reads 100% G/A	No reportada
	gyrA	9304	28 reads 100% G/A	No reportada
	rpoB	760629	202 reads 100% C/G	No reportada
44	gyrA	7362	42 reads 100% G/C	No reportada
	gyrA	7585	63 reads 100% G/C	No reportada
	gyrA	8040	45 reads 100% G/A	No reportada
	gyrA	9304	32 reads 100% G/A	No reportada
45	rpoB	764995	123 reads 99 % C/G	No reportada
	rrS	1472337	248 reads 100% C/T	Coll, et al. 2014
46	gyrA	7362	3 reads 100% G/C	No reportada

	gyrA	7585	3 reads 100% G/C	No reportada
	gyrA	9304	10 reads 100% G/A	No reportada
	pncA	2289017	434 reads 100% A/G	No reportada
47	gyrA	7362	71 reads 100% G/C	No reportada
	gyrA	7585	95 reads 100% G/C	No reportada
	gyrA	9304	26 reads 100% G/A	No reportada
	pncA	2289017	5992 reads 100% A/G	No reportada