



EPIDEMIOLOGÍA GENÓMICA DE LA TUBERCULOSIS EN LA COMUNIDAD VALENCIANA

CARLOS FRANCISCO MADRAZO MOYA

TESIS DOCTORAL NOV-2024

EPIDEMIOLOGÍA GENÓMICA DE LA TUBERCULOSIS EN LA COMUNIDAD VALENCIANA

Codirectores:
Iñaki Comas Espadas
Mariana Gabriela López Rodríguez

**CARLOS FRANCISCO
MADRAZO MOYA**

TESIS DOCTORAL NOV-2024
PROGRAMA DE DOCTORADO EN
BIOMEDICINA Y BIOTECNOLOGÍA



VNIVERSITAT
ID VALÈNCIA



VNIVERSITAT
ID VALÈNCIA



VNIVERSITAT DE VALÈNCIA

Epidemiología Genómica de la Tuberculosis en la Comunidad Valenciana

Autor: Carlos Francisco Madrazo Moya

Codirectores

Iñaki Comas Espadas

Mariana Gabriela López Rodríguez

Tesis Doctoral

Programa de Doctorado en Biomedicina y Biotecnología

Valencia, noviembre 2024

Agradecimientos

En primer lugar, quiero expresar mi más sincero agradecimiento a Mariana e Iñaki, mis directores de tesis, por haberme acompañado y desafiado a lo largo de estos años. A Iñaki, gracias por darme la oportunidad de unirme al grupo, por retarme con tus ideas, y por tu constante disposición. Siempre tuviste la puerta abierta —y no solo en el sentido literal—, lo que me dio la confianza para explorar un mundo tan fascinante como exigente: la ciencia.

A Mariana, gracias por tu paciencia inagotable, por saber hacer fácil lo difícil, y por esas largas conversaciones en las que muchas veces dimos vueltas, pero siempre llegamos a lugares más claros y profundos. Ambos lograron sacar lo mejor de mí, empujándome al límite solo para mostrarme que ese límite era más bien una frontera imaginaria. Me enseñaron a mirar con más atención, a cuestionar con rigor y a encontrar sentido incluso en la incertidumbre.

Quiero agradecer también, con especial cariño, a Manoli, Álvaro, Galo, Miguel e Irving, por haber estado presentes en diferentes momentos del camino. Y a mis compañeros de laboratorio, gracias por su compromiso, por el esfuerzo compartido, por la dedicación a las muestras, los datos y los detalles —que nunca son menores—, y por haber creado un ambiente donde el trabajo diario se volvió más llevadero, más humano, y, por momentos, incluso divertido.

A mi madre y a mi hermana, gracias por creer en mí sin condiciones, por animarme a dar un paso más incluso cuando no tenía fuerzas, y por su apoyo constante, especialmente durante este último año. La distancia no impidió que sintiera su presencia y su cariño todos los días.

A mis amigos —que, afortunadamente, son muchos y muy distintos—, gracias por acompañarme en este viaje. Por los momentos compartidos, por las conversaciones sinceras y por ser ese espacio de equilibrio y alegría que tanto necesité en esta etapa.

Agradezco también al CONACYT y al gobierno de México por brindarme la oportunidad de realizar este doctorado en el extranjero. Ese apoyo fue fundamental para que este sueño se hiciera realidad.

También quiero expresar mi gratitud a quienes estuvieron antes que nosotros: a todas las personas que, movidas por la curiosidad, el asombro o la necesidad, dedicaron su vida a observar, preguntar, descubrir y construir conocimiento. Gracias a su esfuerzo acumulado a lo largo de generaciones, hoy podemos asomarnos un poco más al misterio de la vida y avanzar, paso a paso, en el complejo camino de la ciencia.

Y finalmente, gracias a todas las personas que, de una u otra manera, formaron parte de este camino. A quienes me apoyaron, me desafiaron, me inspiraron o simplemente caminaron conmigo por un tramo. Este trabajo no es solo fruto del esfuerzo personal, sino también del acompañamiento, la confianza y el aliento de muchos. Gracias por ser parte de esta historia.

“During the past years, clinicians, pathologist, bacteriologists, immunologists, roentgenologists, epidemiologists and genetics have devoted an immense amount of work to the attempt to understand the disease tuberculosis”

—Arnold R. Rich. The Johns Hopkins Hospital. January 1994.

Índice

1. Introducción general	1
Tuberculosis	1
Complejo <i>Mycobacterium tuberculosis</i>	2
Diversidad dentro del complejo MTB	5
Partículas infecciosas	7
Epidemiología de la tuberculosis	13
Diagnóstico y tratamiento	16
Tuberculosis resistente	18
Aplicaciones de la secuenciación masiva de genomas	20
2. Objetivos	23
Objetivo general	23
Objetivos específicos	24
Esquema de la tesis	25
3. Epidemiología molecular de la tuberculosis en la Comunidad Valenciana, España	26
Introducción	26
Métodos	27
Población de estudio	28
Colección de aislados y extracción de DNA	28
Secuenciación del genoma completo	29
Análisis bioinformático	29
Predicción de la resistencia a fármacos	32
Resultados	33
Población de estudio	33
Estructura poblacional y agrupamiento genómico	36
Caracterización de la resistencia a fármacos	46
Resistencia a isoniazida (INH)	46
Resistencia a rifampicina (RIF)	48
Resistencia a etambutol (EMB)	49
Resistencia a pirazinamida (PZA)	50
Caracterización de Multidrogoresistencia (resistencia simultánea, rifampicina e isoniazida)	51
Caracterización de la resistencia a fármacos de segunda línea	51
Discusión	53
Caracterización de la resistencia a fármacos	57
4. Genotipos endémicos en la Comunidad Valenciana	61
Introducción	61
Métodos	63
Identificación de genotipos en la Comunidad Valenciana	64
Relación entre genotipos y transmisión	67

Estudio filogeográfico e identificación de genotipos endémicos	67
Reconstrucción ancestral	68
Resultados	70
Identificación y distribución geográfica de genotipos en la Comunidad Valenciana	70
Relación entre genotipos, población local y transmisión	75
Estudio Filogeográfico e Identificación de genotipos endémicos	76
Reconstrucción ancestral	78
Discusión	80
5. Estimación del tamaño del cuello de botella en la transmisión de la tuberculosis humana a partir de la diversidad genómica donante-receptor.	84
Introducción	84
Métodos	89
Población de estudio y selección de pares de transmisión donante-receptor	89
Secuenciación de alta profundidad	91
Análisis bioinformático	92
Reconstrucción filogenética	93
Validación de variantes (SNPs)	93
Estimación del tamaño del cuello de botella	95
Comparación de resultados en otra población	96
Resultados	97
Reconstrucción filogenética	97
Diversidad genética en los pares de transmisión	98
Estimación del tamaño del cuello de botella de la transmisión	100
Discusión	105
6. Discusión General	110
Uso de la WGS en el estudio de la tuberculosis	110
Limitaciones y utilidades del catálogo de resistencias a fármacos de la OMS para tuberculosis	112
Uso WGS para la identificación de transmisión en tuberculosis	113
Identificación y monitoreo de genotipos endémicos de tuberculosis en la Comunidad Valenciana.	114
Identificación de pares de transmisión donante-receptor: desafíos y recomendaciones	115
¿Cuántos bacilos son necesarios para iniciar una infección natural de tuberculosis?	
Discusión de las estimaciones y su importancia para el control de la enfermedad.	115
7. Conclusiones	117
8. Bibliografía	119
Abreviaturas	146
Suplementario	147

1. Introducción general

Tuberculosis

La tuberculosis (TB) es una enfermedad infecto-contagiosa crónica inflamatoria, curable en la mayoría de los casos, que si no es tratada adecuadamente puede ser mortal ¹. Se reconoce como un grave problema de salud pública y, en ocasiones, se considera una pandemia debido a su amplia distribución geográfica y su impacto global. Probablemente, sea la enfermedad infecciosa que más muertes ha causado en la historia de la humanidad. La incidencia de la enfermedad varía en diferentes regiones del mundo, sin que ningún país esté exento de ella. Además, la TB está asociada con importantes repercusiones económicas y sociales, ya que afecta principalmente a poblaciones vulnerables y puede tener un impacto negativo en el desarrollo de las comunidades.

La TB afecta principalmente el sistema respiratorio, pero puede presentarse en distintas partes del cuerpo. En este sentido, la forma más frecuente y significativa de esta enfermedad, en términos epidemiológicos, es la TB pulmonar; sin embargo, la TB extrapulmonar afecta, aproximadamente, al 15% de casos de TB a nivel mundial ².

La tuberculosis se presenta en dos formas: activa y latente. La TB activa es sintomática y transmisible. Por otro lado, la TB latente implica la presencia de la bacteria en el cuerpo sin síntomas. No obstante, la TB latente puede progresar a TB activa si no se trata adecuadamente. Los síntomas de la TB activa pueden ser fatiga, debilidad, adelgazamiento, fiebre y sudores nocturnos. La afección de los pulmones origina tos crónica y hemoptisis, que surgen en caso de lesiones avanzadas. Los pacientes sin tratamiento o que presentan resistencia al mismo fallecen de choque séptico, insuficiencia respiratoria u otra complicación causada por la enfermedad. La dicotomía TB latente/TB activa ha sido el paradigma predominante en esta enfermedad; sin embargo, en los últimos años se han identificado nuevos estadios de TB, entre los que se incluye la TB subclínica ^{3,4}. Se trata de una forma asintomática de la enfermedad durante la cual los individuos infectados son potenciales transmisores, por este motivo la TB subclínica ha llamado la atención de muchos grupos de investigación ^{4,3,4}.

La TB es una enfermedad que se puede padecer más de una vez a lo largo de la vida. Si el paciente no padeció anteriormente TB será una TB primaria, si ya ha padecido es llamada TB secundaria; habitualmente es una recaída a partir de una infección original o una reinfección exógena ^{5,6}.

En cuanto al tratamiento, existen terapias altamente efectivas para combatir las cepas sensibles de la bacteria. No obstante, el surgimiento de resistencia a los antibióticos de primera línea ha llevado al uso de tratamientos más complejos de segunda línea, con resultados variables.

Complejo *Mycobacterium tuberculosis*

La tuberculosis es una enfermedad causada por bacterias pertenecientes al género *Mycobacterium*. Las especies más frecuentes que infectan a los seres humanos son parte del complejo *Mycobacterium tuberculosis* (MTBC) e incluyen: *M. tuberculosis*, *M. africanum*, *M. bovis*, *M. canetti*, y *M. microti* ⁷. Entre estas especies, sin duda *M. tuberculosis* y *M. africanum* son los principales patógenos causantes de la TB en humanos, seguido de *M. bovis* ⁸. Adicionalmente, en los últimos años, las infecciones causadas por micobacterias atípicas, caracterizadas como no tuberculosas (NTM, “Nontuberculous mycobacteria”), ha aumentado significativamente ^{9,10,11}.

Mycobacterium tuberculosis es un bacilo gram positivo, aerobio estricto, sin movilidad y de crecimiento lento, protegido por una pared celular serosa, compuesta por abundantes glicolípidos, ácidos micólicos, peptidoglicanos, carbohidratos, lípidos y proteínas ¹². Esta barrera es responsable de muchos de los aspectos fisiológicos e infecciosos en las micobacterias y la hace única entre los patógenos bacterianos ^{13,14}. Su genoma comprende alrededor de 4,411,529 pb, contiene 3,959 genes y tiene un 65.6% de guaninas-citosinas. *M. tuberculosis* difiere radicalmente de otras bacterias debido a que gran parte de su capacidad codificante se dedica a la producción de enzimas implicadas en la lipogénesis y lipólisis ¹⁵.

Dentro del complejo MTBC, el grupo conocido como *M. tuberculosis* sensu stricto está formado por bacterias que infectan principalmente a humanos y comprende siete linajes filogenéticos (L1-L4 y L7-L9) ¹⁶. El L1 está asociado a las regiones del océano Índico y Filipinas, el L2 ha sido identificado en el este de Asia, el L3 proviene del este de África e

India y el L4 fue descubierto en América y Europa. Los linajes L5 y L6 (*Mycobacterium africanum*) se encuentran en África oriental. El L7 fue identificado en la región del Cuerno de África y el L8 se encuentra limitado a la región de los grandes lagos en África ^{17,18}. El linaje L9 fue recientemente descubierto en la región oriental de África ¹⁹ (ver Figura 1.6).

Los linajes L1-L4 son los responsables del mayor número de infecciones en humanos a nivel global ²⁰⁻²² (Figura 1.6), a diferencia de los linajes L5 y L6; considerados menos virulentos ^{23,24}. Los linajes L7-L9 se encuentran limitados a sus regiones, lo que sugiere una adaptación particular a estas poblaciones ²⁵.

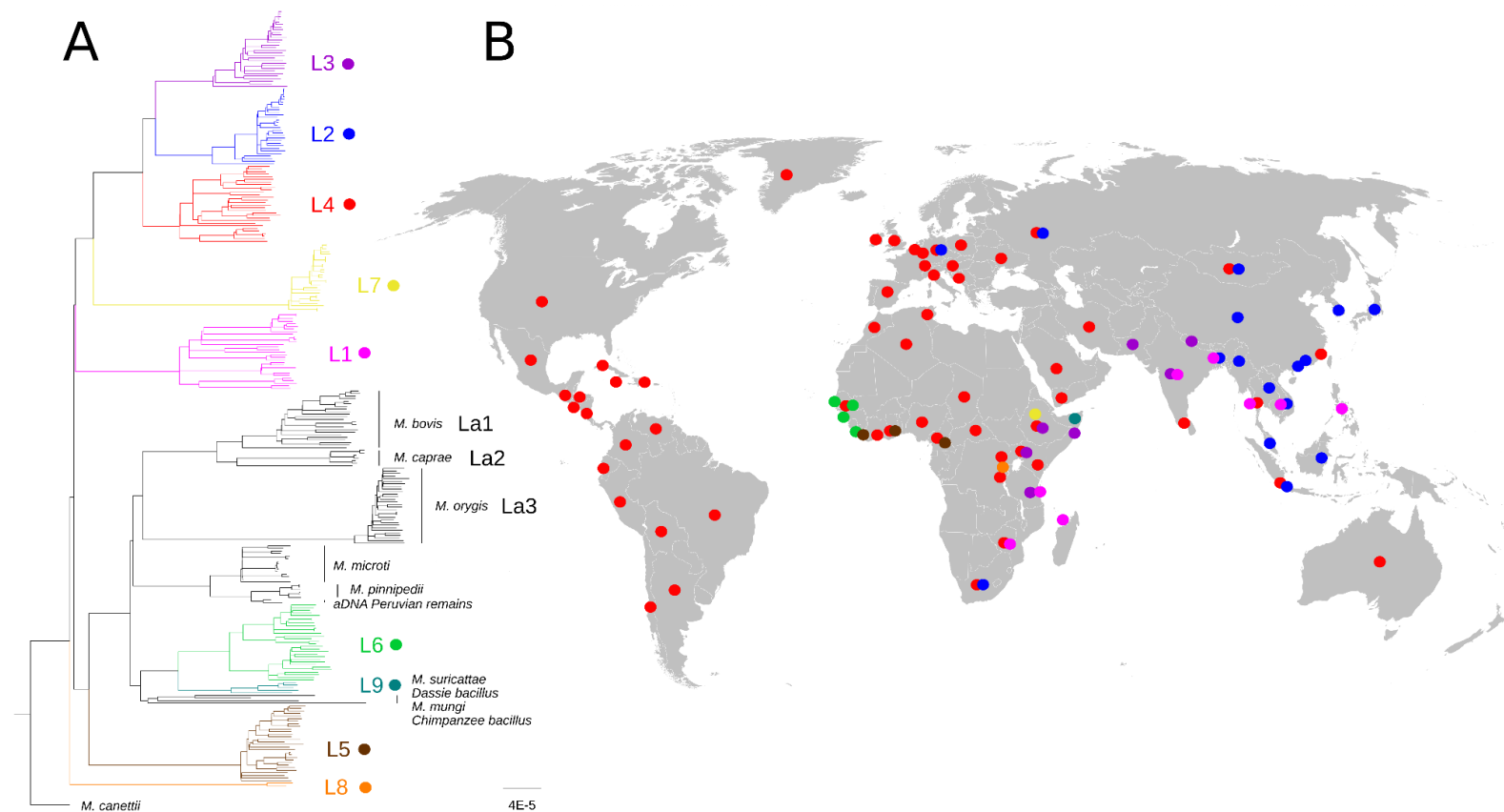


Figura 1.6: Filogeografía global del MTBC adaptado al ser humano.

a) Filogenia de máxima verosimilitud enraizada del MTBC estimada a partir de SNP de todo el genoma (árbol adaptado de ^{26 27}). *M. canettii* es el grupo externo, los linajes adaptados a humanos (L1 a L9) se muestran en colores, los linajes adaptados a animales en negro. Los nombres de especies representan la nomenclatura históricamente establecida, los nombres de linajes son una clasificación más reciente basada en datos genómicos. Los linajes 1 a 4 y 7 a 9 también se conocen como *M. tuberculosis* sensu stricto, los linajes 5 y 6 como *M. africanum*.

b) Distribución global de los siete principales linajes de MTBC adaptados al ser humano. Adaptada de (²⁸).

Diversidad dentro del complejo MTB

La variabilidad genética del MTBC es un factor importante en la patogénesis que puede afectar a la virulencia, la transmisibilidad, la respuesta del huésped y la aparición de farmacorresistencia ²⁹. La baja diversidad en *M. tuberculosis* se debe, entre otros factores, a la ausencia de eventos de recombinación o de transferencia horizontal de genes, que son los principales contribuyentes a la diversidad en otros patógenos bacterianos. Se estima que la tasa de sustitución de *M. tuberculosis* es baja, con un promedio de aproximadamente 4.6×10^{-8} sustituciones por sitio por año ³⁰. Sin embargo, la tasa de mutación es mayor y varía en diferentes linajes o en presencia de factores selectivos, como la exposición a medicamentos³¹. Se estima que *M. tuberculosis* adquiere mutaciones a una tasa promedio de 0.2-0.5 nucleótidos por genoma por año, aunque esta tasa puede variar dependiendo de las características del individuo infectado y la cepa específica ³²⁻³⁵.

La diversidad de *M. tuberculosis* ha sido estudiada desde el nivel global ³⁶ hasta el intra hospedador ³⁷. Esta diversidad se traduce en una importante variación fenotípica que tiene implicaciones en el diagnóstico y evolución de la enfermedad ^{36,38}. A pesar de la limitada variación en comparación con otras bacterias ³⁹, se han observado diferencias fenotípicas entre cepas ⁴⁰, así como diferencias en la expresión génica, la resistencia a fármacos, la transmisibilidad y la respuesta inmune. Estos pequeños cambios en su genoma suponen una adaptación específica, lo que explicaría la distribución global de los linajes del complejo MTBC ³⁶, y sería congruente con las diferencias observadas en la virulencia de cepas clínicas ⁴¹.

La diversidad filogeográfica del MTBC se atribuye a factores ecológicos, eventos demográficos y epidémicos recientes ⁴². Los diferentes linajes del MTBC muestran variaciones en su capacidad de infectar, lo que indica una adaptación a poblaciones humanas específicas ⁴³. En el contexto de India, los linajes Delhi y Beijing son los más predominantes, lo que señala al país como un foco endémico para estos linajes ⁴⁴. Las cepas del linaje Beijing han provocado brotes significativos de tuberculosis a nivel mundial, a veces con resistencia a múltiples medicamentos ⁴⁵. Además, se han identificado asociaciones específicas en distintas poblaciones. Por ejemplo, en Arabia Saudita se ha establecido una relación directa entre los

linajes *M. bovis* y Delhi con la tuberculosis de los ganglios linfáticos y la tuberculosis gastrointestinal ⁴⁶. En Tanzania, se ha demostrado que el linaje 2 se asocia con el sexo femenino y los casos de retratamiento ⁴⁷. En Uruguay, se ha observado un incremento notable del linaje Haarlem entre los pacientes de la unidad de cuidados intensivos (UCI) y en la incidencia de disfunciones hepáticas ⁴⁸. Esto evidencia que los linajes del MTBC influyen en los fenotipos de la enfermedad y en sus consecuencias epidemiológicas.

A nivel genómico, la variabilidad entre los diferentes linajes del complejo MTBC no es considerable (99% de similitud). No obstante, estas diferencias son apreciables a nivel fenotípico y se relacionan con la patogenicidad y virulencia de los linajes ^{28,49}.

La diversidad genética del huésped también desempeña un papel crucial en la manifestación de la tuberculosis. Investigaciones han identificado marcadores genéticos que permiten predecir el desarrollo de la enfermedad y evaluar la respuesta inmunitaria de los pacientes ^{50,51}. Se ha observado una asociación significativa entre la capacidad de las poblaciones humanas para producir vitamina D y su susceptibilidad a la infección ⁵². Estas variaciones genéticas podrían utilizarse como indicadores en la toma de decisiones clínicas y epidemiológicas en el futuro.

La diversidad genética de *Mycobacterium tuberculosis* en individuos infectados puede variar, con la presencia de un solo foco de infección en el huésped que contiene subpoblaciones clonales de bacterias, o la existencia de múltiples nódulos con infecciones policlonales ^{37,53,54}. Sin embargo, el estudio de la diversidad de variantes de baja frecuencia dentro del huésped infectado ha sido escasamente investigado, principalmente debido a las dificultades en su detección ⁵⁵.

Por otra parte, pocos estudios han analizado la diversidad de *M. tuberculosis* en pares de transmisión donante-receptor humanos ⁵⁶, debido a la dificultad de identificar la transmisión directa y a la limitada resolución de los métodos de secuenciación. Además, el estudio de variantes con frecuencias inferiores al 10%, presenta desafíos adicionales. La detección de SNPs a tan baja frecuencia dentro de las poblaciones bacterianas está influenciada por factores como la profundidad de secuenciación, así como la representatividad, origen y método de obtención de las muestras clínicas. Estos factores deben ser considerados al

analizar la diversidad genética de *Mycobacterium tuberculosis* en contextos de transmisión entre individuos.

Partículas infecciosas

Las partículas infecciosas han sido un tema de investigación desde que Flugge ^{57,58} las identificó en sus experimentos en 1890, en el entendimiento actual existen diferentes tamaños de partículas y éstas, a su vez, tienen una relación directa con la capacidad de penetrar a diferentes profundidades en los pulmones (Figura 1.1) ⁵⁹.

La tuberculosis (TB) se propaga eficazmente a través de la transmisión aérea, mediante aerosoles, que son diminutas partículas expulsadas al hablar, toser, estornudar o simplemente al respirar ⁸. Estos movimientos respiratorios generan un fluido compuesto por partículas de diferentes tamaños y con una compleja mezcla de moco, saliva, sangre, células inmunes, células epiteliales y microorganismos. En el caso de la TB, las microgotas provenientes de la tos de individuos infectados suelen ser más pesadas, lo que dificulta su permanencia en el aire y reduce su capacidad de penetración en las vías respiratorias.

Por otro lado, existe un grupo de partículas en el rango de 0.05-100 µm conocidas como bioaerosoles. Estos bioaerosoles, generados durante la exhalación, tienen una alta capacidad de penetración en los pulmones y se mantienen suspendidos en el ambiente con facilidad. Estudios realizados en personas con TB han demostrado que la mayoría de los bacilos exhalados, más del 90%, provienen de la respiración normal ^{60,61}. Incluso durante la noche, en ausencia de tos, se ha observado la exhalación de bacilos ⁶⁰, lo que resalta la importancia de la transmisión de la enfermedad a través de la respiración cotidiana.

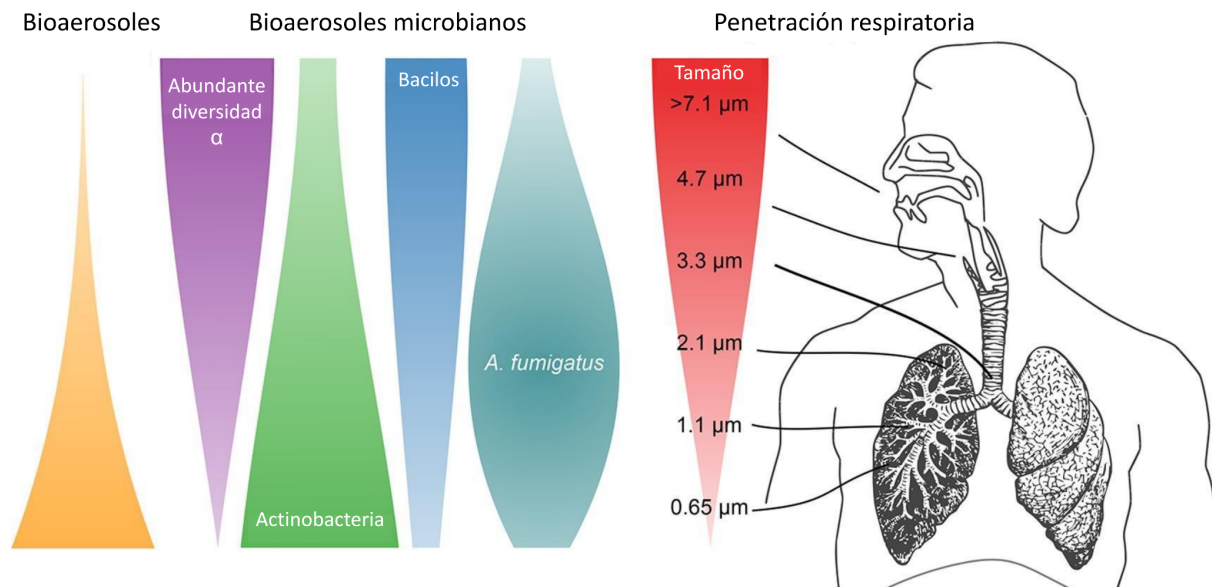


Figura 1.1: Intervalo de penetración respiratoria de diferentes microorganismos en bioaerosol. Adaptado de ⁽⁵⁹⁾

En relación con la dosis de infección, actualmente desconocemos el número de bacilos exactos capaces de fundar una infección en humanos. En modelos animales (cobayas), estudios a principios del siglo XX, han demostrado dosis infecciosas bajas (20-40 bacilos) ⁶² y, más recientemente, se ha logrado establecer una infección con dosis ultrabajas (1-3 bacilos) en modelo murino ⁶³. Los resultados de estos estudios plantean diferentes preguntas desde conocer si la dosis de infección es similar en la transmisión natural en humanos, y si ésta puede estar relacionada con la severidad de la enfermedad.

Es importante mencionar que se desconoce hasta el momento la variabilidad genética que pueda existir entre las bacterias en el interior de los pulmones de un individuo, el esputo, el bioaerosol y las bacterias inhaladas por un nuevo individuo. Aunque existen diversas aproximaciones sobre la diversidad en esputo y cultivo, es un tema complejo y supondrá un reto para los próximos años entender las reglas del juego de la variabilidad genética en la transmisión.

Infección y transmisión

El órgano diana principal de *M. tuberculosis* es el pulmón, el cual inhala aproximadamente 10,000 litros de aire al día, por lo que está expuesto a una enorme cantidad de

microorganismos que la mayoría de las veces son incapaces de colonizar, debido a una respuesta inmune innata efectiva ⁶⁴. La respuesta inmune del pulmón, comprende barreras físicas, así como células de defensa como los neutrófilos, macrófagos, células cebadas, basófilos, eosinófilos y células asesinas (*NK*). Los conductos del pulmón se ramifican en bronquios primarios (3.3 μm), bronquios secundarios (2.1 μm), bronquiolos terminales (1.1 μm) y alvéolos (0.65 μm) (Figura 1.2). Los bronquios primarios y secundarios en el pulmón están revestidos por células epiteliales que desempeñan un papel crucial como barrera física y química. Estas células secretan moco y diversas proteínas vinculadas a la respuesta inmune, como enzimas, quimiocinas, citocinas y defensinas ⁶⁵. Los bronquios terminales y los alvéolos, debido a su fisiología, carecen de epitelio y son, por lo tanto, un sitio idóneo para la proliferación de *M. tuberculosis*.

La dinámica de transmisión más frecuente de la TB consiste en que un individuo infectado (donante) exhala un bioaerosol con un número y viabilidad desconocidos de bacilos; una parte de este es inhalada por un segundo individuo (receptor), donde pequeñas partículas (1.1 – 2.1 μm ⁶⁶) se depositan en los bronquiolos terminales (diámetro 0,5 mm). Una vez allí, los bacilos son fagocitados por los macrófagos, donde inhiben la formación del fagolisosoma e inician su replicación ⁶⁷.

Si la respuesta inmune del individuo infectado es exitosa, la infección no progresa y no se desarrollará TB activa, ya que las micobacterias serán destruidas o quedarán en estado latente dentro de los macrófagos ⁶⁸. Sin embargo, *M. tuberculosis* ha desarrollado mecanismos para contrarrestar la capacidad microbicida de los macrófagos, algunos de estos son: el aumento de la resistencia a compuestos tóxicos del huésped, detención de la progresión normal del fagosoma e inducción de apoptosis ¹³. En este contexto, la infección se volverá activa, progresará hacia la generación de granulomas, la respectiva destrucción tisular y, en algunos casos, la diseminación de los bacilos a otros órganos por vía hematógica ⁶⁹.

El individuo infectado puede transitar diferentes estados de la enfermedad. La infección puede no darse debido a la acción del sistema inmune, mantenerse en un estado latente, empezar a desarrollarse de manera subclínica o avanzar a una infección activa (Figura 1.2).

El momento exacto cuando un individuo es capaz de exhalar bacilos y volverse infeccioso es difícil de identificar. Tradicionalmente se ha pensado que sólo los individuos sintomáticos

son capaces de transmitir la enfermedad, sin embargo la evidencia acumulada sugiere que existen estados subclínicos, asintomáticos, donde los individuos son bacilíferos. Independientemente del estado de la enfermedad en que se encuentre el individuo, si este es capaz de exhalar bacilos viables, la transmisión es posible.

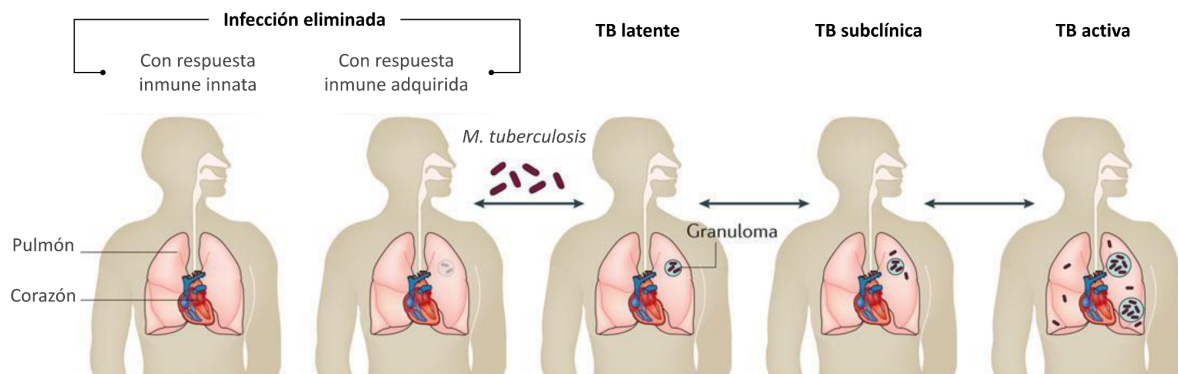


Figura 1.2: Espectro de la TB, desde la infección por *Mycobacterium tuberculosis* hasta la TB activa (pulmonar). Tomado y adaptado de ⁽⁷⁰⁾.

Es importante mencionar que existen numerosas variables que pueden interferir en el proceso de la transmisión e infección de la TB. En un individuo intervienen una gama de factores como; la resistencia nativa a la infección, la resistencia adquirida, la hipersensibilidad y la virulencia del bacilo. Además, factores de riesgo como VIH, diabetes, desnutrición, uso de drogas, inmunodepresión, neoplasia, entre otros; pueden deteriorar el estado físico del hospedador; comprometiendo su capacidad para evitar o controlar una infección microbiana, y favoreciendo su capacidad transmisora. Sin mencionar factores geográficos o sociales ^{71,72}.

Estudio de la transmisión de TB

Asimismo, la transmisión de la tuberculosis es un factor determinante en el estudio de esta enfermedad. Para su evaluación y cuantificación se utilizan diversas medidas epidemiológicas. La tasa de incidencia proporciona una estimación de la propagación de la enfermedad al contar los nuevos casos diagnosticados en un período específico. El número básico de reproducción (R_0) permite evaluar la capacidad de la tuberculosis para propagarse al indicar cuántas personas, en promedio, puede infectar un individuo enfermo. El seguimiento de contactos proporciona información sobre las personas en contacto directo con pacientes de tuberculosis, y el análisis de clusters genéticos identifica agrupaciones de

aislados con similitudes genéticas, revelando posibles cadenas de transmisión. Adicionalmente, las tasas de incidencia y prevalencia también pueden revelar tendencias temporales en la transmisión ⁷³, además de identificar subpoblaciones en riesgo, como personas nacidas en el extranjero, privadas de su libertad o personas sin hogar ⁷⁴. Asimismo, permiten asociar la transmisión a lugares públicos, como medios de transporte, iglesias, escuelas, bares y barrios marginales ⁷⁵.

La evaluación de la transmisión de la tuberculosis presenta notables desafíos. La diversidad del espectro clínico y la variabilidad en el periodo de activación de la enfermedad implica que los casos observados reflejen tanto transmisiones recientes como de mediano plazo, así como reactivaciones de infecciones previas. La complejidad en el diagnóstico conduce a una subestimación de la verdadera prevalencia de la infección, ya que muchos enfoques no logran capturar la totalidad de los casos ni identificar a individuos en estadios clínicos transmisibles. Además, la precisión en la inferencia de infecciones recientes se ve comprometida por la falta de sensibilidad de pruebas como la tuberculina, “IGRA”, radiografía de tórax y la tinción de esputo ⁷⁶. A pesar de estos desafíos, se implementan diversas estrategias para estimar la transmisión de la tuberculosis.

Un enfoque común es el análisis de la notificación de casos y el estudio de sus contactos mediante encuestas. La dificultad de éste radica en diagnosticar la infección en los contactos recientes. En la práctica clínica, se utiliza ampliamente el enfoque de pruebas diagnósticas, como la prueba de la tuberculina y las pruebas de liberación de interferón-gamma, para detectar la exposición a *M. tuberculosis* y evaluar la respuesta inmune, incluso en ausencia de síntomas clínicos. Estas pruebas son especialmente útiles para estimar las infecciones adquiridas recientemente y realizar cribado en investigaciones de contactos y en grupos de alto riesgo, como inmigrantes o trabajadores de la salud ⁷⁵.

Sin embargo, es importante tener en cuenta que estas pruebas tienen limitaciones. No pueden distinguir entre infección latente y enfermedad activa, lo cual es una restricción significativa en entornos de alta transmisión de la enfermedad. En estas situaciones, la capacidad de las pruebas puede verse disminuida debido a la posibilidad de falsos negativos. Por lo tanto, aunque las pruebas de la tuberculina y las pruebas de liberación de interferón-gamma son herramientas valiosas en la clínica, es esencial considerar su rendimiento en contextos

específicos y complementar su uso con otros enfoques para lograr una evaluación integral de la transmisión de la tuberculosis.

La disponibilidad de métodos moleculares ha permitido un enfoque ampliamente utilizado en el estudio de la tuberculosis ⁷⁷. Mediante la tipificación molecular de cepas, es posible evaluar los vínculos genotípicos entre los casos índice de tuberculosis y los posibles casos secundarios. A nivel de población, se han desarrollado métodos como el enfoque "n - 1" y otras técnicas basadas en modelos para estimar los niveles de transmisión reciente, utilizando agrupamientos genotípicos como referencia ⁷⁸⁻⁸¹.

Los estudios epidemiológicos moleculares han proporcionado evidencia confirmatoria sobre la transmisión de la tuberculosis en grupos de alto riesgo, como los adolescentes. Además, han establecido la importancia de la transmisión nosocomial ⁸² y han identificado otros factores de riesgo para la transmisión reciente, como la residencia urbana, la falta de vivienda ⁸³ y la exposición en entornos de alta densidad poblacional, como las prisiones ⁸⁴.

Las técnicas moleculares más utilizadas hasta ahora (espilgotipos, "MIRU", "RFLP-IS6110") tienen como principal limitación que interrogan una pequeña porción del genoma, de modo que los links genómicos que pueden identificarse, a menudo corresponden a eventos de transmisión muy antiguos ⁸⁵. En la actualidad, la epidemiología genómica ha revolucionado los estudios de transmisión a nivel de pacientes y población mediante la secuenciación completa del genoma. Se ha observado que casos con vínculos genéticos a menudo carecen de vínculos epidemiológicos tradicionales ⁸⁶. Además, se ha identificado que la transmisión de la tuberculosis en entornos con alta carga de la enfermedad ocurre principalmente fuera de los hogares o de los contactos cercanos conocidos ⁸⁷. Asimismo, esta técnica ha permitido visualizar con precisión las ubicaciones donde ocurren las infecciones en comunidades específicas. Se ha encontrado que lugares públicos como centros de transporte, aulas escolares y prisiones facilitan el intercambio de aire entre casos infecciosos y personas no infectadas, convirtiéndose en puntos críticos de transmisión. Estos hallazgos brindan oportunidades para implementar intervenciones específicas en dichos lugares ^{88,89}.

Tras analizar el panorama actual del estudio de la transmisión de la tuberculosis, es evidente que se requiere un enfoque integrado que combine la epidemiología tradicional con métodos derivados del análisis espacial, demográfico, de grupos de transmisión y de datos de

secuenciación genómica. Esta combinación de enfoques permitirá identificar de manera precisa los factores que favorecen la transmisión de la enfermedad.

Epidemiología de la tuberculosis

La TB es una de las principales causas de muerte en todo el mundo y una de las principales causas de muerte por un agente infeccioso. De acuerdo con el reporte 2022 de la Organización Mundial de la Salud (OMS), en 2021 se registraron 10.6 millones de nuevos casos de TB y 1.6 millones de defunciones por este padecimiento (1.4 millones en personas VIH negativas y 187,000 en personas VIH positivas), por lo que se ubicó como la enfermedad infecciosa con mayor impacto en la salud humana. Siendo los países con mayor incidencia: India, Indonesia, Filipinas, China, Pakistán, Nigeria, Bangladesh y el Congo ⁹⁰ (ver Figura 1.3).

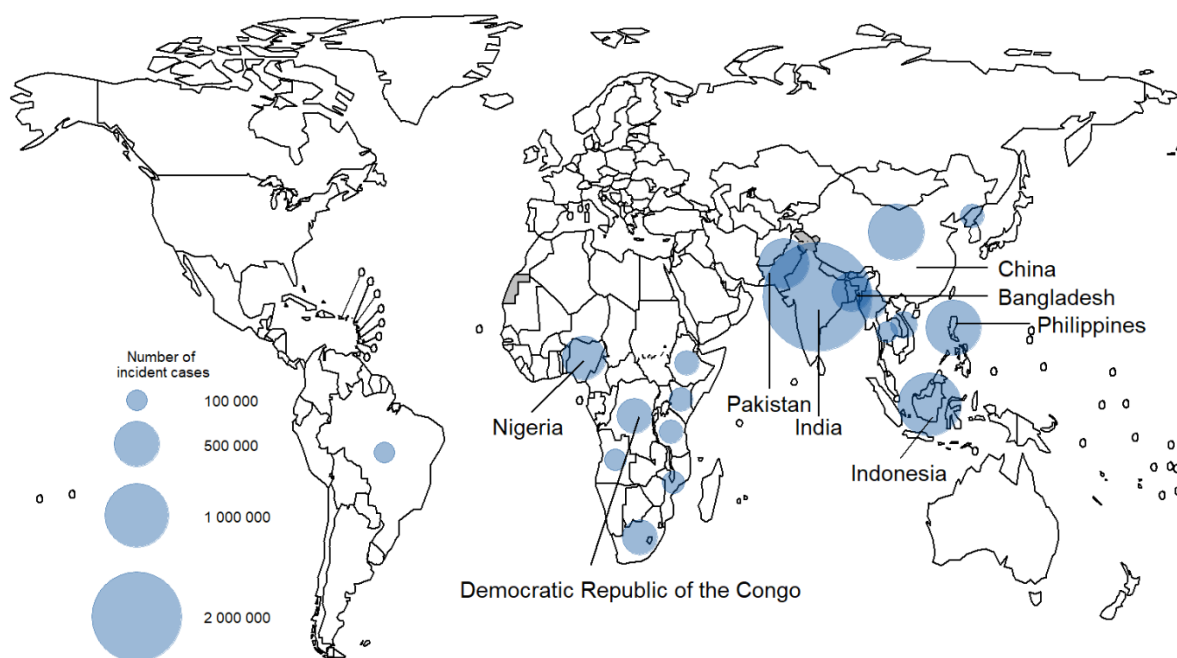


Figura 1.3: Incidencia estimada de TB en 2021, para países con al menos 100.000 casos

La TB resistente a fármacos representa un problema de salud en constante aumento. En el año 2021, se informaron aproximadamente 450 mil nuevos casos de tuberculosis MDR/RR. Es importante destacar que la mitad de estos (42%) ocurrieron en India (26%), Rusia (8%) y Pakistán (7.9) (ver Figura 1.4). Además, se estima que solo el 43% de los individuos que

necesitaba tratamiento contra la TB-MDR lo recibieron, lo que indica una crisis en la detección, prevención y tratamiento de la TB-MDR ⁹⁰.

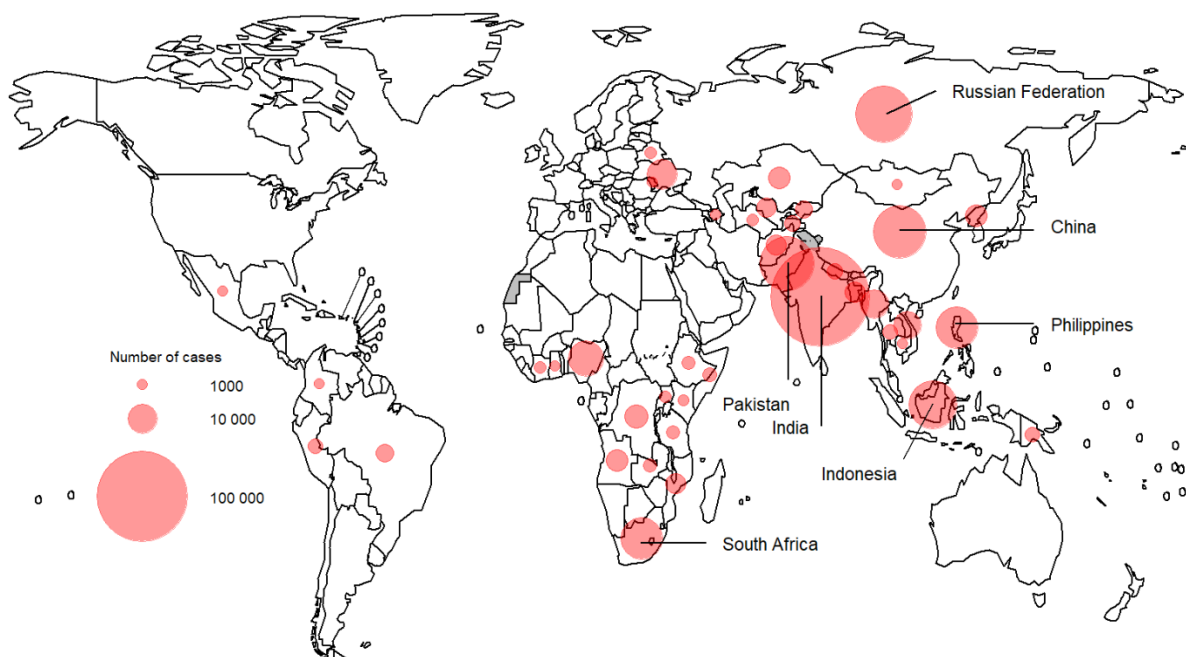


Figura 1.4: Incidencia estimada de TB-MDR/RR en 2021, para países con al menos 1000 casos

Los principales factores de riesgo de contraer tuberculosis incluyen la exposición directa a pacientes con tuberculosis activa, especialmente aquellos que tienen un contacto cercano y prolongado, como el personal penitenciario, los reclusos ⁹¹ y los trabajadores sanitarios ⁹². Además, las personas con un sistema inmunológico debilitado, como los infectados con el virus de la inmunodeficiencia humana (VIH) ⁹³ o que reciben tratamiento inmunosupresor ⁹⁴, tienen un mayor riesgo de desarrollar la enfermedad. La malnutrición, la falta de acceso a atención médica adecuada, las condiciones de vida desfavorables y el consumo de tabaco también se han identificado como factores de riesgo significativos para la tuberculosis ^{92,95}. La edad avanzada, la presencia de diabetes y el uso de drogas intravenosas son otros factores que aumentan la susceptibilidad a la infección. Además, las poblaciones migrantes y desplazadas son especialmente vulnerables debido a las condiciones precarias en las que viven y a la limitada accesibilidad a la atención médica ^{94,96}. Es importante tener en cuenta que estos factores de riesgo pueden interactuar y aumentar el riesgo de desarrollar la enfermedad.

La Organización Mundial de la Salud (OMS) ha presentado un plan estratégico “End TB” con el objetivo de controlar y erradicar la pandemia de tuberculosis (TB). Este plan busca reducir la mortalidad por TB en un 90% y disminuir la incidencia en un 80% para el año 2030. Las metas establecidas en este programa incluyen el desarrollo de nuevos fármacos y la generación de datos genotípicos, fenotípicos y metadatos relacionados con las diferentes cepas de *Mycobacterium tuberculosis*. El propósito de recopilar esta información es permitir la creación de pruebas de diagnóstico de resistencia a los medicamentos que sean clínicamente útiles, rápidas y altamente sensibles y específicas.

El objetivo final es mejorar la capacidad de detectar y tratar de manera eficaz los casos de TB, especialmente aquellos que presentan resistencia a los medicamentos existentes. Con el desarrollo de pruebas de diagnóstico más precisas y eficientes, se espera lograr una detección temprana de la enfermedad y proporcionar un tratamiento adecuado de manera oportuna. Esto contribuirá a frenar la propagación de la TB y reducir su impacto en la salud pública a nivel mundial.

A pesar de los esfuerzos realizados por la Organización Mundial de la Salud (OMS), la tuberculosis (TB) sigue siendo una enfermedad prevalente en varias partes del mundo, especialmente en países de ingresos bajos y medianos. Estas regiones, como África, el sudeste asiático y el Pacífico occidental, concentran la mayor carga de casos de TB a nivel mundial. No obstante, es relevante destacar que la TB también afecta a otros países, incluyendo aquellos con sistemas de salud más desarrollados.

La falta de avances significativos en el control de la TB se debe a diversos factores. Uno de ellos es la aparición de cepas resistentes a los medicamentos, como la tuberculosis resistente a los medicamentos (TB-MDR) y la tuberculosis resistente a múltiples medicamentos (TB-XDR). Además, existen desafíos en cuanto al acceso a pruebas de diagnóstico precisas y sensibles, especialmente en áreas con recursos limitados. Esto dificulta la detección temprana de la enfermedad y, consecuentemente, retrasa el inicio del tratamiento adecuado.

En resumen, la situación actual de la tuberculosis a nivel mundial se caracteriza por una carga de enfermedad considerable y un progreso insuficiente en su control y erradicación. Es importante resaltar que en el año 2020, la tuberculosis sea la segunda causa principal de muerte causada por un agente infeccioso, después del COVID-19 ⁹⁷.

En España la incidencia de TB durante el 2020 fue de 7.3 casos por cada 100.000 habitantes, los principales factores de riesgo asociados a estos casos fueron; tabaquismo, VIH positivo, desnutrición, alcoholismo y diabetes ⁹⁸. En la Comunidad Valenciana se reportaron 377 casos nuevos de TB en el año 2020, con una incidencia de 7.5 casos por 100.000 habitantes. El 88.7% del total de pacientes con TB en la Comunidad Valenciana completa con éxito el tratamiento ⁹⁹.

Diagnóstico y tratamiento

El diagnóstico clínico de TB es el resultado de la observación de signos y síntomas como fatiga, debilidad, adelgazamiento, fiebre y sudores nocturnos. El dato clínico característico de la TB pulmonar es una tos crónica acompañada de hemoptisis que surge en el caso de lesiones avanzadas. El médico puede aplicar la prueba de tuberculina para indicar si el paciente ha estado en contacto con el bacilo y apoyarse en una placa torácica de rayos X para observar si existe la formación de un tubérculo en los pulmones del individuo.

La prueba diagnóstica de laboratorio más empleada es la microscopía de frotis del esputo, desarrollada hace más de 100 años, cuenta con una sensibilidad entre el 50 y 80% ¹⁰⁰. Las muestras de esputo son teñidas y examinadas bajo el microscopio para apreciar la ausencia o presencia de bacilos. En países de alta incidencia, la microscopía sigue siendo la principal técnica de diagnóstico para pacientes sintomáticos ¹⁰¹.

El cultivo de aislados de *M. tuberculosis* en los medios selectivos como *Löwenstein-Jensen* o *Middlebrook* (“7H10/7H11”) es el estándar actual para el diagnóstico de la TB, pero requiere una infraestructura desarrollada y un tiempo de 6 semanas en obtener un diagnóstico ¹⁰¹. Una vez identificado el bacilo, toma entre 4 a 13 días realizar las pruebas de sensibilidad o resistencia a fármacos.

El cultivo es uno de los métodos más sensibles en la identificación de *M. tuberculosis* (>90%), a partir de él se pueden efectuarse pruebas de sensibilidad a fármacos; por lo tanto, no solo detectan la presencia o ausencia del bacilo en la muestra, también proporciona un perfil de resistencia a fármacos el cual es tomado en cuenta para establecer el esquema de tratamiento del paciente. También es necesario para el diagnóstico pediátrico y

extrapulmonar, así como para el diagnóstico diferencial de NTM. Actualmente, existen equipos automatizados como el BD BACTEC™ MGIT™ que facilitan dicha labor.

La prueba rápida para el diagnóstico de TB con resistencia a fármacos actualmente recomendada por la OMS es el ensayo “Cepheid GeneXpert® MTB/RIF” (Cepheid, Sunnyvale USA). Inicialmente, se recomendó (en 2010) para la detección de TB pulmonar en adultos. Desde 2013, también se ha recomendado en niños y formas específicas de TB extra pulmonar ¹⁰¹. La prueba tiene una sensibilidad del 90% para muestras con baciloscopia positiva y un 70% para muestras con baciloscopia negativa ^{102,103}. Las desventajas del ensayo “Cepheid GeneXpert® MTB/RIF” ¹⁰⁴ son: alto costo, la detección única de resistencia a rifampicina fármacos y el diagnóstico de un único paciente por prueba. Recientemente, la misma empresa lanzó “Xpert® MTB/XDR”, un cartucho que se utiliza en combinación con el anterior y permite evaluar la resistencia a isoniazida, fluoroquinolonas, etionamida, amikacina, kanamicina y capreomicina. En todos los casos, el método de “GeneXpert®” se basa en la amplificación de una región de los genes de interés, y posterior hibridación con sondas específicas para las mutaciones asociadas a resistencia más conocidas; siendo esta una de las principales limitaciones del método.

Existen otras metodologías aplicadas al diagnóstico molecular de *M. tuberculosis* ¹⁰⁵, desde el uso de hibridación con sondas específicas de secuencias del bacilo hasta la creación del perfil de ácidos micólicos para su identificación por cromatografía líquida de alta resolución (HPLC) ¹⁰⁶. Aunque estos ensayos tienen un valor significativo y permiten un diagnóstico rápido de la TB, es evidente que se requieren mejorar los métodos de diagnóstico moleculares ¹⁰⁷. La OMS también recomienda una prueba rápida “TB-LAMP” como sustituto de la tinción de esputo, la cual puede obtener resultados en una hora (sensibilidad, 0.78 y especificidad, 0.98) en adultos sintomáticos ¹⁰¹.

En relación con el tratamiento, en la actualidad hay más de 30 medicamentos aprobados por la OMS para el tratamiento de la tuberculosis, los cuales son empleados en todo el mundo. De estos fármacos, los de primera elección al inicio del tratamiento son ¹⁰⁸:

- Isoniazida (INH)
- Rifampicina (RIF)
- Etambutol (EMB)
- Pirazinamida (PZA)

La administración de estos fármacos sigue un régimen principal de seis meses y medio para los nuevos pacientes, con una fase intensiva de 2 meses y una de continuación por 4.5 meses (Tabla 1.1) ¹⁰⁸.

Tabla 1.1. Regímenes de tratamiento en tuberculosis

Fase	Fármaco	Tiempo en semanas	Régimen preferido	Régimen alternativo	Régimen VIH
Intensiva	INH	8	56 dosis, 7 d/sem	40 dosis, 5 d/sem	24 dosis, 3 d/sem
	RIF				
	PZA				
	EMB				
Continuación	INH	18	126 dosis, 7 d/sem	90 dosis, 5 d/sem	54 dosis, 3 d/sem
	RIF				

Fuente: Adaptado de Nahid, P. et al.¹⁰⁹. INH: isoniazida, RIF: rifampicina, EMB: etambutol, PZA: pirazinamida, d/sem: días a la semana.

Se han diseñado varios ensayos para evaluar si un régimen de tratamiento más corto puede seguir siendo altamente eficaz y mantengan la seguridad del tratamiento. Basándose en los resultados de estos ensayos, la OMS ha recomendado dos regímenes diferentes de 4 meses: uno basado en el “Study 31 trial” ¹¹⁰ y otro basado en el “SHINE trial” ¹¹¹.

Las pautas de tratamiento pueden cambiar dependiendo de la evolución, comorbilidades, estado y perfil de resistencia/sensibilidad a fármacos del paciente. Empleando fármacos de segunda línea como kanamicina (KM), amikacina (AM), capreomicina (CM), ofloxacino (OFX), levofloxacino (LFX), etionamida (ETO) y linezolid (LZD).

Tuberculosis resistente

La resistencia a los medicamentos, en *M. tuberculosis*, complica tanto el diagnóstico como el tratamiento de la enfermedad, favoreciendo la transmisión del bacilo y la aparición de nuevos casos que portan de inicio una *M. tuberculosis* resistente.

Cuando la TB sensible a fármacos es tratada con un régimen basado en un único fármaco, es factible que las bacterias con polimorfismos de resistencia sean seleccionadas, proliferen durante el tratamiento y se conviertan en la población dominante. Si un individuo con una TB resistente a un medicamento específico es tratado con un nuevo medicamento, es probable que desarrolle resistencia al medicamento adicional. Las adiciones de fármacos pueden conducir a patrones severos de resistencia y a formas no curables de TB ¹¹².

De acuerdo con esto, se ha observado, mediante secuenciación masiva de aislados de *M. tuberculosis* de pacientes con TB, el desarrollo de la resistencia a fármacos durante el tratamiento y la aparición de polimorfismos compensadores relacionados con dichas resistencias ^{113–116}. En un estudio llevado a cabo en Argentina sobre un brote de cepas pre-XDR, se examinaron 252 aislamientos y se determinó que el progenitor responsable del brote adquirió resistencia a la isoniazida, estreptomicina y rifampicina alrededor de 1973. Además, hacia 1979, la cepa desarrolló resistencia a otras tres drogas. Estos hallazgos revelan una persistente circulación de la cepa durante un período de cuatro décadas ¹¹⁷. Por otra parte, modelos matemáticos han inferido que aun con los planes de contingencia propuestos por la organización mundial de la salud (OMS) es muy probable que cepas XDR y MDR, sean cada vez más frecuentes entre la población con TB dentro de algunos años ¹¹⁸.

La OMS clasifica la resistencia a fármacos de TB de la siguiente manera ¹¹⁹.

- Resistencia (RR): resistencia, al menos, a rifampicina.
- Multi-drogoresistencia (MDR): resistencia combinada a isoniazida y rifampicina.
- Pre-Extremo-resistencia (pre-XDR): resistencia a isoniazida y rifampicina, más resistencia a cualquier fluoroquinolona.
- Extremo-resistencia (XDR): pre-XDR y resistencia a al menos uno de los antituberculosos del grupo A: bedaquilina y/o linezolida,

Fluoroquinolona: Se refiere a levofloxacino y moxifloxacino que son las utilizadas actualmente en el tratamiento de la TB ¹¹³.

Aplicaciones de la secuenciación masiva de genomas

Las aplicaciones más comunes de la secuenciación masiva de genomas (WGS) en TB están relacionadas con la vigilancia epidemiológica, la tipificación de especie y linaje, la generación de perfiles de resistencia y sensibilidad a fármacos y la detección de la transmisión ¹²⁰ (Figura 1.5).

La WGS se ha empleado para investigar la evolución del MTBC, clasificarlo en linajes y sublinajes ¹²¹, así como identificar regiones genómicas relevantes para el desarrollo de pruebas diagnósticas ¹²². Además, se ha empleado para estudiar la evolución de mutaciones en cepas MDR y XDR ^{123,124,125}, describir la transmisión de estas cepas ¹²⁶ y estudiar e identificar factores de riesgo asociados ¹²⁷. De igual manera, se ha usado para comprender el ritmo y el patrón de evolución molecular *in vivo* de la bacteria durante la transmisión ¹²⁸.

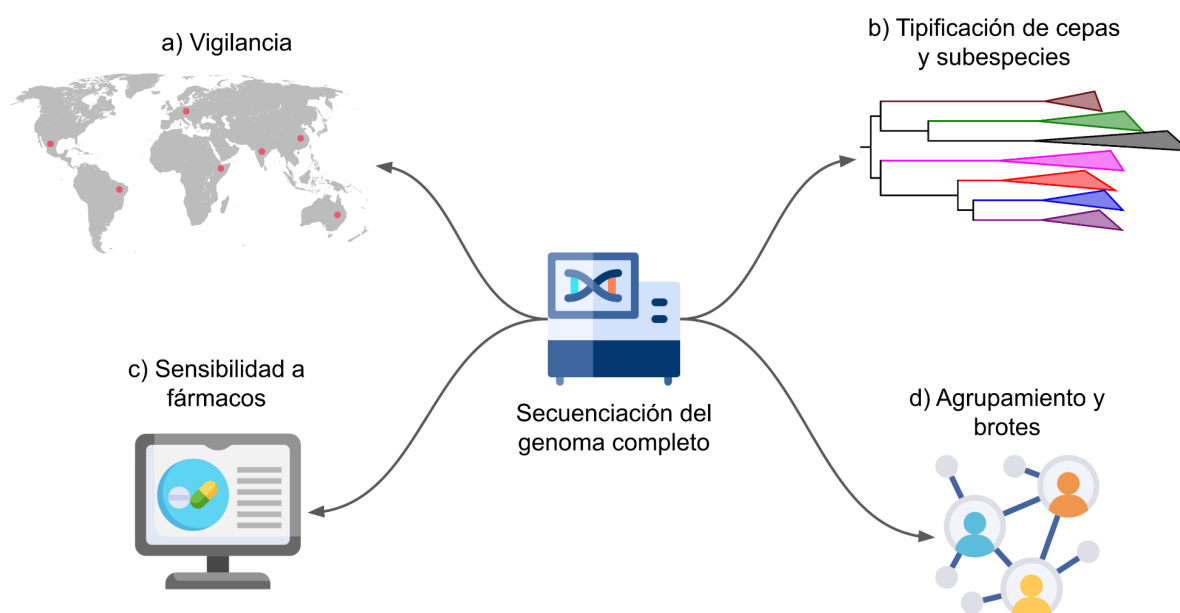


Figura 1.5. Principales aplicaciones de la secuenciación masiva de genomas de *M. tuberculosis*.

a) Vigilancia epidemiológica. b) Determinación de la especie o subespecie de los aislados. c) Determinación de los perfiles de resistencia/sensibilidad a fármacos. d) Identificación de grupos de transmisión y brotes. Adaptada de (¹²⁹).

Debido a lo anterior, la técnica se ha trasladado en corto tiempo del laboratorio de investigación a la práctica clínica ^{130–132}. Además de tener importantes implicaciones en la epidemiología y salud pública ¹³³. Existen reportes de casos en los que la secuenciación se ha

realizado directamente de esputo, ayudando en el diagnóstico y tratamiento de pacientes ¹³⁴, aunque los resultados aún no son lo suficientemente robustos para su uso clínico ¹³⁵. En la actualidad existen proyectos para utilizar la secuenciación de genomas en el diagnóstico rutinario de TB y otras enfermedades, diferentes gobiernos han sido pioneros en su implementación ^{136–138}.

Por otra parte, se han desarrollado diferentes propuestas a nivel global para construir plataformas de datos derivados de la secuenciación de genomas de aislados de *M. tuberculosis*, las cuales tiene como finalidad recabar información y emplearla en la profundización de estudios de resistencia, virulencia, epidemiología molecular, diagnóstico, entre otros aspectos ^{139,140}.

En relación con la vigilancia epidemiológica de enfermedades infecciosas, ésta desempeña un papel fundamental en los sistemas de salud pública, ya que permite detectar de manera oportuna nuevos brotes, microorganismos o variantes en una región, lo que a su vez puede reducir la mortalidad y los costos de atención médica ¹⁴¹. En este sentido, la WGS ha demostrado ser altamente efectiva en la detección de variantes en diferentes microorganismos, incluido *M. tuberculosis* ¹⁴². Su combinación con datos epidemiológicos, como la temporalidad y la geolocalización, la convierte en una poderosa herramienta para la trazabilidad y el control de enfermedades infecciosas ^{71,143}.

En cuanto a los métodos de tipificación de *M. tuberculosis*, en el pasado se han utilizado técnicas como los fragmentos de restricción de longitud polimórfica (RFLPs) IS6110 ⁷⁷, el número variable de repeticiones en tándem (MIRU-VNTR) ^{144,145} o el espigotipado ¹⁴⁶. Las cuales basan su clasificación en regiones específicas del genoma. La WGS ha revolucionado la caracterización molecular de *M. tuberculosis* ¹⁴⁷. En comparación con los métodos previos, que examinan <1% del genoma de la bacteria, la secuenciación masiva inspecciona >90% del genoma, permitiendo una tipificación más precisa de la o las cepas presentes en las muestras clínicas ^{147,148}. La WGS presenta diferentes ventajas técnicas al compararla con los métodos anteriores de tipificación ³⁵, es posible obtener la secuencia de un genoma completo de *M. tuberculosis* en horas y a bajo costo ^{123,130}. El análisis de las secuencias proporciona un gran número de datos, tanto SNPs como InDels, a partir de los cuales se pueden identificar y analizar marcadores de virulencia, predecir resistencias y realizar estudios de epidemiología

molecular ^{54,149}. Además, posee una alta resolución en la identificación de linajes y sublinajes de *M. tuberculosis* y especies relacionadas, a partir de SNPs específicos ²⁹.

Asimismo, la secuenciación masiva de genomas también ha demostrado su utilidad en el diagnóstico de resistencia a fármacos ^{132,150,151}, ya que puede identificar de manera simultánea mutaciones, inserciones y deleciones en todos los genes asociados con la resistencia a fármacos en aislados de *Mycobacterium tuberculosis*. Esto proporciona un perfil de resistencia/susceptibilidad a los medicamentos, lo cual es crucial para un manejo clínico adecuado de los pacientes con tuberculosis ^{123,152,153}.

En estudios de transmisión de la TB, la WGS ha sido empleada en la identificación de grupos de transmisión basada en la diferencia de distancias genéticas entre los aislados de *M. tuberculosis* ^{154,155}. De acuerdo con los umbrales estimados de ≤ 5 SNPs para la detección de eventos de transmisión reciente y ≤ 12 SNPs para eventos más antiguos ¹⁵⁶, recomendados y empleados en diversos protocolos ^{129,157}. La combinación de datos epidemiológicos, espaciales y genómicos permite identificar brotes de TB ^{158,159} y comprender la dinámica de la transmisión a nivel poblacional ^{160,161}.

Los estudios de interacción hospedador-patógeno han sugerido una adaptación de ciertos linajes del complejo MTBC a etnias específicas ⁴³. Esto respalda la idea de que el complejo puede considerarse como una serie de ecotipos adaptados a diferentes hospedadores ¹⁶². Aunque se requieren nuevos enfoques experimentales para clasificar a nivel genómico los posibles ecotipos/genotipos que infectan a poblaciones específicas y para explicar las diferencias genéticas que controlan o influyen en esta supuesta adaptación.

2. Objetivos

Objetivo general

La secuenciación masiva de genomas ha destacado por su utilidad en el diagnóstico, vigilancia, predicción de resistencia a antibióticos y estudio de la transmisión de enfermedades infecciosas, incluyendo la tuberculosis. Esta tesis tiene como objetivo aplicar esta poderosa técnica al estudio de la tuberculosis en la Comunidad Valenciana, abordando aspectos clave como la vigilancia de la enfermedad en la región, la comprensión del proceso de infección de *M. tuberculosis* y la exploración de la relación evolutiva patógeno-humano.

En el transcurso de este estudio, hemos empleado la secuenciación masiva para analizar los casos de TB en la Comunidad Valenciana durante un período de cinco años (2014-2019), combinando datos epidemiológicos proporcionados por la Dirección General de Salud Pública de la Comunidad Valenciana con información genómica generada en nuestro laboratorio. Nuestra hipótesis sostiene que esta integración de datos puede proporcionar una perspectiva más completa de la situación de la enfermedad en la región y enriquecer nuestra comprensión de la TB.

Este trabajo persigue principalmente destacar la importancia de la secuenciación masiva en la vigilancia y análisis de la transmisión de la tuberculosis, con el objetivo de fomentar su aplicación tanto en la investigación científica como en la práctica clínica y la salud pública en general. A través de los resultados obtenidos, se aspira a contribuir al fortalecimiento del control de la tuberculosis y a la reducción de su incidencia.

Objetivos específicos

Esta tesis tiene como objetivos específicos la utilización de la secuenciación masiva de genomas para llevar a cabo la vigilancia genómica de la tuberculosis en la Comunidad Valenciana, la identificación de posibles genotipos endémicos en la región y el análisis de la transmisión de la TB entre pares donante-receptor. A continuación, se detallan estos objetivos.

- Describir y analizar la epidemiología de la tuberculosis (TB) en la Comunidad Valenciana mediante la secuenciación masiva de genomas (WGS) (**Capítulo 3**).
- Caracterizar los aislados clínicos del complejo *Mycobacterium tuberculosis* (MTBC) por medio de secuenciación masiva, incluyendo el perfil de resistencia/sensibilidad a fármacos, la estructura poblacional de la bacteria, grupos de transmisión y factores de riesgo asociados (**Capítulo 3**).
- Desarrollar una metodología para identificar genotipos endémicos integrando datos epidemiológicos y de secuenciación masiva (**Capítulo 4**).
- Identificar posibles genotipos endémicos presentes en la Comunidad Valenciana (**Capítulo 4**).
- Desarrollar una metodología basada en datos epidemiológicos, clínicos y genómicos para identificar pares de transmisión donante-receptor de MTBC (**Capítulo 5**).
- Estimar el número de bacilos necesarios para iniciar una nueva infección de TB en humanos (**Capítulo 5**).

Esquema de la tesis

En el primer capítulo, se proporciona una introducción general a la tesis, donde se delinean los objetivos principales tratados en este trabajo, los cuales se detallan en el segundo capítulo.

Los objetivos de este trabajo de tesis se estructuran en tres capítulos que profundizan en distintos aspectos de la epidemiología de la tuberculosis en la Comunidad Valenciana.

En el tercer capítulo, llevamos a cabo un análisis exhaustivo de los datos clínicos provenientes de los sistemas de vigilancia de la TB en esta región. Exploramos posibles factores de riesgo relacionados con la tuberculosis y su transmisión, además de realizar un análisis de epidemiología genómica con el propósito de identificar perfiles de resistencia a fármacos, grupos de transmisión y relaciones filogenéticas.

En el cuarto capítulo, combinamos los datos previamente obtenidos junto con secuencias globales disponibles en repositorios públicos, para identificar los genotipos endémicos en la Comunidad Valenciana y evaluar su relación con los patrones de transmisión.

En el quinto capítulo, desarrollamos una metodología para la selección de parejas de transmisión (donante-receptor) empleando datos de secuenciación masiva de genomas. Esta metodología se basa en información epidemiológica, la distancia genética, el tamaño de los clústeres de transmisión y los datos de rastreo de contactos. Además, aplicamos un enfoque estadístico para estimar el número de bacilos que podrían iniciar una infección en estas parejas de transmisión.

Finalmente, en el sexto capítulo se abordan los resultados obtenidos a nivel global, mientras que el séptimo se presentan las conclusiones finales de la presente investigación.

3. Epidemiología molecular de la tuberculosis en la Comunidad Valenciana, España

Introducción

La Comunidad Valenciana presenta una baja incidencia de tuberculosis, registrando tan solo 7.5 casos por cada 100,000 habitantes en 2020. Este dato se acerca al promedio nacional de España, que es de 8.2 casos. Sin embargo, está significativamente por debajo de la incidencia en la región europea, que asciende a 25 casos por 100,000 habitantes. Además, se distingue aún más de las tasas en Las Américas (30), el Sudeste Asiático (234) y África (212) ¹⁶³.

El control de la tuberculosis, tanto en la Comunidad Valenciana como a nivel global, se enfoca en la detección, tratamiento y seguimiento de nuevos casos y brotes de la enfermedad. En general, la identificación de nuevos casos se basa en el diagnóstico de aquellos individuos que presentan síntomas y buscan atención médica en centros de salud, además de la pronta identificación y diagnóstico de sus contactos cercanos.

El estudio de contactos desempeña un papel crucial en la identificación de posibles nuevos casos de tuberculosis y en la trazabilidad de las cadenas de transmisión. No obstante, en áreas con alta incidencia y densidad poblacional, obtener una imagen precisa de la transmisión se torna desafiante. Detectar todos los casos resulta complicado, lo que dificulta asignar un individuo a una cadena de transmisión y definir la dirección de la transmisión. A pesar de estas limitaciones, el estudio de contactos sigue siendo una estrategia valiosa en el control de la tuberculosis.

Dentro del control de la tuberculosis, también es relevante detectar y tratar de manera adecuada a los pacientes que portan cepas resistentes a medicamentos. Principalmente, aquellos que son resistentes a la rifampicina (RR) o que presentan resistencia a múltiples medicamentos (MDR), ya que la transmisión de estas cepas puede dar lugar a brotes con estas características ¹²⁶, complicando significativamente el tratamiento, empeorando el pronóstico del paciente y generando costos considerables para los sistemas de salud.

En resumen, el control efectivo de la tuberculosis requiere un diagnóstico y tratamiento rápidos y precisos, junto con la detección de resistencia a medicamentos. La secuenciación masiva de genomas emerge como una herramienta poderosa, útil en el diagnóstico, predicción de resistencia a fármacos y vigilancia epidemiológica, así como para el estudio de la transmisión y la evolución del patógeno.

Este capítulo se centra en un estudio de epidemiología genómica mediante WGS realizado en la Comunidad Valenciana entre 2014 y 2018. Su objetivo principal fue estimar la tasa de transmisión de la tuberculosis, explorar factores de riesgo y analizar la prevalencia de resistencia a antibióticos en la región. Este estudio complementa y expande investigaciones previas llevadas a cabo por nuestro grupo de investigación ¹⁵⁴.

En el desarrollo de nuestro estudio, partimos de una colección de aislados secuenciados y llevamos a cabo un análisis descriptivo comparativo de las variables y factores de riesgo asociados con la tuberculosis en la Comunidad Valenciana en comparación con la población total de casos de TB.

Posteriormente, llevamos a cabo la reconstrucción filogenética de nuestra colección, lo que nos permitió visualizar su estructura poblacional. Incorporando a esta filogenia los resultados de un estudio de agrupamiento genómico, previamente efectuado.

A continuación, fusionamos los datos epidemiológicos disponibles con los generados por el análisis genómico para realizar un análisis descriptivo de las principales variables y factores de riesgo relacionados con los casos pertenecientes a agrupamientos genómicos. Finalmente, ejecutamos una predicción de la sensibilidad o resistencia a fármacos en todos los aislados de la colección y describimos las mutaciones asociadas a estas resistencias.

Esta metodología nos brindó una comprensión completa y detallada de diversos aspectos de la tuberculosis en nuestra población de estudio. Los resultados subrayan de manera convincente la eficacia de la secuenciación masiva de genomas en la predicción de la resistencia a fármacos y en la identificación de grupos de transmisión a nivel poblacional. Estos análisis de datos y resultados han servido como base para el desarrollo de los capítulos cuatro y cinco.

Métodos

Población de estudio

Desde 2006 la Comunidad Valenciana cuenta con una red de vigilancia microbiológica ¹⁶⁴, la cual recolecta datos generados por los laboratorios de microbiología de los hospitales valencianos y cuyo objetivo es detectar rápidamente la circulación de diferentes agentes etiológicos, así como definir patrones de resistencia a los antimicrobianos.

La Comunidad Valenciana presenta una baja incidencia de TB, con aproximadamente 7.5 casos por cada 100,000 habitantes. Está formada por las provincias de Alicante, Castellón y Valencia, siendo Alicante la de menor incidencia y Valencia la de mayor. Aunque la incidencia ha mostrado una tendencia a la baja en la última década, se ha observado una desaceleración notoria y un ligero aumento de casos en la población extranjera desde 1998.

La responsabilidad de la vigilancia local de la tuberculosis recae en la Agencia Regional de Salud Pública (DGSP: Dirección General de Salud Pública), que utiliza el rastreo de contactos como método clave para detectar la transmisión de la tuberculosis. En este proceso, cada paciente con un diagnóstico positivo completa un cuestionario estandarizado a nivel local, lo que permite identificar casos epidemiológicamente relacionados, que potencialmente pueden haberse infectado. Además, esta encuesta recopila datos clínicos, microbiológicos y demográficos relevantes.

Este estudio recibió la aprobación del Comité de Ética para la Investigación Clínica de la Agencia Regional de Salud Pública de Valencia (Comité Ético de Investigación Clínica de la Dirección General de Salud Pública y Centro Superior de Investigación en Salud Pública). No se requirió el consentimiento informado, ya que la tuberculosis forma parte del programa de vigilancia obligatoria de enfermedades transmisibles a nivel regional. Se garantizó la anonimización de toda la información personal, sin retener datos que pudieran identificar individualmente a los participantes.

Colección de aislados y extracción de DNA

Para nuestro estudio, recopilamos 1279 muestras de individuos diagnosticados con TB durante un período de cinco años (2014-2018), en 18 hospitales de la Comunidad Valenciana.

Esta cantidad representó el 74% (1279/1727) de todos los casos con cultivos positivos reportados por la DGSP.

El diagnóstico de estas muestras de TB se lleva a cabo en los hospitales mediante cultivo líquido en tubo indicador de crecimiento de micobacterias (MGIT) o en medio sólido *Löwenstein-Jensen* (LJ). Los cultivos positivos se envían a la Unidad Mixta "Infección y Salud Pública" (FISABIO-CSISP). Luego, el cultivo (1 ml en MGIT) se inactiva a 90 °C durante 30 minutos y se extrae el ADN bacteriano automáticamente mediante el sistema *EasyMag* (bioMérieux), siguiendo las indicaciones del fabricante. Para conservar la colección de aislados de la Comunidad Valenciana, se guarda una alícuota del cultivo en 1 ml de glicerol (20%) a -80 °C. Todos los procedimientos se realizaron en un laboratorio de bioseguridad nivel 3 (BSL3) de acuerdo con las recomendaciones de la OMS.

Además, disponemos de datos epidemiológicos para el 93.26% de los casos con cultivos positivos secuenciados (1191/1279), que incluyen información sobre lugar de nacimiento, edad, sexo y factores de riesgo como alcoholismo o diabetes.

Secuenciación del genoma completo

El DNA disponible de 1279 aislados, correspondientes al 74% de los cultivos positivos identificados en el periodo de estudio, fue secuenciado a una profundidad objetivo de 100X. Las bibliotecas se construyeron con Nextera XT/Nextera Flex DNA Library Prep Kit (Illumina). Se secuenciaron con la tecnología Illumina (MiSeq y HiSeq), generando lecturas de longitud 2x300pb/2x150pb respectivamente.

Análisis bioinformático

En la Unidad de Genómica de la Tuberculosis se ha diseñado un “pipeline” bioinformático adaptado al análisis de los genomas de MTBC. Incluye un conjunto de programas, herramientas y código, ha sido denominado “Thepipeline” y se organiza en los siguientes módulos: “Trimclean”, “Kraken”, “Mapping”, “Coverage”, “Calling” y “Annotation filter”. La totalidad del código y parámetros utilizados en “Thepipeline” se encuentran disponibles en el siguiente link: “<https://gitlab.com/tbgenomicsunit/ThePipeline>”.

El primer módulo “Trimclean” trabaja con los archivos “Fastq” generados por el secuenciador, utiliza el programa Trimmomatic ¹⁶⁵, para eliminar lecturas menores a 50 pares de bases (pb) y limpiar lecturas eliminando regiones de baja calidad, excluyendo los adaptadores y otras secuencias específicas del proceso de secuenciación Illumina ¹⁶⁵.

El módulo “Kraken” ¹⁶⁶ evalúa las posibles contaminaciones clasificando taxonómicamente las lecturas con una base de datos personalizada, que comprende todas las secuencias de bacterias, arqueas, virus, protozoos, plásmidos y hongos en RefSeq (versión 78), además del genoma humano (GRCh38, Ensembl versión 81); finalmente se eliminan todas aquellas lecturas que no corresponden a *M. tuberculosis*.

El módulo “Mapping” emplea los programas BWA (Burrows-Wheeler Aligner) ^{167,168}, Picard ¹⁶⁹ y Bedtools ¹⁷⁰ para realizar el mapeo de las secuencias previamente filtradas por “Kraken” respecto al genoma de referencia del ancestro de MTBC ¹⁶⁷.

Una vez mapeadas las lecturas, el módulo “Coverage” calcula la media y la mediana de la cobertura total de cada genoma, así como el parámetro profundidad, el cual indica el número de veces que una posición genómica es leída.

El módulo “Calling” consta de los programas VarScan ¹⁷¹ y GATK ¹⁷², utilizados para identificar posiciones genómicas, inserciones y deleciones que difieren del genoma de referencia. Las posiciones genómicas que difieren del genoma de referencia son denominadas SNPs. Este módulo selecciona los SNPs que tienen una profundidad mínima de 10X y una calidad de mapeo superior a 20.

Dentro de los SNPs identificados, existen dos tipos: los SNPs fijados (fSNPs), que tienen una frecuencia $\geq 90\%$. Estos se emplean en análisis como la reconstrucción filogenética y la identificación de clústeres de transmisión. Por otro lado, los SNPs variables (vSNP) tienen una frecuencia entre 10-89%, y se emplean en la detección de resistencia a fármacos. La identificación de variantes se realizaron siguiendo los estándares validados por nuestro grupo (http://tgu.ibv.csic.es/?page_id=1794) ¹²⁹.

Finalmente, el módulo “Annotation filter” elimina los SNPs ubicados en regiones difíciles de mapear, las secuencias repetitivas y los genes PPE/PE-PGRS. .

Estructura poblacional y agrupamiento genómico

Para determinar la estructura poblacional de los aislados clínicos, realizamos un alineamiento múltiple utilizando fSNPs y construimos una filogenia de *maximum likelihood* (ML) con el programa IQ-TREE 2¹⁷³. Utilizamos el modelo *GTR* y realizamos 1,000 réplicas *bootstrap* para asegurar la robustez del análisis. Posteriormente, clasificamos los aislados de MTBC en linajes y sublinajes utilizando una lista previamente definida de 62 SNPs^{29,174}.

Para evaluar la transmisión basada en el agrupamiento genómico (clústeres genómicos), realizamos un alineamiento múltiple de las secuencias, excluyendo las posiciones relacionadas con la resistencia a fármacos. Calculamos las distancias entre las muestras en pares utilizando el paquete "*ape*" en el lenguaje de programación R. Definimos los clústeres genómicos como grupos de, al menos, dos cepas de MTBC que compartían una distancia genética específica y evaluamos los clústeres utilizando tres puntos de corte de distancias genéticas (0, 5, 10 y 12 SNPs).

Para determinar la relación genética de los aislados clínicos en situaciones donde se detectaron clústeres no monofiléticos, re-evaluamos los aislados excluidos de los clústeres utilizando el siguiente criterio: si la filogenia mostraba que estaban rodeados por otros aislados en un clúster, se integraban en ese clúster. Este proceso generó una variación mínima en el recuento total de aislados en los clústeres.

Análisis de variables y factores de riesgo relacionados con el agrupamiento genómico

Se realizó un análisis descriptivo de la población de estudio, que consistió en la recolección y organización de datos epidemiológicos disponibles. Se trabajó con un total de 1191 aislados, que representan el 93.2% de todos los aislados secuenciados. Se calcularon las frecuencias absolutas y relativas para cada variable, y se presentó la información en tablas y gráficos. Además, se incorporaron datos obtenidos a través del análisis de secuenciación de genomas.

Para evaluar la asociación entre los factores de riesgo y la agrupación genómica en clúster, se utilizó el cálculo de la razón de momios (RM). También se llevó a cabo un análisis multivariado, mediante un modelo de regresión logística, para evaluar la interacción entre

variables como la edad, el sexo, el lugar de nacimiento, la tinción, la localización, el alcoholismo y el estado de privación de libertad, con la probabilidad de pertenecer a un clúster de transmisión.

Predicción de la resistencia a fármacos

En el estudio de predicción de resistencia a fármacos, empleamos el catálogo propuesto por la OMS para identificar variantes asociadas con la resistencia a medicamentos de primera y segunda línea (Tabla 3) ¹⁷⁵. Anteriormente, utilizábamos catálogos validados de las plataformas ReSeqTB ¹⁵⁴ y PhyResSE ¹⁷⁶. Consideramos únicamente las mutaciones catalogadas en los niveles uno (asociado a resistencia) y dos (asociado a resistencia de manera provisional) ¹⁷⁷ del catálogo, excepto la posición 1674481 Ser94Ala en el gen *inhA*, que se tuvo en cuenta en relación con la resistencia a isoniazida por razones detalladas en la sección de resultados. Finalmente, las muestras se clasificaron según su perfil de resistencia como sensibles, resistentes, resistentes a rifampicina (RR), multidrogoresistentes (MDR) o extremo-resistentes (XDR).

En cuanto a la evaluación de la predicción de resistencia a fármacos, utilizamos la tabla tetracórica para calcular la sensibilidad, especificidad, valor predictivo positivo y valor predictivo negativo de los medicamentos: rifampicina, isoniazida, etambutol y pirazinamida (Tabla 3.11). Todos los análisis estadísticos y gráficos se llevaron a cabo utilizando la colección de paquetes "Tydiverse" en el lenguaje R ¹⁷⁸.

Resultados

Población de estudio

Durante el periodo comprendido entre 2014 y 2018, la DGSP registró un total de 2206 casos de tuberculosis, de los cuales 1727 (78.2%) fueron cultivo-positivos. Realizamos secuenciación de genoma completo en 1279 individuos, equivalente al 74% de todos los pacientes con resultados positivos en cultivo (1279/1727). Contamos con información epidemiológica para 1191 individuos, lo cual constituye un subconjunto representativo de la población con tuberculosis (Figura suplementaria 3.0.).

En la Tabla 3.1 se ofrece una descripción detallada de las características demográficas y clínicas, tanto de los casos reportados como de aquellos sometidos a secuenciación de genoma completo, con metadatos disponibles. Cabe resaltar que las características clínicas y demográficas en este conjunto de datos son similares a las del total de casos.

Tabla 3.1. Características demográficas de los individuos con tuberculosis positiva analizados en el estudio. La columna "Total de pacientes" refleja los valores obtenidos mediante el sistema de vigilancia local, mientras que "Pacientes con WGS" identifica los casos de TB que fueron secuenciados y que cuentan con metadatos disponibles

Variables		WGS pacientes (n = 1191)		Total de pacientes (n = 2206)	
	Edad				
	≤ 15	34	2.9%	147	6.7%
	16-24	103	8.6%	195	8.8%
	25-44	469	39.4%	847	38.4%
	45-64	377	31.7%	640	29.0%
	≥ 65	208	17.5%	377	17.1%
	Sexo				
	Hombre	769	64.6%	1379	62.5%
	Mujer	422	35.4%	827	37.5%
	Lugar de nacimiento				
	Español	734	61.6%	1378	62.5%
	Extranjero	457	38.4%	828	37.5%
	Tinción de esputo ^a				
	Positivo	673	56.8%	1002	47.5%
	Negativo	511	43.2%	1107	52.5%
Localización ^b					
Pulmonar	983	82.5%	1644	74.6%	
Extrapulmonar	208	17.5%	559	25.4%	
Factores de riesgo		WGS pacientes (n = 1191)		Total de pacientes (n = 2206)	
	Alcoholismo ^c				
	Si	241	21.5%	351	17.0%
	No	878	78.5%	1709	83.0%
	PDU ^d				
	Si	40	3.6%	59	2.9%
	No	1073	96.4%	2010	97.1%
	PPL ^e				
	Si	63	5.4%	86	4.0%
	No	1104	94.6%	2077	96.0%
	Diabetes ^f				
	Si	120	10.2%	218	10.1%
	No	1056	89.8%	1947	89.9%
	VIH Positivo ^g				
	Si	73	6.6%	113	5.6%

	No	1028	93.4%	1923	94.4%
Exclusión social ^h					
	Si	143	12.5%	218	10.3%
	No	1004	87.5%	1903	89.7%
Trabajador sanitario ⁱ					
	Si	22	1.9%	46	2.1%
	No	1163	98.1%	2140	97.9%
Retraso diagnóstico ^j					
	Si	312	27.5%	560	27.4%
	No	821	72.5%	1485	72.6%
Neoplasia ^k					
	Si	102	8.8%	178	8.3%
	No	1057	91.2%	1960	91.7%
Paciente ambulatorio ^l					
	Si	53	4.5%	81	3.7%
	No	1128	95.5%	2096	96.3%
Hospitalización ^m					
	Si	892	75.0%	1554	70.7%
	No	298	25.0%	645	29.3%
Asilo de ancianos ⁿ					
	Si	24	2.0%	38	1.7%
	No	1154	98.0%	2141	98.3%

^a Datos desconocidos en 97 individuos

^b Datos desconocidos en 3 individuos

^c Datos desconocidos en 146 individuos

^d PDU parenteral drug user (usuario de drogas por vía parenteral). Datos desconocidos en 137 individuos

^e PPL Persona privada de su libertad. Datos desconocidos en 43 individuos

^f Datos desconocidos en 41 individuos

^g VIH virus de inmunodeficiencia humana. Datos desconocidos en 170 individuos

^h Datos desconocidos en 85 individuos

ⁱ Datos desconocidos en 20 individuos

^j Datos desconocidos en 161 individuos

^k Datos desconocidos en 68 individuos

^l Datos desconocidos en 29 individuos

^m Datos desconocidos en 7 individuos

ⁿ Datos desconocidos en 27 individuos

Al explorar las características y factores de riesgo en los 1191 casos, se observan los siguientes datos estadísticos (ver Tabla 3.1). La distribución por sexo reveló que el 64.6% (769) de los casos de TB correspondía a hombres, mientras que el 35.4% (422/1191) eran mujeres. En relación con la edad, el rango abarcó desde 5 meses hasta 100 años, con una edad promedio de 45 años y una mediana de 44 años.

Del total de casos de tuberculosis estudiados, el 61.6% (734/1191) correspondió a individuos nacidos en España, mientras que el 38.4% (457/1191) eran casos de personas nacidas en otros países (ver Tabla 3.1). Entre los casos de no nacidos en España, los países de procedencia más comunes fueron Rumania (125 casos), Marruecos (80 casos), Pakistán (36 casos), Bolivia (31 casos) y Ecuador (19 casos).

En nuestra muestra de 1191 casos, las proporciones de los factores de riesgo clínicos fueron las siguientes: el 21.5% (241/1119) se vio afectado por el alcoholismo, el 3.6% (40/1113) correspondió a usuarios de drogas vía parenteral (PDU), el 10.2% (120/1176) tenía diabetes (DM), el 6.6% (73/1101) era personas que viven con el virus de la inmunodeficiencia humana (VIH), el 27.5% (312/1133) experimentó un retraso en el diagnóstico y el 8.8% (102/1159) tenía neoplasia.

En cuanto a los factores de riesgo socio-demográficos, la distribución fue: el 5.4% (63/1167) eran personas privadas de libertad (PPL), el 12.5% (143/1147) eran excluidos sociales, el 1.9% (22/1185) pertenecían al personal sanitario, el 4.5% (53/1181) eran pacientes transeúntes, el 75% (892/1190) requirió ingreso hospitalario y el 2% (24/1178) vivía en residencias de ancianos.

Estructura poblacional y agrupamiento genómico

La estructura poblacional de los aislados de *M. tuberculosis* se exploró mediante la reconstrucción filogenética de máxima verosimilitud (ML) utilizando 1277 aislados secuenciados. Este número se obtuvo después de excluir dos aislados (1277/1279) debido a la detección de secuencias pertenecientes a más de dos linajes o más de tres sublinajes, lo cual podría comprometer el análisis.

Aunque la región de Valencia se considera un área con baja carga de tuberculosis, nuestro análisis identificó cuatro muestras con posibles infecciones mixtas. De éstas, dos fueron excluidas del análisis filogenético, mientras que las cuatro se incluyeron en el análisis de predicción de resistencias a fármacos.

En lo que respecta a la distribución de linajes, se destacó un claro predominio de cepas del linaje L4 (92%). Le siguieron en menor proporción cepas de L3 (2.6%), L2 (2.4%), L5 (0.5%), L6 (0.5%) y L1 (0.4%). Además, se identificaron cepas de origen animal (1.5%) (Figura 3.1 A).

Los clústeres fueron definidos mediante distintos umbrales de distancia genética entre pares de aislados; evaluamos umbrales de 0, 5, 10, 12 y 15 SNPs. Nuestros resultados revelaron la formación de 106, 174, 181, 186 y 187 clústeres de transmisión, respectivamente. Los aislados en los clústeres correspondientes a los tres primeros umbrales se representan como anillos grises alrededor de la filogenia en la Figura 3.1 A.

Al aplicar un umbral de 0 SNPs, se identificaron 289 (23%) aislados en 106 clústeres. A medida que el umbral se incrementó gradualmente, en el rango de 1 a 5 (SNPs), esta cifra aumentó a 557 (44%), señalando una tendencia hacia la agrupación. Con un umbral de 6 a 10, el número de aislados ascendió a 630 (49%). Al establecer un umbral más alto, en el rango de 11 a 12 (SNPs), se observó un ascenso adicional, llegando a 658 (52%) aislados en clústeres. En el rango superior, de 13 a 15 (SNPs), el valor se elevó a 690 (54%). Sin embargo, en el contexto particular del "Estudio de contactos", únicamente 168 (13%) de los aislados se agruparon en clústeres (Figura 3.1 B), lo que difiere de los resultados obtenidos mediante el método de agrupamiento genómico empleado.

Dentro del espectro de distribución observado en la figura 3.1 B, la elección del umbral de 10 SNPs surgió como una opción equilibrada en el análisis. Al establecer este umbral, la proporción de aislados en clústeres se elevó a 630 (49%), presentando un aumento significativo en relación con los umbrales anteriores y una variación moderada con los umbrales superiores (ver Figura 3.1 B).

Es relevante resaltar que la identificación genómica de clústeres de transmisión, incluso en la resolución más baja de 0 SNP, reveló un total de 289 aislados en clúster (23%). Por otro lado,

el estudio de contactos logró identificar 168 aislados en clúster (13%). Esta diferencia de 10% en la identificación de clústeres subraya la mayor eficacia de la WGS en detectar la transmisión en comparación con el enfoque de estudio de contactos (ver Figura 3.1 B).

Dentro de este contexto, se procedió a calcular los tamaños y la cantidad de clústeres (utilizando el umbral de 10 SNPs). Los resultados revelaron que la mayoría de los clústeres estaban conformados por dos aislados (84 clústeres), mientras que los clústeres de 10, 12, 13, 18 y 19 aislados eran poco comunes (1 clúster cada uno). En la franja intermedia, se encontraron clústeres con 3 aislados (40 clústeres), 4 aislados (20 clústeres), 5 aislados (13 clústeres), 6 aislados (12 clústeres), 7 aislados (3 clústeres) y 8 aislados (4 clústeres). Estos hallazgos aportan una visión detallada de la composición y distribución de los clústeres bajo el umbral seleccionado (ver figura 3.1 C).

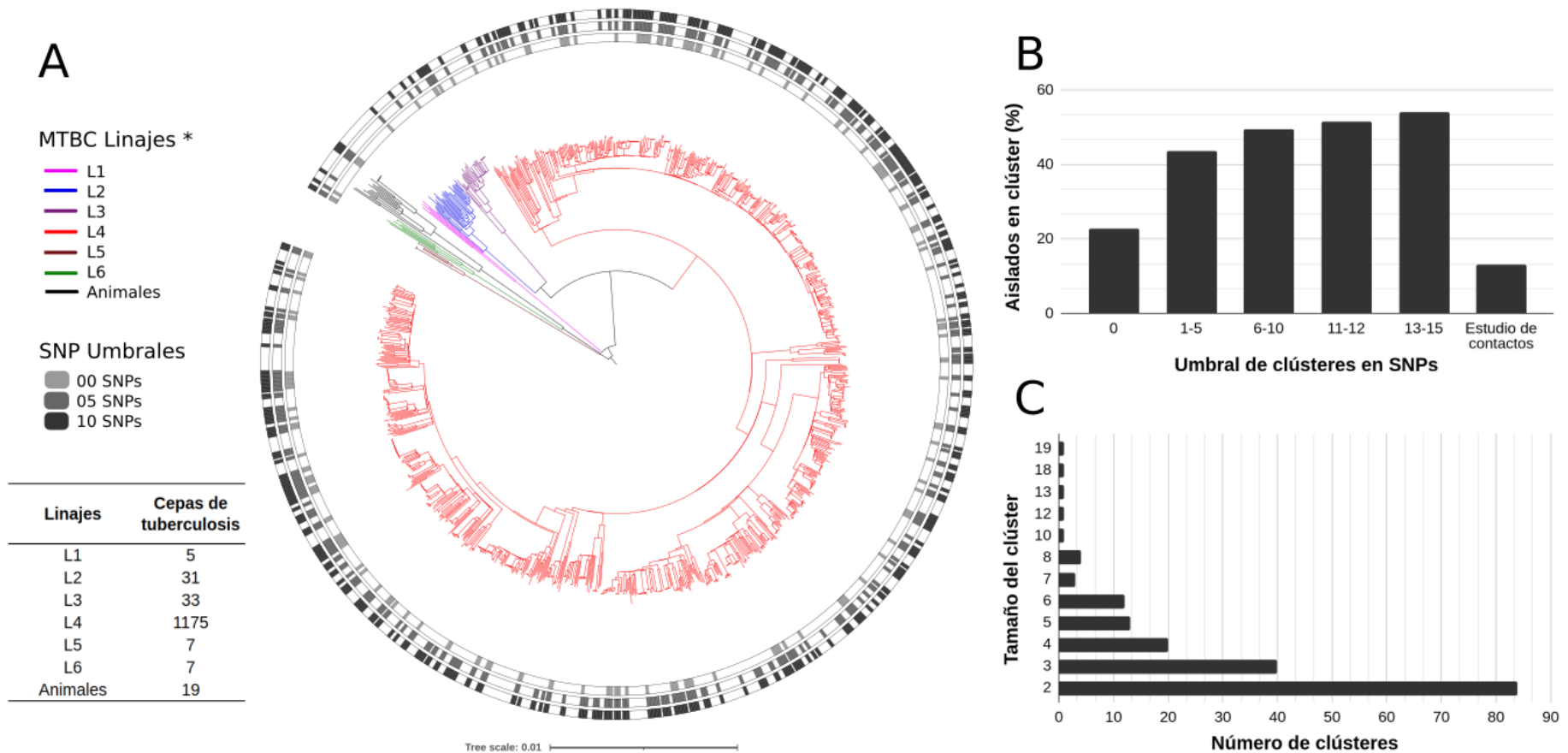


Figura 3.1. (A) Filogenia de 1277 aislados de TB recogidos durante los años 2014 - 2019. Cada anillo representa grupos genómicos detectados por diferentes umbrales de polimorfismo de un solo nucleótido (SNP) (0, 5 y 10 SNP). (B) Porcentaje de muestras dentro de los grupos para diferentes umbrales de SNP. (C) Número de clústeres genómicos por diferentes tamaños de clúster. Se utilizó un umbral de 10 SNP como estándar. No se detectaron tamaños de clúster de 9,11 y 14-17 muestras. *Nomenclatura propuesta por Comas et al., 2013³⁰

Análisis de variables y factores de riesgo relacionados con el agrupamiento genómico

Con el propósito de profundizar en la comprensión de la transmisión, se exploró la relación entre casos agrupados en clústeres y casos únicos, considerando diferentes variables en los 1191 aislados con metadatos disponibles. Los resultados de este análisis se presentan en la Tabla 3.2. Observamos que la edad desempeña un papel crucial en esta agrupación: los pacientes menores de 15 años tienen una probabilidad más alta y significativa de pertenecer a clústeres genómicos (2.4%), en comparación con los casos únicos (0.4%), con un valor $p < 0.001$. También, se destaca una disparidad entre sexos, con una inclinación más pronunciada de los hombres a estar en clústeres (32.2%) en comparación con las mujeres (15.4%), con un valor p de 0.04.

El lugar de nacimiento también muestra influencia significativa: los pacientes nacidos en España tienen más posibilidades de pertenecer a clústeres genómicos (33.9%) en comparación con aquellos nacidos en el extranjero (13.7%), con un valor $p < 0.001$. Además, ciertas variables se asocian con una tendencia más marcada a la agrupación en clústeres, como tener tinción de esputo positivo, localización pulmonar y ciertos grupos de edad.

En cuanto a los distintos factores de riesgo que pueden influir en la probabilidad de pertenecer a clústeres en comparación con ser casos únicos. Dos aspectos destacan de manera significativa: el alcoholismo y las personas privadas de su libertad (PPL). El alcoholismo se relaciona con una mayor probabilidad de pertenecer a clústeres (12.6%) en comparación con casos únicos (8.9%), con un valor $p < 0.001$. De manera similar, las PPL muestran una asociación con una mayor probabilidad de pertenencia a clústeres (3.3%) en comparación con casos únicos (2.1%), con un valor p de 0.04. Además, la tabla 3.2 explora otros factores como el uso de drogas por vía parenteral (PDU), diabetes, VIH positivo, exclusión social, ser trabajador sanitario, retraso en el diagnóstico, neoplasia, ser paciente ambulatorio, hospitalización y residencia en asilo de ancianos. Aunque algunos de estos factores también están vinculados a la pertenencia a clústeres, los resultados recalcan el impacto significativo del alcoholismo y las PPL en esta pertenencia, en comparación con casos únicos. Estos hallazgos contribuyen a la comprensión de cómo los factores de riesgo pueden influir en la propagación de la TB y subrayan la importancia de estas variables y factores de riesgo.

Tabla 3.2. Variables y factores de riesgo asociados con transmisión

Variables	Casos en clúster		Casos únicos		Razón de momios [95 % CI]	Valor <i>p</i>
Edad						
≤ 15	29	2.4%	5	0.4%	5.78 (2.20 - 15.18)	<0.001
16-24	55	4.6%	48	4.0%	1.14 (0.74 - 1.75)	0.55
25-44	235	19.7%	234	19.6%		referencia
45-64	181	15.2%	196	16.5%	0.92 (0.70 - 1.21)	0.54
≥ 65	67	5.6%	141	11.8%	0.47 (0.34 - 0.67)	<0.001
Sexo						
Hombre	383	32.2%	386	32.4%	1.28 (1.01 - 1.63)	0.04
Mujer	184	15.4%	238	20.0%		referencia
Lugar de nacimiento						
Español	404	33.9%	330	27.7%	2.21 (1.74 - 2.81)	<0.001
Extranjero	163	13.7%	294	24.7%		referencia
Tinción de esputo ^a						
Positivo	342	28.9%	331	28.0%	1.33 (1.06 - 1.68)	0.01
Negativo	223	18.8%	288	24.3%		referencia
Localización						
Pulmonar	506	42.5%	477	40.1%	2.56 (1.85 - 3.53)	<0.001
Extrapulmonar	61	5.1%	147	12.3%		referencia
Factores de riesgo	Casos en clúster		Casos únicos		Razón de momios [95 % CI]	Valor <i>p</i>
Alcoholismo ^b						
Si	141	12.6%	100	8.9%	1.65 (1.24 - 2.21)	<0.001
No	404	36.1%	474	42.4%		referencia
PDU ^c						
Si	23	2.1%	17	1.5%	1.47 (0.78 - 2.79)	0.23
No	514	46.2%	559	50.2%		referencia
PPL ^d						
Si	38	3.3%	25	2.1%	1.71 (1.02 - 2.87)	0.04
No	520	44.6%	584	50.0%		referencia
Diabetes ^e						
Si	53	4.5%	67	5.7%	0.84 (0.58 - 1.23)	0.38

	No	511	43.5%	545	46.3%		referencia
VIH Positivo ^f							
	Si	38	3.5%	35	3.2%	1.17 (0.73 - 1.88)	0.52
	No	495	45.0%	533	48.4%		referencia
Exclusión social ^g							
	Si	74	6.5%	69	6.0%	1.21 (0.85 - 1.72)	0.29
	No	472	41.2%	532	46.4%		referencia
Trabajador sanitario ^h							
	Si	14	1.2%	8	0.7%	1.95 (0.81 - 4.68)	0.13
	No	550	46.4%	613	51.7%		referencia
Retraso diagnóstico ⁱ							
	Si	144	12.7%	168	14.8%	0.94 (0.72 - 1.22)	0.63
	No	392	34.6%	429	37.9%		referencia
Neoplasia ^j							
	Si	42	3.6%	60	5.2%	0.73 (0.49 - 1.11)	0.14
	No	516	44.5%	541	46.7%		referencia
Paciente ambulatorio ^k							
	Si	31	2.6%	22	1.9%	1.57 (0.90 - 2.74)	0.11
	No	534	45.2%	594	50.3%		referencia
Hospitalización ^l							
	Si	438	36.8%	454	38.2%	1.26 (0.97 - 1.65)	0.08
	No	129	10.8%	169	14.2%		referencia
Asilo de ancianos ^m							
	Si	9	0.8%	15	1.3%	0.65 (0.28 - 1.50)	0.31
	No	554	47.0%	600	50.9%		referencia

*Pacientes con datos epidemiológicos disponibles (1191/1279)

*En azul, variables o factores de riesgo estadísticamente significativos.

^a Datos desconocidos en 7 individuos

^b Datos desconocidos en 72 individuos

^c Datos desconocidos en 78 individuos

^d Datos desconocidos en 24 individuos

^e Datos desconocidos en 15 individuos

^f Datos desconocidos en 90 individuos

^g Datos desconocidos en 44 individuos

^h Datos desconocidos en 6 individuos

ⁱ Datos desconocidos en 58 individuos

^j Datos desconocidos en 161 individuos

^k Datos desconocidos en 68 individuos

^l Datos desconocidos en 29 individuos

^m Datos desconocidos en 7 individuos

Ante la detección de OR significativos en distintas variables y factores de riesgo, y con la finalidad de comprender cómo estas variables interactúan y afectan la probabilidad de pertenecer a un clúster genómico, se tomó la decisión de realizar un análisis multivariado. Empleamos un modelo de regresión logística para evaluar la existencia de interacciones entre variables como edad, sexo, lugar de nacimiento, tinción de esputo, localización, alcoholismo y personas privadas de su libertad (Tabla 3.3).

Para construir el modelo, se empleó el método *stepwise*, un enfoque de selección de variables paso a paso. Este método implica una selección rigurosa de variables independientes basada en criterios específicos. Iniciando con un modelo que incorpora todas las variables disponibles, el algoritmo evaluó de manera iterativa si agregar o eliminar variables mejoraba la calidad del modelo. Este proceso se repitió hasta alcanzar un modelo óptimo, asegurando la inclusión de las variables más relevantes (ver Anexo 1).

El modelo se construyó utilizando el 79% (937/1191) de los aislados de nuestra colección que no presentaban datos faltantes en ninguna de las variables a evaluar. Se validó mediante el test de Hosmer and Lemeshow *goodness of fit* (GOF) (p-valor = 0.109) y el análisis de curva ROC (*Receiver Operating Characteristic*), con un área bajo la curva de 0.7168. Además, se realizó una validación cruzada (0.2208100 - 0.2208037) (ver Anexo 1).

En el análisis del modelo de regresión logística que hemos llevado a cabo, se ha explorado con detalle la relación entre diversas variables y el agrupamiento genómico. A continuación, resumimos las interpretaciones más relevantes derivadas del análisis y presentadas en la tabla 3.3.

Comenzando con la variable Edad, hemos observado que por cada aumento unitario en la variable, las probabilidades de pertenencia al clúster genómico en cuestión disminuyen en

aproximadamente 0.034 ($p = 1.43 \text{ e-}12$) en el modelo. Esto sugiere que a medida que la edad aumenta, la probabilidad de pertenecer a un clúster genómico disminuye.

Al considerar la variable Lugar de nacimiento-Extranjero, encontramos que si un individuo es extranjero, las probabilidades de pertenencia al clúster genómico se reducen en un factor de alrededor de 1.416 ($p = 1.48 \text{ e-}11$), en comparación con aquellos que no tienen esta condición. En resumen, los individuos extranjeros muestran una probabilidad menor de pertenecer a un clúster genómico.

En relación con la localización de la infección, cuando esta ocurre en los pulmones, las probabilidades de pertenencia al clúster genómico aumentan significativamente en aproximadamente 0.791 ($p = 0.016$). Este hallazgo propone una asociación positiva entre la localización pulmonar y la pertenencia al clúster genómico.

En el análisis multivariado, a diferencia de las observaciones en el análisis bivariado, los factores PPL y Alcoholismo no se presentan como factores de riesgo para la pertenencia a clústeres. Del mismo modo, las variables "sexo" y "tinción del esputo", inicialmente asociadas de manera marginal a la pertenencia a clústeres ($p = 0.04$, $p = 0.01$ respectivamente), ya no mantienen dicha asociación. Esta variación puede atribuirse a la reducción del conjunto de datos en este análisis.

Según los resultados del modelo analizado, el origen español sigue siendo un factor de riesgo para la pertenencia a clúster, coincidiendo con el análisis bivariado. No se identificaron correlaciones significativas adicionales en las interacciones múltiples de las variables en este modelo. Es crucial destacar la aplicabilidad potencial de este modelo a conjuntos de datos similares en el futuro.

Tabla 3.3. Coeficientes y Significancia en el Análisis de Regresión Logística para la Pertenencia a Clústeres Genómicos

Variables	Estimación	Error estándar	Valor z	Pr(> z)
(Intercept)	1.360	0.400	3.396	0.000683 ***
Edad	-0.034	0.005	-7.081	1.43e-12 ***
Sexo-Mujer	0.501	0.399	1.257	0.208622
Lugar de nacimiento-Extranjero	-1.416	0.210	-6.750	1.48e-11 ***
Tinción de esputo-Positivo	-0.603	0.457	-1.318	0.187483
Localización-Pulmonar	0.791	0.328	2.408	0.016049 *
Alcoholismo-Si	0.366	0.268	1.367	0.171747
PPL-Si	0.166	0.562	0.295	0.767743
Sexo-Mujer: Localización-Pulmonar	-0.734	0.439	-1.673	0.094327 .
Tinción-Positivo: Localización-Pulmonar	0.559	0.494	1.131	0.258246
Extranjero-Si: Alcoholismo-Si	0.180	0.422	0.426	0.669800
Sexo-Mujer: PPL-Si	-2.093	1.448	-1.445	0.148371
Lugar de nacimiento-Extranjero: PPL-Si	2.391	1.261	1.896	0.057922 .
Alcoholismo-Si: PPL-Si	1.283	1.214	1.057	0.290526

Códigos de significancia: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Caracterización de la resistencia a fármacos

En total de los 1279 aislados evaluados, se identificaron 125 aislados (9.77%) resistentes al menos a un fármaco y 1154 aislados (90.23%) sensibles. La sensibilidad de la secuenciación de genomas para la predicción de resistencias fue superior al 95% para isoniazida y entre 50-69% para los otros tres fármacos de primera línea (Tabla 3.3).

Tabla 3.3. Desempeño de la WGS como herramienta de diagnóstico de la resistencia a fármacos de primera línea

Fármaco	Aislados resistentes con DST	Aislados resistentes con WGS	Sen	Esp	VPP	VPN
Isoniazida	68	68	96	100	96	100
Rifampicina	29	20	69	100	100	99
Etambutol	20	16	50	99	63	99
Pirazinamida	37	20	51	100	95	98

* Estos valores se extrajeron de 1279 aislados clínicos de TB. Se utilizó la DST (Drug susceptibility testing) test de susceptibilidad a fármacos como método de referencia.

* Abreviaturas: DST, prueba de susceptibilidad a los medicamentos; WGS (Whole genome sequencing): secuenciación del genoma completo; Sen=sensibilidad (%); Esp=Especificidad (%); VPP, valor predictivo positivo (%); VPN: valor predictivo negativo (%).

Resistencia a isoniazida (INH)

Se identificaron 68 aislados con resistencia fenotípica a isoniazida, de los cuales 65 (96%) mostraron los polimorfismos que explicarían su carácter resistente, sin embargo, 3 aislados (4%) con resistencia a este fármaco, no mostraron mutaciones en los genes analizados que pudieran explicar tal condición. Por otra parte, 3 aislados sensibles a este fármaco mostraron mutaciones asociadas a resistencia (Tabla 3.12).

En la Tabla 3.4 se puede apreciar la presencia de seis cambios en dos genes asociados con la resistencia a isoniazida, el más frecuente se presentó en 32 aislados en la posición intergénica 1673425 (c/t) del gen *fabG1*, seguido por en el codón *katG*315Ser/Thr presente en 27 aislados.

Debido a la frecuente aparición de la posición 1674481 Ser94Ala en el gen *inhA* en muestras con resistencia fenotípica a la isoniazida, y en ausencia de otras mutaciones del catálogo de la OMS que expliquen esta resistencia, hemos decidido incluir esta posición en nuestro análisis como un indicador de resistencia a la isoniazida. Aunque en el catálogo se clasifica como una mutación de significado incierto para la isoniazida (Nivel 3), se ha identificado como una mutación que confiere resistencia a la etionamida (Nivel 1). Además, ya se ha documentado previamente como una variante asociada a la resistencia a la isoniazida ^{179,180} y en nuestro laboratorio, hemos evidenciado un incremento en la Concentración Mínima Inhibitoria (CMI) de isoniazida asociada a esta mutación.

Tabla 3.4. Polimorfismos en genes asociados con resistencia a isoniazida encontrados en los aislados analizados (n=68)

Posición	Región	Codón (M.t.)	Cambio de aminoácido	Cambio de base	Fi	Nivel OMS
Isoniazida (INH)						
<i>fabG1</i>						
1673425	Intergénica	---	---	c/t	32	1
1673432	Intergénica	---	---	t/a	2	2
1673432	Intergénica	---	---	t/c	1	2
1674048	Codificante	203	Leu → Leu	g/a	4	1
<i>inhA</i>						
1674481	Codificante	94	Ser → Ala	tcg/gcg	13	---
<i>katG</i>						
2155168	Codificante	315	Ser → Thr	agc/aac	27	1
2155168	Codificante	315	Ser → Asn	agc/acc	1	1
Total					80	

(M.t) *Mycobacterium tuberculosis*. (Fi) Frecuencia absoluta.

(Nivel OMS) 1. Asociado a resistencia y 2. Asociado a resistencia de manera provisional ¹⁷⁷.

Resistencia a rifampicina (RIF)

En cuanto a rifampicina, de los 29 aislados con resistencia fenotípica, en 20 (69%) se observaron mutaciones que explicaran dicha resistencia, nueve aislados (31%) resistentes no mostraron cambios en los genes analizados. Por otra parte, ningún aislado sensible a este fármaco mostró mutaciones asociadas a resistencia (Tabla 3.12).

Por último, se apreciaron en total 14 cambios diferentes en el gen *rpoB*, el más frecuente fue en el codón *rpoB* 450Ser/Leu presente en 15 aislados (Tabla 3.5).

Tabla 3.5. Polimorfismos en genes asociados con resistencia a rifampicina encontrados en los aislados analizados

Posición	Región	Codón (<i>M.t.</i>)	Cambio de aminoácido	Cambio de base	<i>Fi</i>	Nivel OMS
Rifampicina (RIF)						
<i>rpoB</i>						
761089	Codificante	428	Ser → Thr	agc/acc	1	2
761090	Codificante	428	Ser → Arg	agc/agg	1	2
761095	Codificante	430	Leu → Pro	ctg/ccg	2	1
761097	Codificante	431	Ser → Gly	agc/ggc	1	2
761098	Codificante	431	Ser → Thr	agc/acc	1	2
761106	Codificante	434	Met → Val	atg/gtg	1	2
761110	Codificante	435	Asp → Val	gac/gtc	1	1
761127	Codificante	441	Ser → Ala	tcg/gcg	1	2
761139	Codificante	445	His → Asp	cac/gac	1	1
761139	Codificante	445	His → Tyr	cac/tac	1	1
761139	Codificante	445	His → Asn	cac/aac	2	1
761155	Codificante	450	Ser → Leu	tcg/ttg	15	1
761161	Codificante	452	Leu → Pro	ctg/ccg	1	1
761277	Codificante	491	Ile → Phe	atc/ttc	2	1
Total					31	

(*M.t.*) *Mycobacterium tuberculosis*. (*Fi*) Frecuencia absoluta.

(Nivel OMS) 1. Asociado a resistencia y 2. Asociado a resistencia de manera provisional ¹⁷⁷.

Resistencia a etambutol (EMB)

De los 20 aislados con resistencia fenotípica a etambutol en 10 (50%) se observaron polimorfismos conocidos que le proporcionan tal carácter, 10 (50%) aislados no mostraron cambios. Adicionalmente, seis aislados identificados como sensibles presentaron mutaciones asociadas con la resistencia a este fármaco (Tabla 3.12).

Finalmente, se identificaron seis cambios en dos genes (*embA* y *embB*) donde la mutación más frecuente fue en la posición codificante 4247429 (*embB* 306Met/Val) presente en siete aislados, seguido por el cambio del codón *emb* B306Met/Ile, en la misma posición, presente en cinco aislados (Tabla 3.6).

Tabla 3.6. Polimorfismos en genes asociados con resistencia a etambutol encontrados en los aislados analizados

Posición	Región	Codón (<i>M.t.</i>)	Cambio de aminoácido	Cambio de base	<i>Fi</i>	Nivel OMS
Etambutol (EMB)						
<i>embA</i>						
4243221	Intergénica	---	---	c/t	1	1
<i>embB</i>						
4247429	Codificante	306	Met → Val	atg/gtg	7	1
4247431	Codificante	306	Met → Ile	atg/att	5	1
4247431	Codificante	306	Met → Ile	atg/ata	1	1
4247469	Codificante	319	Tyr → Ser	tat/tct	1	1
4247730	Codificante	406	Gly → Asp	ggc/gac	1	1
Total					16	

(*M.t.*) *Mycobacterium tuberculosis*. (*Fi*) Frecuencia absoluta.

(Nivel OMS) 1. Asociado a resistencia y 2. Asociado a resistencia de manera provisional ¹⁷⁷.

Resistencia a pirazinamida (PZA)

En los 37 aislados que mostraron resistencia a este fármaco, 19 (51.35%) contenían un polimorfismo que explicaba dicha resistencia y 18 (48.65%) no mostraron ninguna modificación (Tabla 3.12). Finalmente, un solo aislado considerado como sensible mostró un cambio asociado con esta resistencia. Se identificaron 27 polimorfismos, todos ellos en el gen *pncA*, el más frecuente se presentó en 17 aislados, en el codón *pncA* 57His/Asp (Tabla 3.7).

Tabla 3.7. Polimorfismos en genes asociados con resistencia a pirazinamida encontrados en los aislados analizados

Posición	Región	Codón (<i>M.t.</i>)	Cambio de aminoácido	Cambio de base	<i>Fi</i>	Nivel OMS
Pirazinamida (PZA)						
<i>pncA</i>						
2288805	Codificante	146	Ala → Val	gcg/gtg	1	1
2288930	Codificante	104	Ser → Arg	agc/agg	1	2
2288962	Codificante	94	Phe → Leu	ttc/ctc	1	1
2289058	Codificante	62	Pro → Thr	ccg/acg	1	2
2289073	Codificante	57	His → Asp	cac/gac	17	1
2289081	Codificante	54	Pro → Leu	ccg/ctg	1	1
2289090	Codificante	51	His → Arg	cac/cgc	1	1
2289096	Codificante	49	Asp → Gly	gac/ggc	1	1
2289220	Codificante	8	Asp → Asn	gac/aac	1	1
2289222	Codificante	7	Val → Ala	gtc/gcc	1	1
2289231	Codificante	4	Leu → Ser	ttg/tcg	1	1
Total					27	

(*M.t.*) *Mycobacterium tuberculosis*. (*Fi*) Frecuencia absoluta.

(Nivel OMS) 1. Asociado a resistencia y 2. Asociado a resistencia de manera provisional ¹⁷⁷.

Caracterización de Multidrogoresistencia (resistencia simultánea, rifampicina e isoniazida)

Se identificaron a 21 aislados fenotípicamente multidrogoresistentes de los cuales 19 (86%) presentaron polimorfismos en los genes asociados con la resistencia a rifampicina e isoniazida. Dos aislados (14%) no presentaron cambios en los genes relacionados con la resistencia a los dos fármacos implicados en esta agravada forma de TB resistente. Finalmente, se identificó un aislado con mutaciones asociadas a resistencia para ambos fármacos, pero no se disponía de perfil fenotípico.

Caracterización de la resistencia a fármacos de segunda línea

No fue posible realizar la identificación de la resistencia fenotípica a fármacos de segunda línea en los aislados incluidos en el estudio; sin embargo, el análisis de genes asociados con la resistencia a estos fármacos mostró lo siguiente: 27 aislados presentaron mutaciones que les conferían resistencia a fluoroquinolonas y 10 a kanamicina, amikacina o capreomicina, (ver Tablas 3.8, 3.9).

Los polimorfismos presentes en estos aislados corresponden a 10 cambios dentro de los genes *gyrB* y *gyrA*, así como cinco posiciones en los genes *eis* y *rrs* (Tabla 3.8), en este sentido es importante mencionar que algunas de estas mutaciones pueden generar resistencia a más de un fármaco.

Tabla 3.8. Polimorfismos en genes asociados con resistencia a fluoroquinolonas encontrados en los aislados analizados

Posición	Región	Codón (<i>M.t.</i>)	Cambio de aminoácido	Cambio de base	<i>Fi</i>	Nivel OMS
Fluoroquinolonas (FQ)						
<i>gyrB</i>						
6734	Codificante	499	Asn → Asp	aac/gac	1	2
6750	Codificante	504	Ala → Val	gcg/gtg	1	2
<i>gyrA</i>						
7564	Codificante	88	Gly → Ala	ggc/gcc	1	2
7570	Codificante	90	Ala → Val	gcg/gtg	7	1
7572	Codificante	91	Ser → Pro	tcg/ccg	1	1
7581	Codificante	94	Asp → Asn	gac/aac	5	1
7581	Codificante	94	Asp → His	gac/cac	1	1
7581	Codificante	94	Asp → Tyr	gac/tac	1	1
7582	Codificante	94	Asp → Gly	gac/ggc	2	1
7582	Codificante	94	Asp → Ala	gac/gcc	12	1
Total					32	

(*M.t.*) *Mycobacterium tuberculosis*. (*Fi*) Frecuencia absoluta.

(Nivel OMS) 1. Asociado a resistencia y 2. Asociado a resistencia de manera provisional ¹⁷⁷.

Tabla 3.9. Polimorfismos en genes asociados con resistencia a aminoglucósidos (amikacina, kanamicina y capreomicina) encontrados en los aislados analizados

Posición	Región	Codón (<i>M.t.</i>)	Cambio de aminoácido	Cambio de base	<i>Fi</i>	Nivel OMS
Amikacina (AM), Kanamicina (KM), Capreomicina (CM)						
<i>rrs</i>						
1472359	Ribosomal	---	---	a/c	1	1
1473246	Ribosomal	---	---	a/g	5	1
1473329	Ribosomal	---	---	g/t	1	1
<i>eis</i>						
2715344	Intergénica	---	---	c/t	2	1
2715346	Intergénica	---	---	c/t	2	1
Total					11	

(*M.t.*) *Mycobacterium tuberculosis*. (*Fi*) Frecuencia absoluta.

(Nivel OMS) 1. Asociado a resistencia y 2. Asociado a resistencia de manera provisional ¹⁷⁷.

Discusión

El trabajo actual es la continuación de la implementación de la epidemiología genómica como una herramienta para la vigilancia de la tuberculosis , estudio de prevalencia de resistencia a fármacos y análisis de la estructura poblacional de la bacteria, en la Comunidad Valenciana.

Los datos demográficos de la población de estudio no difieren en gran medida con las estadísticas globales de España ¹⁸¹, con una distribución de pacientes de sexo masculino (64.56%) superior al femenino (35.43%), una concentración de casos en individuos mayores o iguales a 65 años y la localización, preferentemente pulmonar, de la infección. Por otro lado, las proporciones de variables y factores de riesgo en nuestra muestra de 1191 aislados secuenciados, no difiere significativamente de la población total (2206 casos) en la Comunidad Valenciana, lo que confirma la representatividad de nuestra muestra para el conjunto de casos en la región.

Estructura poblacional y agrupamiento genómico

Mediante la reconstrucción filogenética de nuestros aislados, fueron identificadas muestras de los linajes 1-6 y algunos aislados de origen animal. Con una representación mayor del linaje 4, el más distribuido en Europa ²⁸. Así también no se observaron muestras del linaje 7 u 8 posiblemente por su restringida ubicación geográfica ^{18,182}.

Nuestra investigación se enfocó en el período de 2014 a 2018 en la Comunidad Valenciana. Al comparar nuestros hallazgos con un estudio previo, que abarcó los años 2014 a 2016 en la misma población ¹⁸³, observamos resultados similares. Analizamos 1279 aislados de TB en contraste con los 785 del estudio anterior. Encontramos que en nuestro estudio, el 74% de las muestras resultaron positivas en el cultivo, mientras que en el estudio anterior este porcentaje fue del 77%. También identificamos similitudes en la agrupación genómica, con un 49% en nuestro estudio (utilizando 10 SNPs), en comparación con el 47% (utilizando 12 SNPs) informado previamente. Además, las proporciones de variables y factores de riesgo se mantuvieron semejantes tanto en las poblaciones totales como en los análisis estadísticos del agrupamiento genómico contra los casos individuales.

En este estudio, decidimos utilizar un umbral de 10 SNPs como punto de referencia para analizar los eventos de transmisión, tanto recientes como antiguos, basándonos en nuestra propia distribución de datos y en investigaciones previas realizadas en nuestra población ¹⁸³. Estos umbrales de distancia genética nos permitieron distinguir entre transmisión reciente (0 SNPs), ocurrida en la última década (5 SNPs) y más antigua (15 SNPs) ⁸⁵. La elección de estos umbrales ha sido respaldada por su utilidad en investigaciones previas en el ámbito de la salud pública ^{184,185}, donde han demostrado ser valiosos para evaluar la carga de transmisión, medir el impacto de los programas de control y detectar sitios de transmisión y factores de riesgo que a menudo pasan desapercibidos al rastreo de contactos.

En este contexto, se destaca una diferencia significativa entre el 49% de los aislados que formaban clústeres genómicos (630 aislados), en comparación con el 13% (168 aislados) identificados mediante el estudio de contactos. Este hallazgo sugiere que el enfoque de agrupamiento genómico ofrece una perspectiva más amplia de la transmisión de la tuberculosis en comparación con el estudio de contactos, resaltando la utilidad de la secuenciación genómica en la identificación de conexiones epidemiológicas.

A pesar de la eficacia demostrada del rastreo de contactos en la detección de nuevos casos de tuberculosis ¹⁸⁶, este método se ve afectado por diversas variables, como los retrasos en atención, diagnóstico y tratamiento, así como la baja sensibilidad de las pruebas diagnósticas iniciales ¹⁸⁷. Además, el comportamiento de las personas al no informar sobre sus contactos cercanos, junto con la presencia de casos asintomáticos y la posibilidad de contactos más allá del círculo cercano del paciente, dificultan la detección de casos relacionados con un individuo.

En contraste, el análisis genómico en el contexto de la tuberculosis ofrece ventajas notables al identificar con precisión la transmisión mediante distancias genéticas entre muestras, permitiendo la diferenciación entre transmisión reciente y antigua. Proporciona una resolución superior en la tipificación, lo cual es crucial para una enfermedad tan diversa como la tuberculosis en términos de virulencia y resistencia. Facilita la investigación de brotes al identificar grupos de transmisión, permitiendo comprender la propagación y evolución del patógeno, y encuentra casos que podrían pasar desapercibidos para el rastreo de contactos tradicional, especialmente aquellos sin vínculos evidentes.

A pesar de estas diferencias en la detección de casos entre el rastreo de contactos y la secuenciación genómica, es fundamental destacar que el análisis genómico no pretende reemplazar al estudio de contactos, sino complementarlo. Su valor radica en su capacidad para evaluar la eficacia y el alcance de las estrategias de rastreo de contactos existentes, lo que conduce a la identificación de áreas de mejora. Al combinar estos dos enfoques, se logra una comprensión más completa de la transmisión de la tuberculosis, proporcionando una sólida base para la toma de decisiones en el ámbito de la salud pública.

Por otro lado, es crucial tener en cuenta que las discrepancias en la detección de casos entre el rastreo de contactos y la secuenciación genómica varían según la carga de tuberculosis en una determinada región. En entornos con una baja carga de casos, existe un mayor grado de coincidencia entre casos genéticamente relacionados y vínculos epidemiológicos (38-57%)^{184,185,188}, lo que sugiere que el rastreo de contactos es más efectivo en estas condiciones. En contraste, en áreas con alta carga de casos y transmisión comunitaria descontrolada¹⁸⁹, la concordancia entre ambos enfoques es significativamente menor (8-19%)¹⁹⁰⁻¹⁹², lo que plantea desafíos adicionales para el control de la enfermedad. En el contexto de la región de Valencia, hemos observado una coincidencia similar a la que se encuentra en áreas con alta carga de tuberculosis, aproximadamente un 13.2%. Este hallazgo sugiere que una parte significativa de la transmisión de la enfermedad no se identifica a través del seguimiento de contactos, a pesar de su implementación.

En estos contextos de alta carga de TB, el seguimiento de contactos cercanos en hogares tiene un impacto limitado en la incidencia comunitaria de la tuberculosis^{193,194}, ya que gran parte de la transmisión resulta de encuentros casuales entre individuos desconocidos en la comunidad^{33,190}. Además, la transmisión está más vinculada a factores sociales, como las interacciones y congregaciones^{75,195}. Nuestros resultados respaldan la noción de que la transmisión de tuberculosis en la Comunidad Valenciana es la causa subyacente de la discrepancia entre el número de casos detectados mediante agrupación genómica y el estudio de contactos. Por lo tanto, es de suma importancia identificar con precisión los lugares donde se produce la transmisión de la tuberculosis. Esto implica investigar no sólo los sitios obvios, como el círculo cercano del paciente, sino también sitios menos convencionales, como el transporte público o establecimientos de hostelería.

Análisis de variables y factores de riesgo relacionados con el agrupamiento genómico

Al analizar la relación entre diferentes variables y la transmisión de la enfermedad (ver Tabla 3.2), observamos asociaciones significativas en dos grupos de edad: menores de 15 años y mayores de 65 años. Los menores de 15 años presentan una asociación con la pertenencia a agrupaciones genómicas, mientras que los mayores de 65 años están más relacionados con casos individuales. Nuestro modelo de regresión logística respalda esta tendencia. Estos hallazgos sugieren que la transmisión es más común entre individuos jóvenes, sobre todo en niños, donde la infección se produce por transmisión de su círculo familiar. En contraste, en adultos mayores, la mayoría de los casos podrían ser reactivaciones de infecciones previas. Las asociaciones observadas en ambos grupos de edad ≤ 15 y ≥ 65 años pueden explicarse por la situación inmunológica particular de estos individuos, que los hace más susceptibles al desarrollo de infecciones microbianas ^{196–198}.

La observación de una asociación entre el sexo masculino y la transmisión ($p=0.04$) (Tabla 3.2) podría tener múltiples explicaciones, que van desde posibles sesgos epidemiológicos, características biológicas ¹⁹⁹, barreras socioeconómicas y culturales, hasta diferencias en el acceso a la atención médica y la exposición a entornos de transmisión. A pesar de que se ha comprobado que el sexo influye en las manifestaciones clínicas de la enfermedad ²⁰⁰, en nuestro modelo de regresión logística no pudimos detectar una interacción específica con el sexo que explique el aumento en las proporciones observado en el modelo bivariado. Identificar la razón detrás de esta asociación, que se ha observado en varios estudios ^{199,201} y en nuestra población, es esencial para mejorar la comprensión de la enfermedad.

La asociación significativa entre “Lugar de nacimiento - Español” (nacidos en España) ($p=0.001$) (Tabla 3.2) podría atribuirse a una predisposición de los individuos nacidos en España a la infección por TB o a una posible adaptación de *M. tuberculosis* a este grupo de población. Nuestros estudios previos han detectado una alta tasa de transmisión en la población local en comparación con poblaciones de baja incidencia, donde la transmisión se observa principalmente en población extranjera ^{183,202}. Este hallazgo se abordará en detalle en el capítulo cuatro de este estudio.

La asociación significativa entre la localización pulmonar de la infección y la transmisión ($p=0.001$), así como la asociación entre una muestra de tinción de esputo positiva y la transmisión ($p=0.01$), se justifica por la naturaleza de la tuberculosis, que se transmite por vía aérea y tiene como diana el espacio alveolar del pulmón, específicamente los macrófagos alveolares. Debido a esto, la infección pulmonar por TB es más frecuente que la extrapulmonar, la historia natural de la enfermedad implica la expulsión de bacilos a través de la tos como un mecanismo de defensa del individuo. En consecuencia, es de esperar que la mayoría de las personas infectadas desarrollen la enfermedad pulmonar y presenten carga bacteriana al exhalar o toser.

En cuanto a factores de riesgo asociados con transmisión: alcoholismo ($p=0.001$) y personas privadas de libertad con enfermedad tuberculosa ($p=0.04$). Es importante mencionar que cerca del 80% de los individuos privados de su libertad son hombres ²⁰³ y que el alcoholismo es más frecuente en hombres que en mujeres ²⁰⁴. Aunque nuestro análisis de regresión logística no encontró una interacción entre estas dos variables, no podemos descartar la posibilidad de que en otras poblaciones esta interacción genere sesgos en las asociaciones observadas.

Caracterización de la resistencia a fármacos

En cuanto a la predicción de resistencias a fármacos de primera línea, la sensibilidad, especificidad, VPP (Valor predictivo positivo) y VPN (Valor predictivo negativo) obtenidos con este catálogo, incluyendo la posición 1674481 en el gen *inhA* para la predicción de resistencia a isoniazida, son equiparables a los reportados por estudios similares ^{205,130,206,207,208} (Tabla 3.11).

El análisis de la sensibilidad para el diagnóstico de la resistencia a isoniazida mostró un valor del 96%, lo cual es hasta un 12% superior al compararse con estudios similares ^{205,130,206,207,208}. La mayoría de los aislados presentaron una mutación conocida que explica la resistencia fenotípica observada (65/68). Se detectó resistencia genotípica en tres aislados catalogados fenotípicamente sensibles al fármaco y en otros tres no se identificaron mutaciones conocidas asociadas a la resistencia fenotípica reportada.

Un aspecto significativo a destacar es la consideración de la posición codificante 1674481 Ser94Ala en el gen *inhA*, para predecir la resistencia a la isoniazida. Aunque esta posición se clasifica en la categoría tres (significado incierto) según los criterios de la OMS para predecir la resistencia a la isoniazida, es esencial señalar que ha sido detectada en varios estudios y se ha asociado con la resistencia a la isoniazida ^{151,179,180,209}; asimismo, los análisis de la Concentración Mínima Inhibitoria (CMI) llevados a cabo por nuestro grupo aportan evidencia adicional que respalda dicha asociación. La detección de esta mutación y el aumento en la sensibilidad de la predicción de resistencia a la isoniazida, del 76% al 96%, posiblemente se debe a que las mutaciones asociadas a la resistencia a INH ya están ampliamente caracterizadas a nivel global en el catálogo de la OMS. Aunque aún falte evidencia para clasificarla en los niveles 1 o 2 de dicho catálogo. Este descubrimiento plantea la posibilidad de que, en el futuro, esta mutación pueda ser elevada a un nivel más alto en el catálogo de la OMS, permitiendo una mayor precisión en la predicción de resistencia a la isoniazida.

En el contexto del análisis de la resistencia a la rifampicina, es importante destacar que todos los aislados previamente clasificados como sensibles mediante el diagnóstico fenotípico también se confirmaron como sensibles mediante secuenciación masiva de genomas (WGS). Sin embargo, surge una observación interesante: entre los aislados con resistencia fenotípica a la rifampicina, nueve de ellos no presentaron mutaciones conocidas en los genes asociados a la resistencia, lo que plantea incertidumbres respecto a su estado de resistencia. Una posible explicación para estos nueve casos de falsos negativos podría estar relacionada con la falta de consideración de posiciones específicas para la resistencia a la rifampicina, que no se encuentran contempladas en el catálogo de la OMS. A pesar de estos casos individuales, en términos generales, el catálogo ha demostrado ser efectivo en la predicción de la resistencia a este medicamento.

La caracterización de la resistencia al etambutol plantea importantes consideraciones. Por un lado, la detección de falsos positivos sugiere posibles errores en el diagnóstico fenotípico. Por otro lado, los falsos negativos generan incertidumbre sobre el perfil genotípico de resistencia al etambutol en nuestra población. Estos falsos negativos podrían estar relacionados con mutaciones no incluidas en el catálogo, como se ha mencionado anteriormente. Además, investigaciones similares indican una disminución en la sensibilidad para este fármaco (Tabla

3.11), lo que sugiere que la investigación de mutaciones y mecanismos de resistencia al etambutol podría no estar tan avanzada como la de isoniazida o rifampicina.

En cuanto a la caracterización de la resistencia a pirazinamida, se identificó un aislado como falso positivo y en 18 aislados no se pudieron detectar los polimorfismos que explican la resistencia fenotípica. Nuestros valores de sensibilidad son inferiores en comparación con estudios similares (Tabla 3.11) ^{205,130,206,207,208}. Esto posiblemente se deba al uso restrictivo de las posiciones en nuestro catálogo. Sería prudente considerar la expansión de estas posiciones específicamente para este fármaco, dado que se ha propuesto previamente que cualquier mutación no sinónima dentro del gen *pncA* puede generar resistencia ²¹⁰.

En relación con las resistencias de segunda línea, se detectaron 27 aislados resistentes a fluoroquinolonas (Tabla 3.8) y 10 a aminoglucósidos (Tabla 3.9) mediante el uso de la WGS. Esto resalta la utilidad de la WGS en la predicción de resistencias, ya que estos aislados no contaban con un diagnóstico fenotípico. Además, estos hallazgos subrayan la baja incidencia de aislados resistentes a fármacos de segunda línea en la Comunidad Valenciana.

En el contexto de nuestro estudio sobre resistencias a fármacos, hemos identificado una tendencia general hacia un aumento en la especificidad y una disminución en la sensibilidad en la predicción de resistencia. Esta tendencia parece estar relacionada con la selección restrictiva de mutaciones que hemos utilizado, basada en los niveles uno y dos del catálogo de la OMS. Además, aunque poco probable, no podemos descartar la posibilidad de que la variabilidad genética presente en nuestra población haya dado lugar a mutaciones no documentadas o con evidencia limitada que no se encuentran en el catálogo de referencia.

En conclusión, a pesar de que el catálogo restrictivo empleado en nuestro estudio ha demostrado ser efectivo para predecir la resistencia o susceptibilidad en las muestras analizadas, aún dista de ser perfecto. Estos resultados resaltan la necesidad de una evaluación continua y ajustes en la aplicación de la secuenciación masiva de genomas (WGS) en la predicción de resistencia a fármacos. La constante acumulación de evidencia, a través de estudios similares al nuestro en diversas poblaciones, es esencial para el desarrollo y la mejora del catálogo de mutaciones de la OMS. A medida que se acumule nueva evidencia, se espera una expansión en el número de mutaciones asociadas a la resistencia a fármacos. Este

proceso de mejora continua permitirá un mejor manejo de la resistencia a los medicamentos y una toma de decisiones clínica más precisa en el futuro.

4. Genotipos endémicos en la Comunidad Valenciana

Introducción

La patogénesis de la enfermedad causada por *Mycobacterium tuberculosis* involucra una compleja interacción entre las variaciones genéticas del huésped y la bacteria. A lo largo de siglos de coexistencia, la humanidad ha desarrollado mecanismos para prevenir la infección y el desarrollo de la enfermedad, indicando una adaptación bacteria-huésped ^{50,51,211,212}.

Mycobacterium tuberculosis comparte procesos metabólicos similares con micobacterias no patógenas, pero demuestra una habilidad excepcional para persistir y propagarse en el huésped humano ²¹³, lo cual lo convierte en un patógeno especializado de considerable peligrosidad.

A pesar de su diversidad genética, es probable que haya una selección de su carga genética a través de la adaptación entre las poblaciones bacterianas y humanas. Esta asociación entre *M. tuberculosis* y los seres humanos tiene una historia milenaria, documentada en antiguas civilizaciones como Egipto ²¹⁴ y culturas precolombinas ²¹⁵. Los modernos métodos de datación han permitido inferir que dicha asociación probablemente ya existía al menos desde hace unos 5000 años.

Los avances tecnológicos recientes permiten un estudio más detallado y local de este fenómeno. Distintos estudios han demostrado que *M. tuberculosis* se encuentra estructurado geográficamente, con ciertos linajes asociados a regiones geográficas específicas ^{216,28} o a rutas migratorias ²¹⁷. Por ejemplo, se ha observado que el L1 infecta principalmente a individuos del subcontinente indio o del sudeste asiático ²¹⁸ y ciertos linajes se transmiten de manera más eficiente dentro de la población huésped con la que han coexistido históricamente ^{219 43}.

La diversidad genética en los seres humanos y la naturaleza crónica de la TB dificultan el estudio de las interacciones entre el huésped humano y el bacilo, debido a la dificultad para identificar el estado clínico de la enfermedad. Sin embargo, modelos murinos han proporcionado una comprensión adecuada de esta relación, revelando variabilidad en la virulencia entre cepas del mismo linaje ²²⁰ y mutaciones específicas que aumentan la virulencia ²²¹.

Hasta el momento, pocos estudios han abordado la identificación de genotipos de *M. tuberculosis* asociados a poblaciones específicas, así como las consecuencias de esta asociación; esto se debe, principalmente, a la disponibilidad de herramientas de tipificación de baja resolución y la dificultad para rastrear la transmisión en humanos.

El objetivo principal de este estudio consiste en emplear la secuenciación del genoma completo con el fin de investigar la existencia de genotipos particulares de *M. tuberculosis* asociados a poblaciones humanas específicas. En este contexto, se propone una metodología para la identificación de genotipos circulantes, utilizando una colección de aislados clínicos de la Comunidad Valenciana, y se definen criterios específicos para determinar la presencia de genotipos endémicos.

Los resultados obtenidos en este estudio respaldan las hipótesis previas acerca de la adaptación local entre el huésped y la bacteria. No obstante, cabe destacar que estos hallazgos se basan en marcadores de alta resolución, lo que los diferencia de investigaciones previas llevadas a cabo en San Francisco, Canadá, Suiza y Tailandia ^{43,222,223,224}.

Métodos

En este capítulo, hemos desarrollado una metodología para identificar genotipos endémicos de *Mycobacterium tuberculosis* en la Comunidad Valenciana. Utilizamos como punto de partida nuestra colección previamente analizada en el capítulo tres, que consta de 1277 aislados clínicos. Nuestra metodología consta de varios pasos fundamentales.

Primero identificamos los genotipos presentes en la Comunidad Valenciana y analizamos su distribución geográfica. Seleccionamos cuidadosamente los genotipos relevantes para nuestro estudio y exploramos su relación con la transmisión. Posteriormente, llevamos a cabo un estudio filogeográfico que implicó la representación geográfica de estos genotipos en una filogenia global representativa. El objetivo principal de esta fase fue identificar los genotipos endémicos. Este análisis nos permitió reconstruir la ancestralidad de los clados de interés y realizar un análisis estadístico de sus variables. También nos permitió reexaminar su relación con la transmisión.

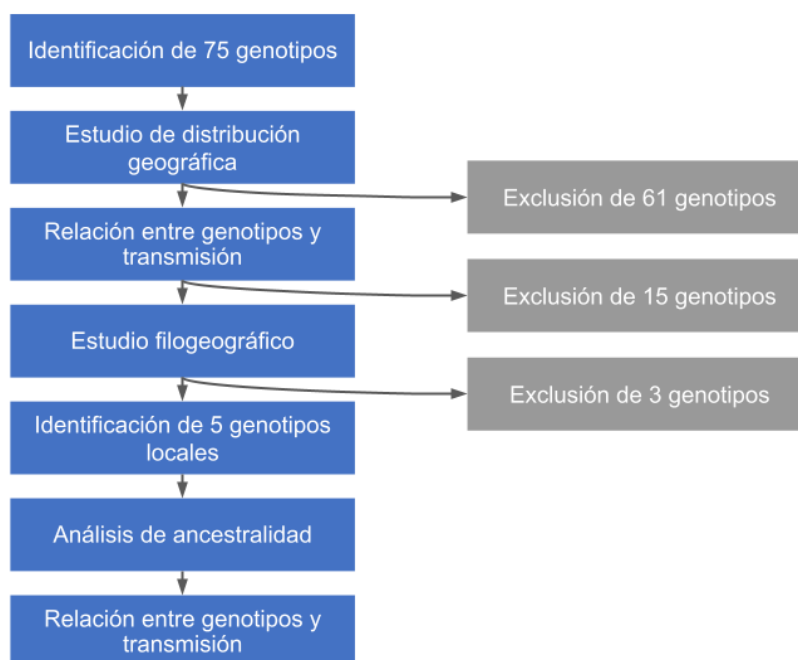


Figura 4.1. Metodología para la identificación de genotipos endémicos de la Comunidad Valenciana

Para llevar a cabo análisis estadísticos y crear gráficos claros y representativos, hemos utilizado la colección de paquetes "Tydiverse" ¹⁷⁸ en el lenguaje R, además de librerías especializadas en análisis filogenéticos y visualización de datos.

Este enfoque metodológico nos ha permitido obtener una comprensión más profunda de la diversidad genética de *M. tuberculosis* en la Comunidad Valenciana. Además, hemos identificado genotipos endémicos que podrían desempeñar un papel crucial en la transmisión de la enfermedad en esta región.

Identificación de genotipos en la Comunidad Valenciana

Para la identificación de genotipos, hemos empleado el método computacional FastBAPS ²²⁵ (Fast Bayesian Analysis of Population Structure, por sus siglas en inglés). Este método estadístico ha sido diseñado específicamente para analizar la estructura genética de poblaciones, identificando grupos de individuos con similitudes genéticas en grandes conjuntos de datos genómicos.

FastBAPS utiliza un enfoque bayesiano y una técnica de agrupamiento jerárquico para asignar individuos a diferentes genotipos. Es especialmente eficiente en el análisis de extensos conjuntos de datos genómicos, como aquellos obtenidos mediante secuenciación masiva. Es por esta razón que ha sido ampliamente adoptado en estudios genéticos de poblaciones bacterianas, debido a su velocidad y eficacia en comparación con otras herramientas de análisis de estructura poblacional ^{226,227}.

Para llevar a cabo la clasificación, suministramos al programa un alineamiento multifasta junto con su correspondiente árbol filogenético. Hemos utilizado el alineamiento de secuencias en formato multifasta (sparse.data) como entrada para la función "optimise_prior", que asigna cada aislado a su respectivo grupo genético. Los parámetros que aplicamos fueron los siguientes:

- `optimise_prior(sparse.data, type = "optimise.baps")`

Luego, proporcionamos al programa la reconstrucción filogenética en formato Newick, la cual fue generada en el capítulo tres de nuestro estudio. Esta reconstrucción se utilizó como entrada para la función "multi_level_best_baps_partition". Dicha función tiene la tarea de redefinir los grupos con base en la información contenida en la filogenia. Los parámetros que configuramos para esta etapa fueron los siguientes:

- `multi_level_best_baps_partition(sparse.data, Newick)`

Nuestro análisis nos permitió identificar genotipos con una mayor resolución que la de linaje y sublinaje. Es relevante destacar que FastBAPS tiene la capacidad de generar particiones en varios niveles de resolución, incluso asignando un genotipo por muestra si es necesario. En nuestra aproximación, generamos tres niveles de partición y seleccionamos el segundo nivel para nuestros análisis.

La elección del segundo nivel se basó en que el primer nivel ofrecía una resolución baja, lo que resultaba en un número limitado de genotipos con un gran número de muestras, similar a la distribución de linajes y sublinajes. Por otro lado, optamos por descartar el tercer nivel de partición debido a que proporcionaba una resolución demasiado alta, lo que resultaba en un gran número de genotipos con un número reducido de aislados.

Para la identificación de genotipos endémicos en la Comunidad Valenciana, llevamos a cabo un análisis estadístico descriptivo. Inicialmente, nos centramos en los genotipos del linaje cuatro, que constituían el 92% de nuestra muestra (1175 de 1277 aislados). A continuación, seleccionamos aquellos genotipos que presentaban más de 20 aislados, que nos permitieran realizar inferencias estadísticas.

Distribución geográfica de genotipos de *M. tuberculosis* en la Comunidad Valenciana

A continuación y en concordancia con el análisis previo, llevamos a cabo un análisis de distribución geográfica, con el propósito de examinar la disposición espacial de todos los genotipos. Esta investigación nos brindó una comprensión más profunda de cómo se distribuyen los genotipos en la Comunidad Valenciana, además de identificar posibles relaciones entre ubicaciones y genotipos.

Para abordar este punto, empleamos la herramienta Microreact ²²⁸ para cartografiar la distribución geográfica de los aislados agrupados en genotipos. Utilizamos las coordenadas de latitud y longitud correspondientes a cada muestra, las cuales se obtuvieron a partir de los códigos postales (consultar Figura 4.b). Además, cargamos en la interfaz de la plataforma tanto el árbol filogenético, como la información temporal de los aislados.

Con el propósito de identificar y descartar genotipos que pudieran haber surgido por eventos de transmisión locales y, por lo tanto, no estar dispersos en toda la Comunidad Valenciana, se examinó la asociación entre la distancia genética y la distancia geográfica usando el Test de Mantel. Esta prueba estadística, ampliamente empleada en biología y genómica, ha sido utilizada en otros estudios para analizar la correlación espacial entre poblaciones y la distancia genética, así como para evaluar la correlación entre variables como la urbanización y la resistencia a fármacos ^{229,230}.

El Test de Mantel se basa en el cálculo de la correlación de rangos de Spearman entre dos matrices y su significancia se determina mediante la prueba chi-cuadrado. La interpretación se realiza a partir del coeficiente de correlación de Mantel (r_M) y el valor p correspondiente. Donde un r_M cercano a 1 indica una fuerte correlación positiva entre la distancia genética y la distancia geográfica, mientras que un r_M cercano a -1 sugiere una correlación negativa y un valor r_M cercano a 0 indica falta de correlación. Se consideró que un valor p pequeño (generalmente <0.05) indicaba una asociación significativa.

Se efectuó el Test de Mantel utilizando las matrices de distancia genética y geográfica, empleando la librería "vegan" versión 2.6-4 en el lenguaje R. La matriz de distancia genética, previamente empleada en la construcción de clústeres de transmisión en el capítulo tres, se reestructuró para este análisis. Adicionalmente, se diseñó una matriz de distancia geográfica en kilómetros exclusivamente para este estudio.

La prueba se llevó a cabo tanto a nivel global con todos los aislados, como de forma individual para cada genotipo, utilizando los siguientes parámetros:

- `mantel(matriz_distancias_geneticas, matriz_distancias_geograficas, permutations=999, method = "spearman", na.rm = TRUE)`

Relación entre genotipos y transmisión

En un estudio complementario de nuestra investigación, nos enfocamos en analizar la transmisión de los genotipos potencialmente endémicos. Para llevar a cabo esta evaluación, examinamos la relación entre la transmisión y la distribución de casos locales (nacidos en España) y extranjeros.

Utilizamos un diagrama de dispersión para analizar esta relación en cada genotipo. En el eje Y, mostramos la proporción de casos no nacidos en España, mientras que en el eje X, representamos la proporción de casos nacidos en España que pertenecen a un clúster genómico de transmisión.

Estudio filogeográfico e identificación de genotipos endémicos

Con el objetivo de evitar la sobrerrepresentación de genotipos, debido a muestras similares resultantes de eventos de transmisión, en el análisis filogeográfico, se procedió a eliminar la redundancia seleccionando una muestra representativa por cluster de transmisión. Esta selección se llevó a cabo mediante el programa Seqtk²³¹ utilizando el comando "subseq" y un alineamiento de multiseuencias previamente generado (in.fq), con el fin de extraer las secuencias de interés (name.lst) y, posteriormente, generar un nuevo alineamiento (out.fq) aplicando los siguientes parámetros:

- seqtk subseq in.fq name.lst > out.fq

A continuación, se procedió a reconstruir una filogenia de máxima verosimilitud siguiendo los parámetros descritos en el capítulo tres. Esta reconstrucción permitió identificar nuevamente los genotipos previamente seleccionados en una filogenia reducida, excluyendo las muestras en clústeres de transmisión.

Una vez implementada la estrategia mencionada, procedimos a realizar un mapeo de los 14 genotipos locales en una filogenia global que representaba el linaje cuatro. Esta filogenia global se construyó con un total de 9640 genomas procedentes de 85 países distintos, incluyendo las muestras de nuestra colección.

La reconstrucción de esta filogenia global se llevó a cabo mediante el método de neighbor-joining utilizando el programa MEGA ²³². Esto nos permitió determinar si estos genotipos se agrupaban en clados específicos, lo que podría indicar una posible restricción geográfica en la Comunidad Valenciana, o si, por el contrario, se distribuían de manera aleatoria en diferentes regiones del mundo. Es importante resaltar que en este análisis se excluyeron las muestras con baja calidad de secuenciación.

Una vez completada la reconstrucción de la filogenia global, procedimos a identificar los clados en los que se ubicaban los genotipos previamente seleccionados, seguido de un análisis estadístico para describir la proporción de aislados pertenecientes a estos genotipos en relación con el total de aislados en cada clado. Prestamos especial atención a los clados en los que esta proporción alcanzaba o superaba el 20% (consultar Tabla 4.1). Ofrecimos información detallada sobre las variables de linaje/sublinaje, la cantidad y el porcentaje de aislados nacidos en España y en otros lugares, así como el número total de aislados. Vale la pena señalar que identificamos aquellos genotipos que mostraban una proporción igual o superior al 65% de casos nacidos en España en la filogenia original (ver Figura 4.2 y Tabla 4.1).

Finalmente, se identificaron como genotipos endémicos aquellos pertenecientes al linaje cuatro que cumplían con los siguientes criterios: un número de aislados igual o superior a 20, una proporción igual o superior al 65% de casos nacidos en España en la filogenia original y, al mapearse en la filogenia global, presentaban una proporción de aislados de nuestra colección igual o superior al 20%.

Reconstrucción ancestral

Para refinar nuestra selección, utilizamos el programa RASP4 ²³³ una herramienta de reconstrucción de estados ancestrales que nos permitió inferir la ancestralidad de los genotipos. Nuestro racional fue que el ancestro común de los genotipos endémicos debía tener su origen en la Comunidad Valenciana.

RASP4 es una herramienta especializada en la reconstrucción simultánea de estados ancestrales de múltiples genes/genomas y caracteres a partir de datos genómicos. Este

software se fundamenta en modelos estadísticos avanzados y emplea métodos de inferencia como la máxima verosimilitud, el muestreo de Markov Monte Carlo y el análisis bayesiano para estimar los estados ancestrales de los caracteres. Además, presenta una interfaz gráfica amigable que facilita la carga de datos y la personalización de los parámetros de análisis ²³⁴.

En nuestro estudio, utilizamos la versión 4.2 del programa RASP4 para llevar a cabo la reconstrucción ancestral de los clados de los genotipos endémicos. Aplicamos el análisis BBM (Bayesian Binary MCMC) para categorizar el origen geográfico de los aislados de los clados en seis áreas distintas: Comunidad Valenciana (A), Oceanía (B), Europa (C), América (D), Asia (E) y África (F).

Llevamos a cabo una minuciosa exploración e identificación de los nodos que presentaban una ancestralidad igual o superior al 80% en la Comunidad Valenciana. Finalmente, marcamos estos nodos en la filogenia global que habíamos construido previamente.

Resultados

Identificación y distribución geográfica de genotipos en la Comunidad Valenciana

En nuestro estudio, hemos utilizado el software FastBaps para llevar a cabo la identificación de genotipos en la región de la Comunidad Valenciana. El análisis nos condujo a la identificación de un conjunto de 75 genotipos únicos, como se muestra de manera gráfica en la Figura 4.2.

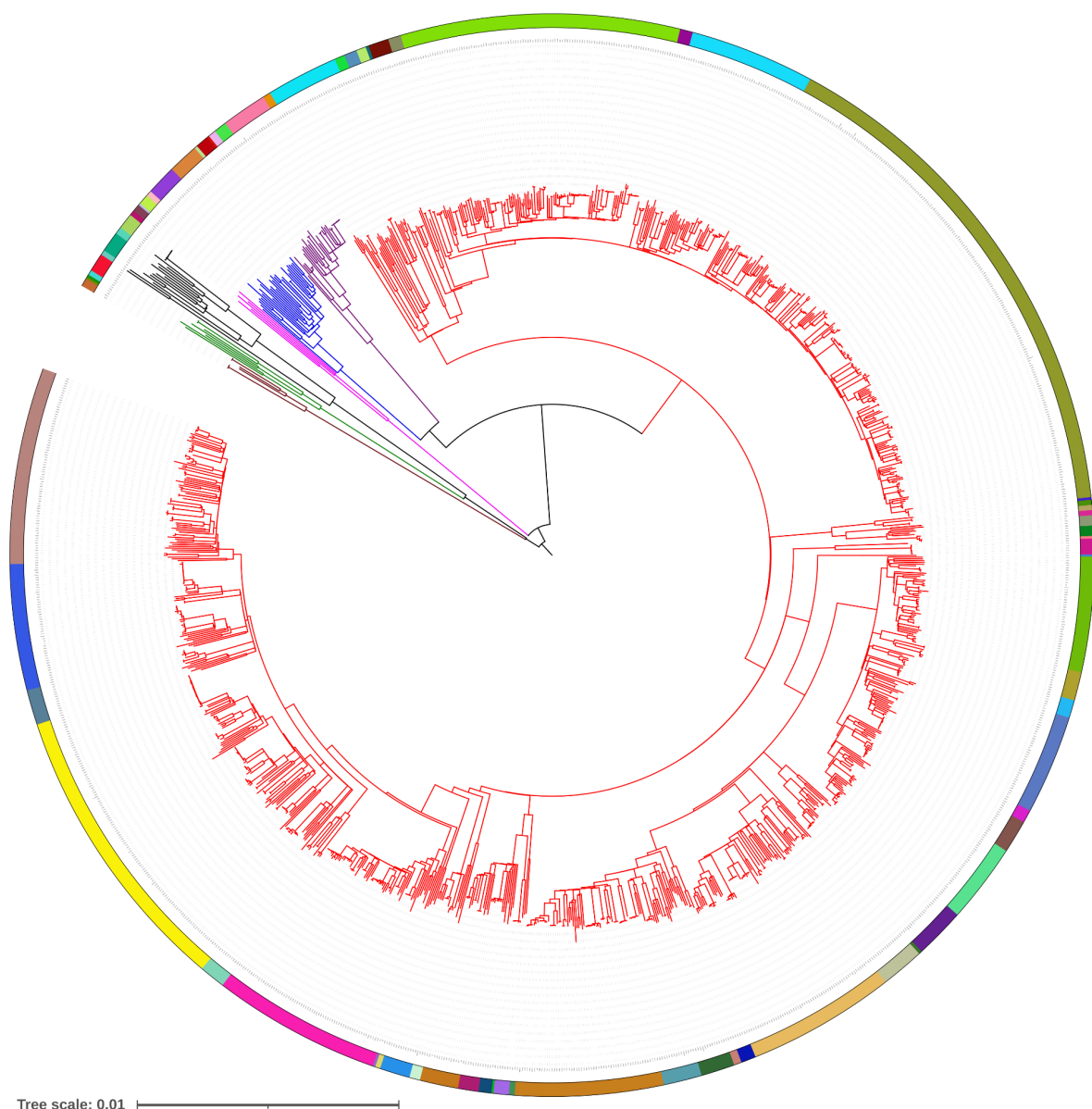


Figura 4.2. Filogenia de 1277 aislados de TB recogidos durante los años 2014 - 2019. El anillo representa los 75 genotipos locales identificados por el análisis del programa FastBAPS. Linajes: L1 en rosa, L2 en azul, L3 en púrpura, L4 en rojo, L5 en café, L6 en verde y las cepas animales en negro.

En este estudio, se realizó una evaluación de la distribución geográfica y temporal de los casos de tuberculosis (TB) en la Comunidad Valenciana. Para este propósito, se utilizaron las coordenadas geográficas correspondientes a los códigos postales de los casos registrados. Los resultados revelaron que los casos se dispersaron a lo largo de las tres provincias que componen la comunidad (Castellón, Valencia y Alicante), siendo Valencia la provincia con el mayor número de casos, destacando la ciudad de Valencia como la de mayor incidencia. En relación con la dimensión temporal, se documentaron los casos por año de declaración, observándose una variación a lo largo de los años: 233 casos en 2014, 253 en 2015, 262 en 2016, 232 en 2017 y 211 en 2018.

Posteriormente, se procedió a mapear la distribución geográfica de los 1277 aislados, clasificados en 75 genotipos, empleando la plataforma Microreact ²²⁸. Este análisis permitió visualizar la distribución de casos y genotipos en la Comunidad Valenciana. Se destacó una concentración de casos en las respectivas capitales de las tres provincias, siendo Valencia ciudad la que presentó la mayor concentración, con aproximadamente 400 casos. Los resultados están representados de forma gráfica en la Figura 4.3 de este capítulo y pueden ser visualizados en:

<https://microreact.org/project/jmJZB3HNAMMcepV9j3VE3L-baps75valencia>.

Para evaluar si existe una relación entre la distancia genética y geográfica en nuestra muestra de genotipos, se llevó a cabo un análisis de correlación utilizando el Test de Mantel. Los resultados revelaron una correlación nula ($r_m = 0.024$; $p < 0.036$) al considerar todas las muestras en conjunto, así como al analizar los 75 genotipos de manera individual. Estos resultados indican que no existe una relación directa entre la distancia genética y la distancia geográfica para el conjunto de datos, ni para cada genotipo identificado (ver Tabla Suplementaria 4.3).

En resumen, los resultados del test de Mantel no respaldan una asociación significativa entre la distribución de los genotipos de tuberculosis en la Comunidad Valenciana y la proximidad geográfica o una región geográfica específica. Este descubrimiento indica que factores adicionales, como la movilidad de las personas o la capacidad infecciosa de los genotipos, ejercen una influencia importante en la propagación de la tuberculosis en esta comunidad.



Figura 4.3. Localización geográfica de la comunidad valenciana y distribución de los 75 genotipos identificados por el programa FastBAPS.

Al analizar la distribución de genotipos, observamos que, en promedio, hay 17 aislados por genotipo, con un rango que varía de 1 a 202. Destacamos que 61 genotipos tienen menos de 20 aislados, de los cuales 49 tienen menos de 10. Estos hallazgos sugieren una amplia diversidad genética, sin embargo, se observa una predominancia de ciertos genotipos, lo que indica una representación desigual de la variabilidad genética.

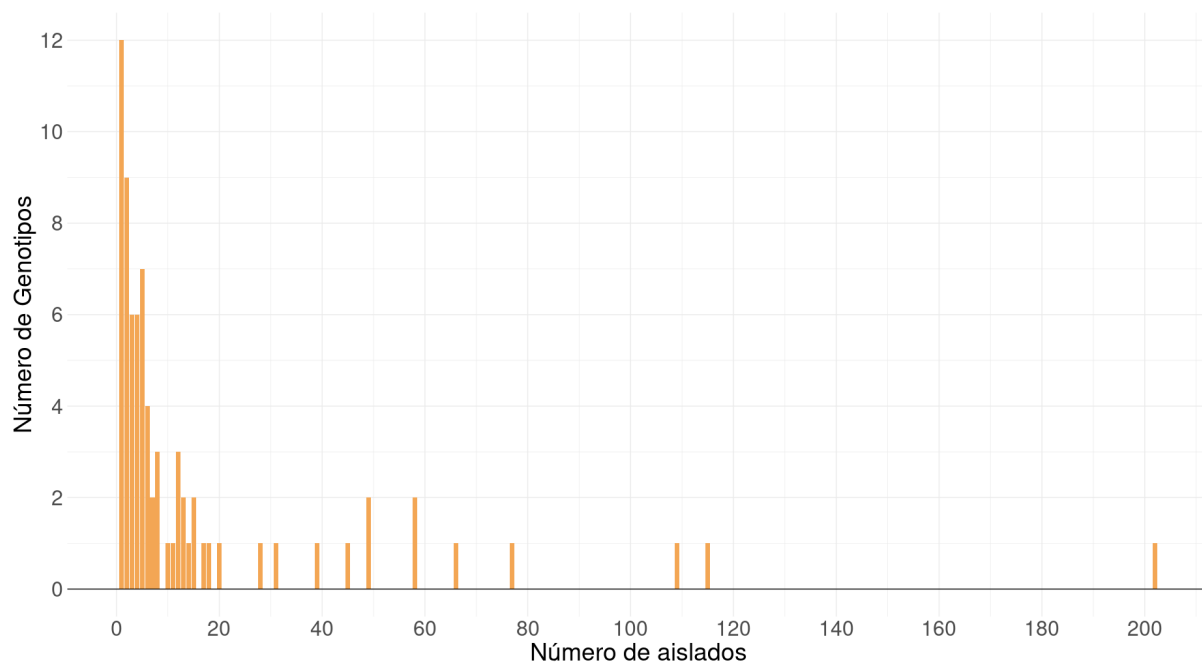


Figura 4.4. Distribución del número de aislados en los genotipos.

A partir de los resultados previos, hemos seleccionado 14 genotipos del linaje 4, los cuales cuentan con más de 20 aislados, como se muestra en la Figura 4.4 y se ilustra en la Figura 4.5. Tras la selección de estos 14 genotipos, realizamos una actualización en la reconstrucción filogenética de nuestra colección. En esta instancia, optamos por utilizar exclusivamente una muestra representativa de cada clúster de transmisión, reduciendo así el número de aislados de 1277 a 831, como se detalla en la Figura 4.5. Esta decisión fue tomada con el objetivo de evitar una posible sobrerrepresentación de los genotipos al incluirlos en nuestro estudio filogeográfico global.

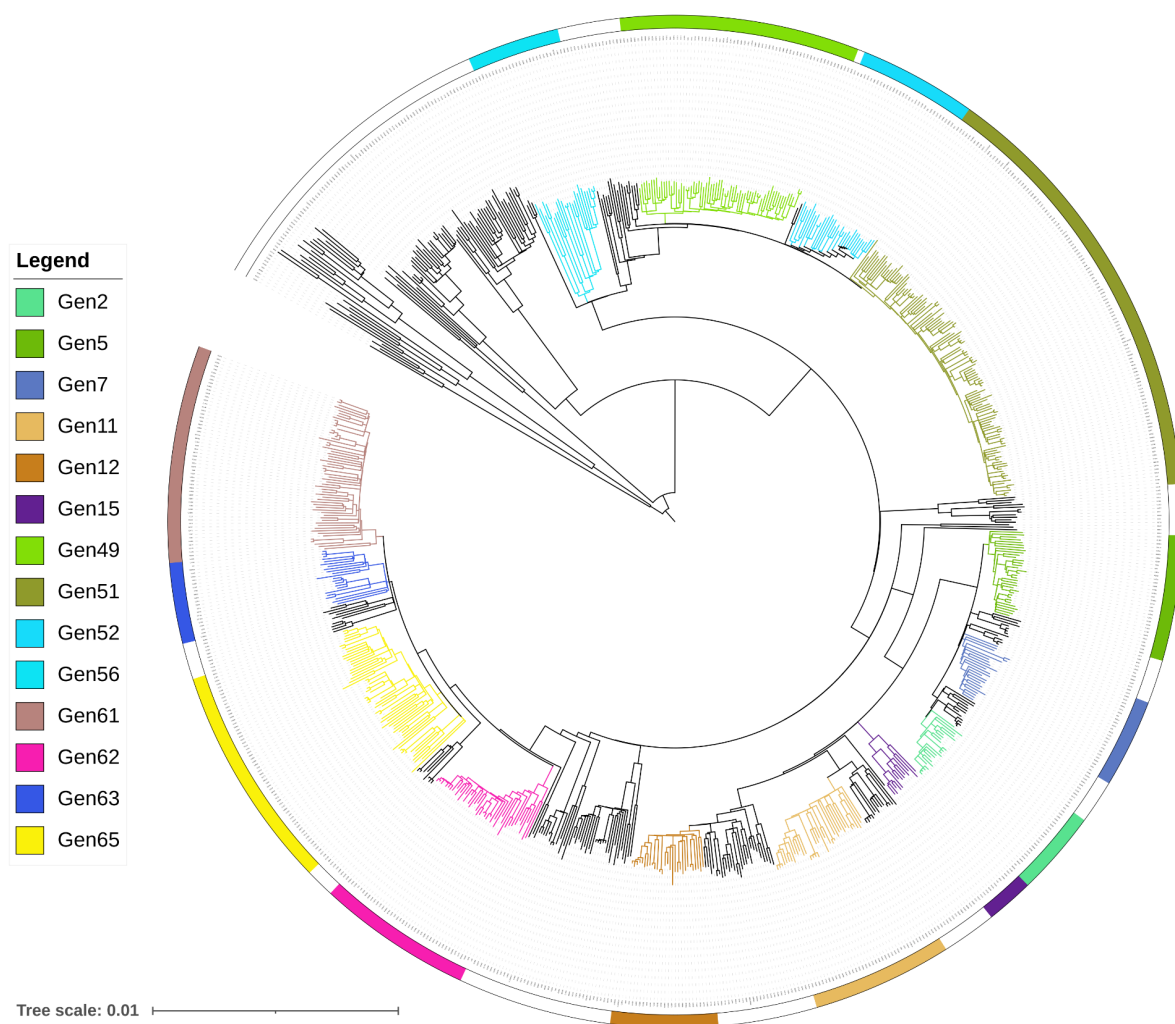


Figura 4.5. Filogenia de 831 aislados de tuberculosis recopilados entre los años 2014 y 2019. El anillo visualiza los 14 genotipos identificados mediante el análisis del programa Fastbaps, con una presencia mínima de 20 aislados por genotipo.

Relación entre genotipos, población local y transmisión

Entre los 14 genotipos previamente seleccionados, identificamos 8 genotipos locales (2, 5, 7, 11, 56, 61, 62, 63, Fig. 4.6), es decir, aquellos con una proporción baja de casos no nacidos en España ($<30\%$). La Fig. 4.6 muestra una clara relación entre los 8 genotipos y transmisión dentro de la población nacida en España, particularmente para cuatro de ellos (2, 11, 62, 63). Sin embargo, no podemos descartar que alguno de estos genotipos sea no endémico con una amplia distribución global.

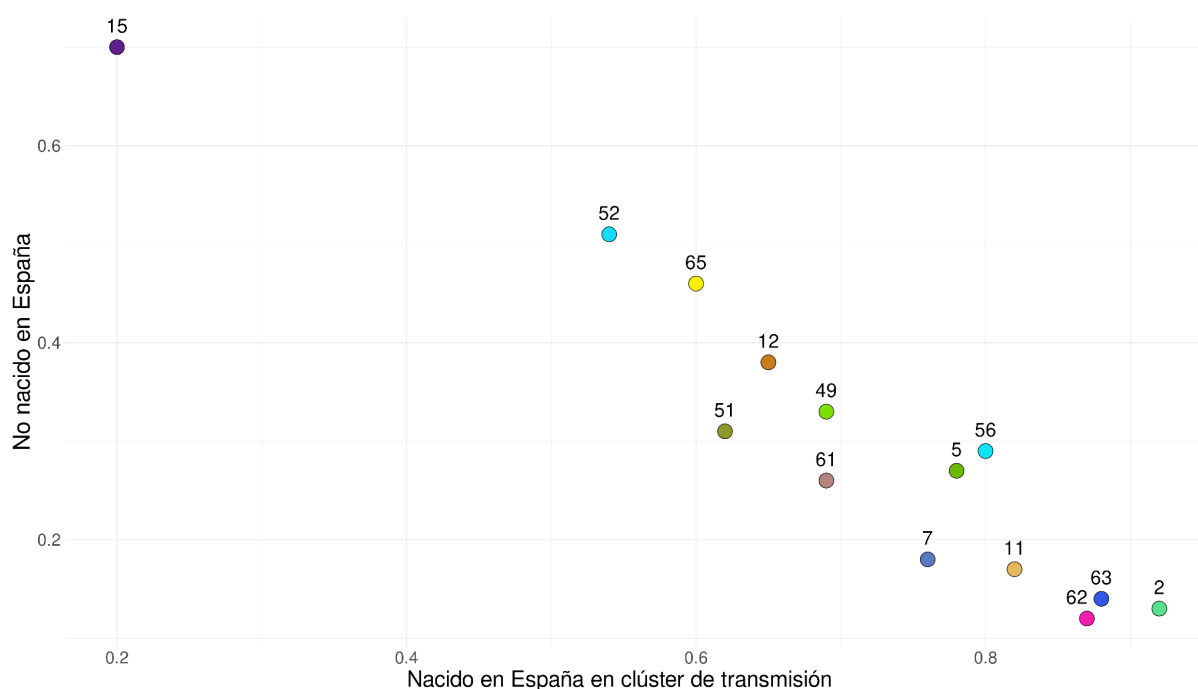


Figura 4.6. Relación genotipos vs. transmisión. Ratio de casos no nacidos en España frente a ratio de casos nacidos en España en transmisión de 14 genotipos locales de *M. tuberculosis* en la Comunidad Valenciana. Eje Y: Proporción de casos no nacidos en España en el genotipo. Eje X: Proporción de casos nacidos en España en el clúster de transmisión en el genotipo. Se identifican los genotipos 2, 11, 62 y 63 como aquellos con baja proporción de casos no nacidos en España ($<30\%$) y además los nacidos en España están mayoritariamente en un cluster de transmisión ($>80\%$ de los casos nacidos en España).

Estudio Filogeográfico e Identificación de genotipos endémicos

Para evaluar la distribución global de los posibles genotipos endémicos identificados en el apartado anterior, hemos mapeado los 8 genotipos seleccionados en una filogenia global, lo cual ha permitido visualizar su distribución en los diversos clados de la filogenia. Esta representación ofrece una perspectiva de la posición de estos genotipos en un contexto global, como se ilustra en la Figura 4.7.

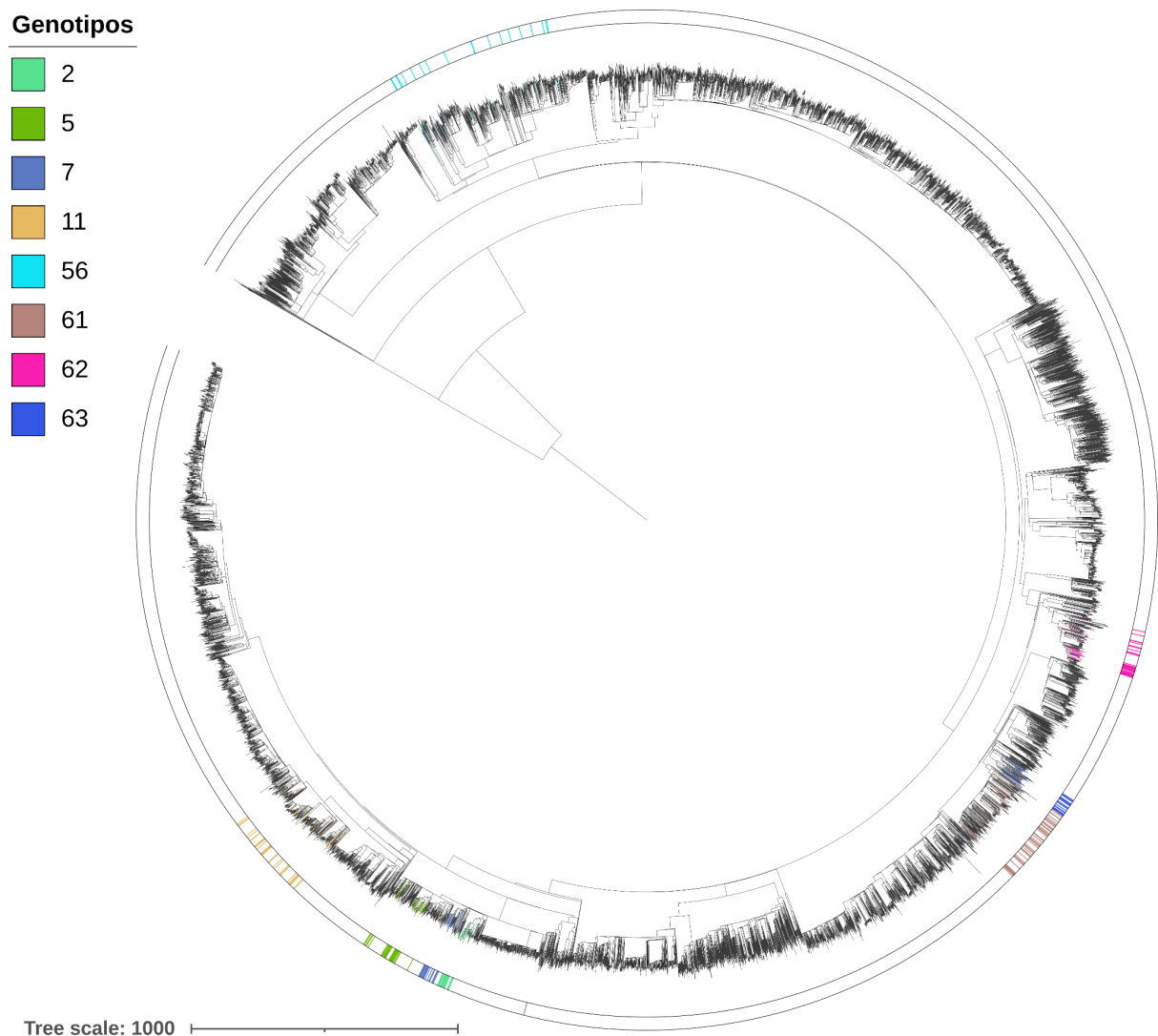


Figura 4.7. Filogenia global representativa del linaje 4, con el anillo exterior que destaca 8 genotipos locales específicos de la Comunidad de Valencia.

Los criterios empleados para la identificación de genotipos endémicos implican una proporción de aislados en la filogenia global que sea $\geq 20\%$ de casos procedentes de la Comunidad Valenciana, con al menos un 65% de esos casos siendo nacidos en España (ver Figura 4.6). Los detalles específicos de estas proporciones se presentan de forma concisa en la Tabla 4.1, subrayando la relevancia epidemiológica de estos genotipos en la región. Como resultado de esta evaluación, hemos identificado cinco genotipos específicos, a saber, los genotipos 2, 7, 61, 62 y 63, que cumplen con los criterios mencionados.

Tabla 4.1. Estadísticas de los 8 genotipos potencialmente endémicos

Genotipo	Linaje	Total de aislados	Comunidad Valenciana (%) ¹	Nacidos en España (%) ²	Selección Genotipos
2	L - 4.3.2	38	21 (55%)	17 (81%)	Si
5	L - 4.3.2	221	34 (15%)	24 (71%)	
7	L - 4.3.2	52	23 (44%)	20 (87%)	
11	L - 4.3.4.2	782	36 (5%)	28 (78%)	
56	L - 4.10	618	21 (3%)	16 (76%)	Si
61	L - 4.10	246	54 (22%)	41 (76%)	
62	L - 4.10	174	39 (22%)	32 (83%)	
63	L - 4.10	61	22 (36%)	18 (82%)	

Los datos de esta tabla fueron obtenidos de la Filogenia global sin transmisión (Figura 4.7). Se destacan en negrita las proporciones $\geq 20\%$ de aislados procedentes de la Comunidad Valenciana y $\geq 65\%$ de casos Nacidos en España. ¹ El porcentaje está calculado respecto al total de aislados incluidos en el genotipo. ² El porcentaje está calculado respecto del número de aislados de la Comunidad Valenciana.

Reconstrucción ancestral

Para delimitar el momento de introducción del clado endémico en la Comunidad Valenciana, llevamos a cabo un análisis filogeográfico. Seleccionamos los nodos más antiguos que predicen un ancestro en la región. Al concluir la reconstrucción de la ancestralidad en los clados de los cinco genotipos seleccionados (genotipos 2, 7, 61, 62, 63), identificamos nodos donde la proporción de ancestralidad del área "A" (Comunidad Valenciana) es igual o superior al 80% (nodos: 2_491, 7_539, 61_851, 62_406, 63_673) (ver Figura 4.8).

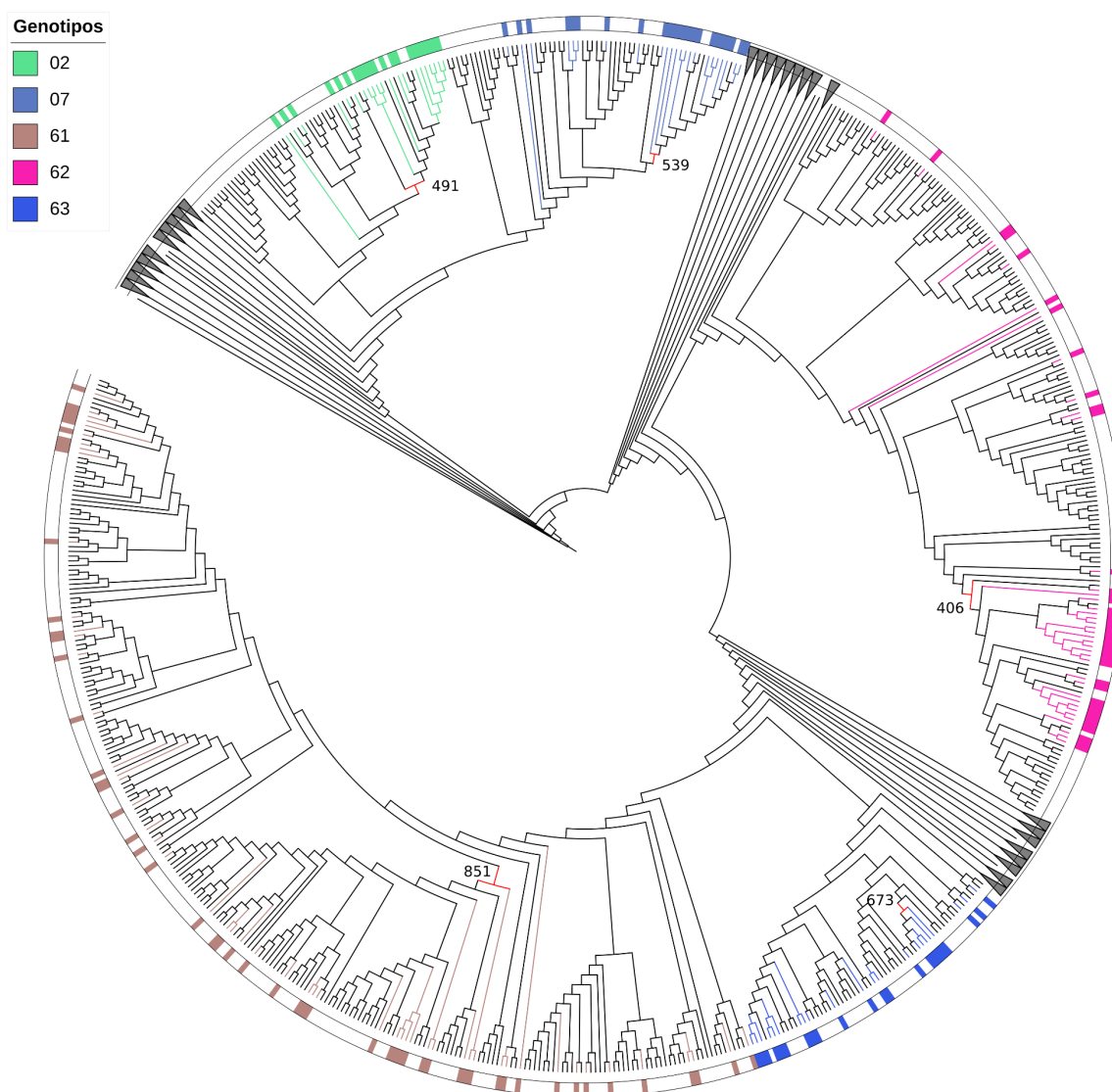


Figura 4.8. Resultados gráficos de las áreas de distribución ancestral en cada nodo de la filogenia de *M. tuberculosis*. Resultados gráficos obtenidos por modelo BBM (RASP4). Frecuencias de ancestralidad de área "A" (Comunidad Valenciana) de nodos identificados; 2_491 (82%), 7_539 (98%), 61_851 (97%), 62_406 (85%), 63_673 (95%). Los clados no pertenecientes a los genotipos endémicos se muestran colapsados.

El nodo 2_491 presentó una ancestralidad del área “A” del 82%, el mismo incluye el 55% (21/38) de los aislados totales del genotipo 2. Por su parte, el nodo 7_539 mostró una ancestralidad del área “A” del 98%, incluyendo el 33% (17/52) de los aislados totales del genotipo 7. Ambos genotipos pertenecen al linaje 4.3.2, como se detalla en la Tabla 4.2.

Además, en el nodo 61_851, se observó una ancestralidad del área “A” del 90%, abarcando el 33% (81/246) de los aislados totales del genotipo 61. Por su parte, el nodo 62_406 presentó una ancestralidad del área “A” del 85%, incluyendo el 29% (51/174) de los aislados totales del genotipo 62 (véase la Tabla 4.2). Ambos genotipos pertenecen al linaje 4.10 (véase la Tabla 4.2). Cabe destacar que los nodos 62_406, 2_491 aparecen asociados a mayor transmisión en población local en la Fig. 4.9.

Tabla 4.2. Estadísticas de los nodos identificados como genotipos endémicos

Genotipos con ancestralidad > 80% de área "A"				
Nodo	Total de aislados en el nodo	Total de la Comunidad Valenciana	Comunidad Valenciana (%)	Anc. Área "A" (%)
2_491	21	16	0.76	0.82
7_539	17	14	0.82	0.98
61_851	81	21	0.26	0.90
62_406	51	27	0.53	0.85
63_673	6	5	0.83	0.95

*Anc. Ancestralidad del área “A” (Comunidad Valenciana). Total de aislados, número total de aislados que están presentes en nodo (ver Figura 4.8).

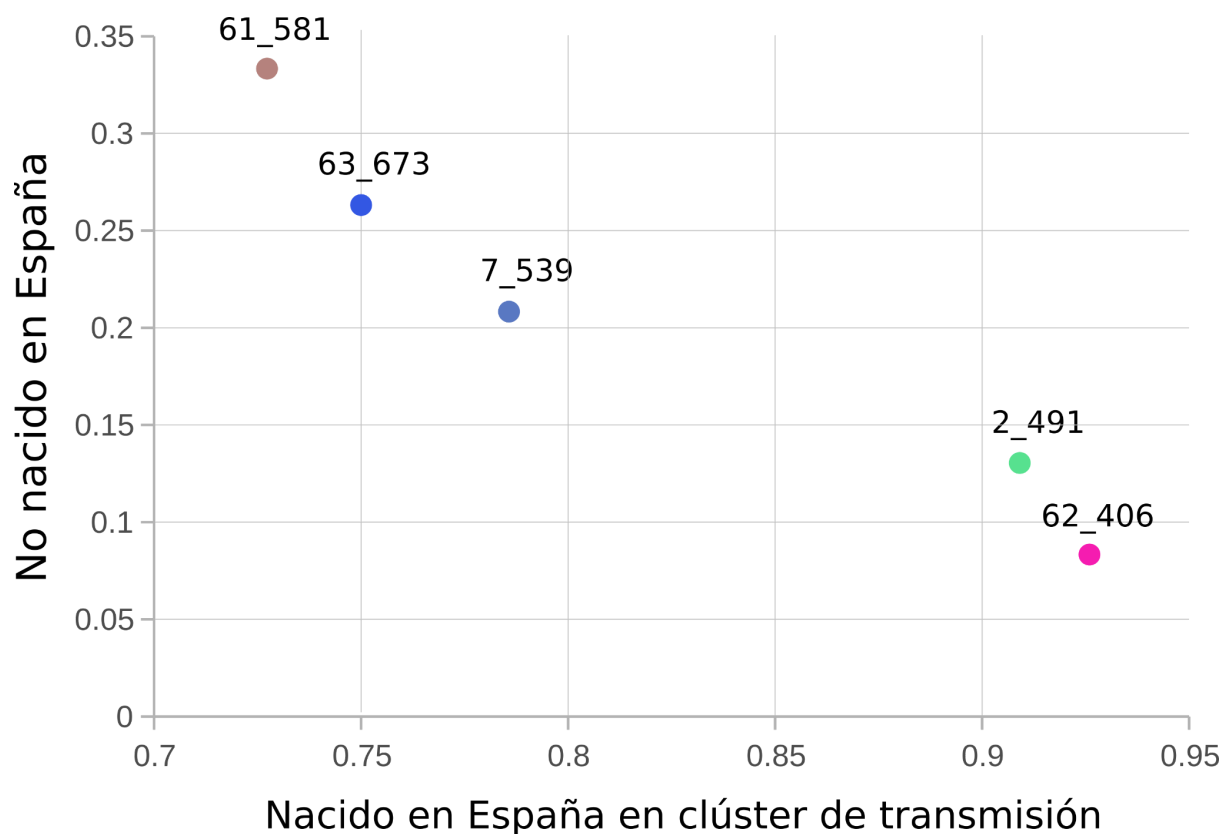


Figura 4.9. Relación entre genotipos y transmisión: Comparativa de casos no nacidos en España versus casos nacidos en España en transmisión de 5 nodos redefinidos como genotipos locales de *M. tuberculosis* en la Comunidad Valenciana. El eje Y representa la proporción de casos no nacidos en España en el genotipo, mientras que el eje X muestra la proporción de casos nacidos en España en clústeres de transmisión dentro del genotipo. Se destacan los nodos 2_491 y 62_406, con la menor proporción de casos no nacidos en España (<15%), y la mayor proporción de casos nacidos en España en clústeres de transmisión (>90% de los casos nacidos en España).

Discusión

En este estudio, hemos investigado la presencia de genotipos endémicos de tuberculosis en la Comunidad Valenciana mediante el uso de diversos análisis. Nuestra metodología ha permitido superar las limitaciones de estudios similares, los cuales se han visto restringidos en términos de muestras y técnicas de tipificación.

En primer lugar, hemos llevado a cabo una clasificación de una colección de 1277 aislados de *M. tuberculosis* en 75 genotipos, evidenciando así la diversidad genética del MTBC en la Comunidad Valenciana (Figura 4.3). Nuestros hallazgos indican que los genotipos de MTBC presentes en esta región pertenecen a siete de los nueve linajes principales, siendo el L4 el más abundante. Estos resultados concuerdan con la distribución global del L4¹⁸³.

Un elemento destacado en nuestro estudio es la detección de una amplia distribución de los genotipos en toda la región, sugiriendo una transmisión efectiva de la enfermedad sin sesgo geográfico aparente. Esto plantea la idea de que el panorama de la tuberculosis en la Comunidad Valenciana es dinámico, caracterizado por una amplia diversidad de genotipos dispersos a lo largo de la región, lo cual coincide con hallazgos previos en investigaciones similares²³⁵.

Esta observación genera preguntas acerca de los mecanismos responsables de la exitosa distribución de estos genotipos. Además, descarta la hipótesis de que la proximidad geográfica entre individuos sea el factor principal en la transmisión de los genotipos, lo que nos lleva a considerar que otros factores desempeñan un papel más significativo en este proceso.

En relación con los genotipos endémicos, estos pertenecen al linaje cuatro, específicamente a los sublinajes 4.3.2 (genotipos 2_491 y 7_539,) y 4.10 (genotipos 61_851, 62_406, 63_673). Es relevante señalar que el sublinaje 4.10, en general, tiene una frecuencia alta en España^{28,174}. En contraste, la mayoría del resto de genotipos muestran una estrecha relación genética con genotipos de otras partes del mundo. Esto es comprensible, ya que la Comunidad Valenciana no se encuentra aislada, sino que experimenta un alto flujo de personas de diversas regiones.

En cuanto a las limitaciones de la clasificación de los genotipos, no hemos determinado si los genotipos identificados en nuestra investigación fueron introducidos recientemente o han estado presentes durante mucho tiempo en la región. Por lo tanto, no podemos discernir si los genotipos endémicos son el resultado de una coevolución con la población o de una introducción mejor adaptada. Un análisis temporal podría contribuir a una mejor comprensión de algunos de nuestros genotipos endémicos. Además, recomendamos que futuras investigaciones contemplen la expansión de la filogenia global de referencia, lo que podría reevaluar el estatus de algunos de nuestros genotipos endémicos o definirlos con mayor precisión.

En cuanto a las áreas geográficas utilizadas en nuestro análisis de RASP4, las definimos como Comunidad Valenciana (A), Oceanía (B), Europa (C), América (D), Asia (E) y África (F). Ha surgido la posibilidad de reconsiderar la inclusión del área "A" (Comunidad Valenciana) dentro del área "C" (Europa). Llevamos a cabo diversas combinaciones y obtuvimos resultados similares en nuestros estudios. Sin embargo, es importante destacar que la limitación de no poder identificar el país de origen de todas las muestras en nuestra colección y la escasez de secuencias genómicas de la península ibérica disponibles en la actualidad han sido obstáculos significativos en este análisis.

En el análisis de la transmisión de los nodos endémicos (Figura 4.9), se ha identificado una mayor tasa de transmisión de los genotipos 2_491 y 62_406 entre los individuos nacidos en España en comparación con los genotipos 7_539, 61_851 y 63_673. Aunque desconocemos los factores sociales y biológicos que expliquen esta asociación, consideramos que esta mayor transmisión observada en estos genotipos podría estar relacionada con su condición de endémicos. Las diferencias entre los genotipos 61_851 y 62_406, sugieren que el genotipo 61_851 se ha expandido tanto en población local como en nacidos fuera de España. Podría tratarse de un genotipo más generalista (usando la terminología de Stucki ¹⁷⁴). En contraste, el genotipo 62_406 parece haberse expandido sobre todo entre individuos nacidos en España. No obstante, es importante considerar que esta diferencia podría estar influida por prácticas sociales, donde los portadores del genotipo 62_406 podrían tener interacciones más limitadas en entornos locales. Esto sugiere que asociar un genotipo, linaje o sublinaje específico con huéspedes humanos puede generar sesgos si no se tienen en cuenta estos factores.

En el análisis refinado filogeográfico se destaca la presencia notoria de los genotipos 2_491 y 7_539 pertenecientes al linaje 4.3.2, los cuales exhiben clados de ancestralidad muy similares. Es plausible que estos genotipos compartan un fondo genético común que les confiere una mayor eficacia en su capacidad de transmisión entre la población nacida en España. Por otro lado, el genotipo 63_673 se caracteriza por presentar únicamente un pequeño clado compuesto por seis aislados y no lo podemos considerar en esta discusión.

Nuestra investigación, en última instancia, coloca al genotipo 62_406 como el protagonista central. Esto se debe a su notable cantidad de aislados, su marcada ancestralidad proveniente de la Comunidad Valenciana y su eficaz capacidad de transmisión entre individuos nacidos en España. Estos factores respaldan la consideración de este genotipo como endémico de la región, dado que cumple con una mayor cantidad de los criterios establecidos en nuestro estudio.

Es relevante destacar que la adaptación genética de los genotipos de *M. tuberculosis* a la población de huéspedes ha sido propuesta en investigaciones previas, lo que podría explicar la prevalencia de ciertas variantes ^{174,236}. No obstante, es importante tener en cuenta que estudios en otras regiones han sugerido que no existe una diferencia significativa en cuanto a la virulencia entre los distintos genotipos ²³⁷.

En resumen, mediante rigurosos criterios de selección, hemos identificado cinco genotipos endémicos (2_491, 7_539, 61_851, 62_406 y 63_673) en la Comunidad Valenciana. De estos, dos muestran una mayor afinidad por infectar a la población de origen español (genotipos 2_491 y 62_406), y uno de ellos (genotipo 62_406) cumple de manera satisfactoria con todos nuestros criterios de selección. En nuestra investigación, el próximo paso consistirá en realizar pruebas *in vitro* con el propósito de evaluar la coevolución. Durante estas pruebas, observaremos la interacción entre los genotipos catalogados como endémicos y las células inmunitarias humanas. Durante el proceso, mediremos la respuesta inflamatoria y formularemos la hipótesis de que una respuesta inmunitaria menor (menos inflamación) podría correlacionarse con una mayor adaptabilidad de los genotipos al huésped humano.

5. Estimación del tamaño del cuello de botella en la transmisión de la tuberculosis humana a partir de la diversidad genómica donante-receptor.

Introducción

La transmisión de *M. tuberculosis* se produce cuando un individuo infectado (donante) exhala un bioaerosol que contiene una cantidad y viabilidad desconocidas de bacilos. Una parte variable de estos bacilos es inhalada por un segundo individuo (receptor). Durante la transmisión, la población bacteriana que se transfiere del donante al receptor pasa por distintas fases en las que se filtran y seleccionan diferentes subpoblaciones. En la literatura científica, este proceso se conoce como "bottleneck" o "cuello de botella" ²³⁸.

El cuello de botella está influenciado por diversos factores y variables. En primer lugar, únicamente una pequeña proporción de la población bacteriana presente en el donante se transmite al receptor. Dentro de esta población transmitida, sólo una fracción logra establecerse en los bronquiolos terminales y dar origen a la infección ⁵⁹. A esta fracción de la población transmitida, que tiene éxito en el establecimiento de la infección y persiste hasta el momento del muestreo, se le denomina población fundadora ²³⁹ (ver Figura 5.1).

En el presente capítulo, los términos "población fundadora" y "tamaño del cuello de botella" serán tratados como sinónimos debido a su estrecha relación. Ambos términos se refieren al número de bacilos que inician la infección y se utilizarán indistintamente en este contexto.

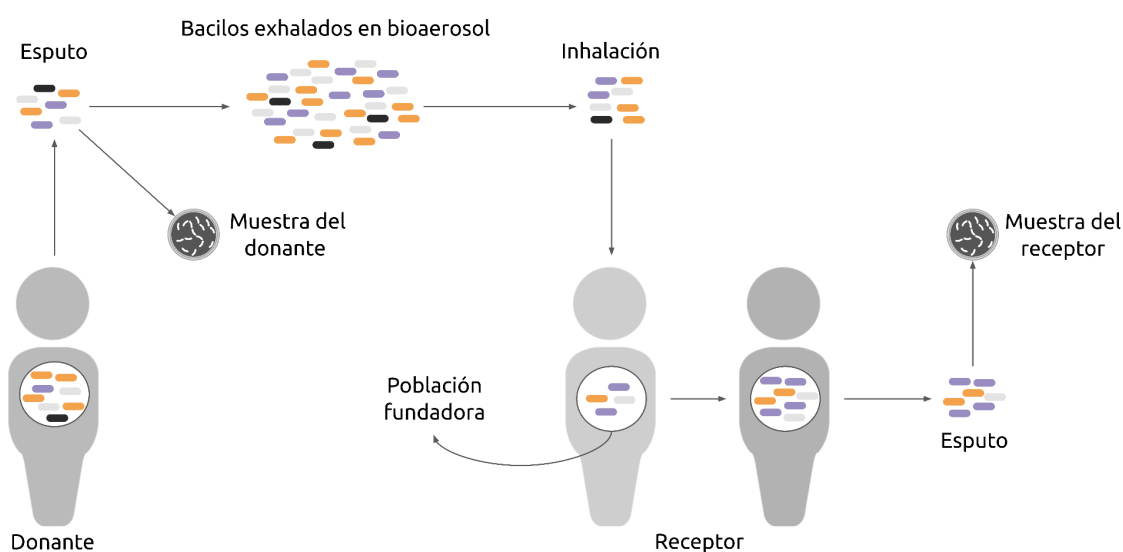


Figura 5.1. Esquema de la transmisión de la TB en una pareja donante - receptor.

El número de bacilos que componen la población fundadora en la infección de la TB humana ha sido una incógnita desde que Robert Koch anunció, en 1882, el descubrimiento del agente causante de la TB, *Mycobacterium tuberculosis* ²⁴⁰. Las primeras aproximaciones para cuantificar bacilos en pacientes con TB positiva se enfocaron en cuantificar, por microscopía, el número de bacilos en las microgotas de tos expulsadas. Se observó que estas gotas presentaban un intervalo de tamaños que va desde $< 30 \mu\text{m}$ hasta $500 \mu\text{m}$ de diámetro. En las gotas más pequeñas se observaron de 1 a 2 bacilos, mientras que en las más grandes se encontraron más de 200 ^{241,242}. Otros estudios han informado la presencia de un solo bacilo en microgotas de $15\text{-}20 \mu\text{m}$ y de 15-30 bacilos en gotas de $100 \mu\text{m}$ ²⁴³.

Durante años se ha considerado que un solo bacilo era capaz de infectar y causar TB, debido a las pruebas presentadas en modelos animales utilizando vías de infección artificiales, como la inyección intra-traqueal ²⁴¹. Experimentos realizados en cobayas, mediante la exposición a aerosoles de diferentes dosis (20-290.000 bacilos), han demostrado que incluso a dosis muy bajas (20-40 bacilos) se produce infección; en contraposición dosis orales conteniendo hasta 290.000 bacilos no han resultado exitosas ⁶². Recientemente, se han logrado infecciones artificiales por aerosoles en ratones con dosis ultrabajas (1-3 bacilos) ²⁴⁴. Estos ensayos han permitido el control meticuloso de diversas variables, tales como la dosis de infección, el sitio inicial de la misma, así como la evaluación de la resistencia, hipersensibilidad y virulencia. La información derivada de estos modelos animales se vuelve de gran relevancia para una

mejor comprensión de la tuberculosis en seres humanos. Sin embargo, al tratarse de infecciones provocadas, los valores del tamaño de la población fundadora de la infección no pueden ser extrapolados a las infecciones naturales.

En los últimos años se ha avanzado mucho en la detección de bioaerosoles exhalados por pacientes de tuberculosis. Diferentes estudios han estimado que un individuo infectado puede exhalar una media de 336 bacilos vivos al día ^{245,246}; además, se ha demostrado la exhalación de bacilos durante los períodos nocturnos y también en ausencia de tos ⁶⁰. Finalmente, un nuevo modelo estadístico indica que la respiración de marea (respiración normal) aporta más del 90% de los bacilos exhalados al día ²⁴⁷.

Aunque todos estos estudios son muy útiles para comprender la disponibilidad/accesibilidad de las bacterias en el ambiente (aire), no profundizan en el destino de estos bacilos una vez han sido inhalados por un individuo (receptor). Conocer el número de microorganismos inhalados y capaces de fundar una infección es importante para comprender su evolución, transmisión e historia natural de la enfermedad y tiene implicaciones para el diagnóstico, tratamiento, control y vigilancia de la tuberculosis ^{70,248}.

En la infección natural (es decir, accidental) en el ser humano, es imposible cuantificar la dosis inicial de bacilos con algún grado de precisión ²⁴⁹. En los últimos años, y utilizando datos de secuenciación masiva, se han desarrollado algunos métodos estadísticos para estimar los tamaños de los cuellos de botella, entre los cuales se encuentran; el método de presencia/ausencia ^{250,251}, el método de muestreo binomial ^{252,253} y el método de muestreo beta-binomial ²³⁹.

Estos métodos fundamentan sus estimaciones en la identificación y comparación de variantes (SNPs) entre el donante y el receptor. Es relevante destacar que los tres métodos mencionados parten del supuesto de que la diversidad genética del patógeno es completamente neutra, es decir, que la selección no ejerce ningún impacto en la dinámica de la frecuencia de las variantes. Además, dichos métodos presuponen la independencia entre los distintos sitios de los polimorfismos (SNPs). Por último, conviene resaltar la relevancia de la profundidad de secuenciación en las estimaciones de dichos métodos, una elevada profundidad (>400X) disminuye considerablemente la presencia de falsos negativos y falsos

positivos, los cuales tienen efectos significativos en el cálculo del tamaño del cuello de botella, especialmente en el caso de variantes de baja frecuencia.

El método de presencia/ausencia destaca por su simplicidad; se basa en determinar si los SNPs detectados en los donantes están presentes o ausentes en los receptores. Sin embargo, es importante tener en cuenta que debido al uso frecuente de umbrales restrictivos para el llamado de las variantes, este método es especialmente propenso a acumular falsos negativos, lo que significa que no se llaman variantes que sí están presentes en el receptor. La falta de consideración de estos falsos negativos en el receptor conduce a una subestimación del tamaño del cuello de botella.

El método de muestreo binomial, por el contrario, analiza las variantes comunes, es decir las que se llaman tanto en el donante como en el receptor. Considera las frecuencias de estas variantes, teniendo en cuenta el número de lecturas conteniendo las variantes y el número total de lecturas en cada posición. Este método asume que la población de patógenos en el receptor es una muestra representativa de la población del donante. La principal limitación de este método es que no considera las variantes del donante que no se llaman en el receptor, a pesar de que esta ausencia aporta información relevante. Pasar por alto estas variantes puede introducir un sesgo significativo en las estimaciones del tamaño del cuello de botella.

El método de muestreo beta-binomial es similar al método de muestreo binomial, pero con la particularidad de que tiene en cuenta la replicación aleatoria del patógeno al comienzo de la infección. Esto permite que las frecuencias de las variantes (SNPs) del receptor puedan diferir entre el momento del muestreo y la infección. Además, este método considera tanto la posibilidad de que una variante del donante esté realmente ausente en el receptor como la posibilidad de que esté por debajo del umbral de llamado de variantes. Como resultado, proporciona estimaciones más precisas en comparación con el método de muestreo binomial.

Los estudios que emplean el método de muestreo beta-binomial ^{239,254} han informado de una baja diversidad genética en el huésped y tamaños de cuello de botella de transmisión estrechos en virus como la influenza A y el SARS-CoV-2, enfermedades agudas transmitidas por aerosoles. Esto indica que, en estos casos, una cantidad reducida de viriones puede ser suficiente para establecer la infección. Sin embargo, estos estudios se han centrado en infecciones agudas, y hasta el momento el método de muestreo beta-binomial no ha sido

utilizado para estimar el tamaño del cuello de botella en infecciones crónicas con periodos de latencia variables y potencialmente largos, como en el caso de la tuberculosis.

La diversidad de *M. tuberculosis* ha sido estudiada desde el nivel global ³⁶ hasta el intra hospedador ³⁷, sin embargo, pocos estudios han analizado la diversidad de *M. tuberculosis* en pares de transmisión donante-receptor ⁵⁶. El estudio de la diversidad es un desafío debido a que la detección de variantes (SNPs) a frecuencias mínimas (<5%) dentro de las poblaciones bacterianas está condicionada por factores como la profundidad de secuenciación, los cambios en el tamaño de la población y la selección positiva y/o purificadora ⁵³.

En este capítulo, nuestro objetivo es describir la diversidad genética de *M. tuberculosis* durante la transmisión reciente empleando el método de muestreo beta-binomial, descrito en la literatura ²⁵⁵, para estimar el tamaño del cuello de botella en pares de transmisión donante-receptor. Cuantificar el tamaño de la población causante de la infección es de gran interés para aumentar nuestro conocimiento sobre la dinámica de la transmisión de la TB.

Métodos

Nuestra metodología se resume en los siguientes puntos: población de estudio y selección de muestras, identificación de pares de transmisión, extracción y secuenciación de DNA, análisis bioinformático, reconstrucción filogenética, validación de variantes (SNPs) , y estimación del tamaño del cuello de botella.

Población de estudio y selección de pares de transmisión donante-receptor

En este capítulo se utilizaron los datos analizados en el capítulo tres, que incluyen información sobre el linaje genético, el perfil de resistencia a los fármacos, la presencia de infecciones mixtas, las fechas de diagnóstico e inicio de tratamiento, y la identificación de grupos de transmisión a diferentes umbrales. Estos datos se emplearon para identificar de manera precisa los pares de transmisión donante-receptor.

Primero, se seleccionó el umbral de distancia genética igual o inferior a 10 SNPs ¹⁵⁷ para establecer los clústeres de transmisión en el análisis. Empleando un total de 181 clústeres de transmisión que involucraron 627 casos de tuberculosis (TB). En 107 de estos casos, se pudo rastrear y recopilar información epidemiológica, clínica y demográfica, incluyendo las fechas de diagnóstico y de inicio de los síntomas (Figura 5.2).

De los 107 casos analizados, se realizó una cuidadosa selección de 14 pares de transmisión donante-receptor, lo que implicó un total de 28 muestras. Se aplicaron criterios rigurosos para la identificación de estos pares, a saber: sólo se consideraron aquellos pares que pertenecieran a un clúster compuesto por dos o tres muestras, lo cual indica una alta probabilidad de transmisión directa entre los individuos del grupo. Para confirmar la relación de transmisión reciente, se estableció un umbral de distancia genética entre 0-2 SNPs.

En el proceso de identificación, se utilizaron las fechas de inicio de los síntomas y de diagnóstico como criterios esenciales para determinar el donante y el receptor en cada par de transmisión. Asimismo, se excluyeron las muestras en las cuales el donante presentaba una infección extrapulmonar, debido a que se consideró que la probabilidad de transmisión en esos casos era baja. Además, se llevó a cabo una exhaustiva revisión de los perfiles de

sensibilidad/resistencia a fármacos de los aislados presentes en los pares de transmisión. Se descartaron aquellos aislados que presentaran resistencia a los medicamentos.

Finalmente, hemos incorporado en nuestro estudio 7 pares de transmisión adicionales (14 muestras) identificados en una investigación previa llevada a cabo por nuestro grupo ¹⁵⁴. En este contexto, se ha utilizado TransPhylo, un paquete de software especializado en la reconstrucción de la transmisión de enfermedades infecciosas mediante el análisis de datos genómicos. TransPhylo se ha desarrollado específicamente para el análisis de filogenias temporales con el propósito de determinar la direccionalidad de la transmisión de patógenos entre distintos individuos (Figura 5.2).

Es relevante destacar que la baja incidencia de tuberculosis en la Comunidad Valenciana, conjuntamente con la implementación de la vigilancia epidemiológica y la vigilancia genómica correspondiente, ha favorecido la identificación de pares de transmisión, al tiempo que ha disminuido la probabilidad de incorporar individuos que cursen con una posible reinfección dentro de dichos pares.

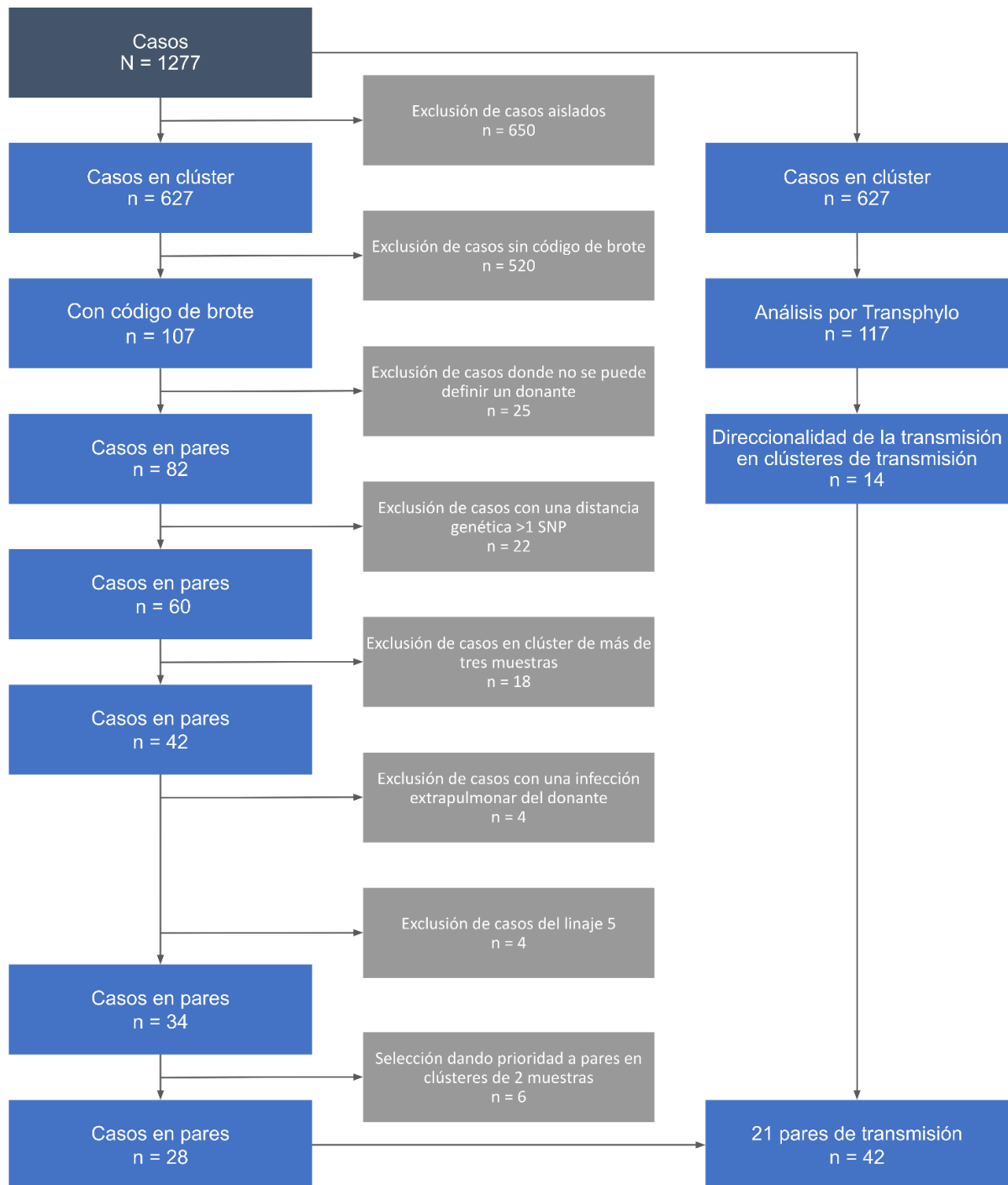


Figura 5.2. Diagrama de flujo de criterios de inclusión y exclusión

Secuenciación de alta profundidad

A partir del ADN de los aislados seleccionados se construyeron bibliotecas de secuenciación *Nextera XT DNA Library Prep Kit* (Illumina). Las mismas fueron secuenciadas en un instrumento Illumina NextSeq, utilizando una longitud de lecturas de 2 x 150 pb. La carrera fue optimizada para alcanzar una profundidad de secuenciación de 500X.

Análisis bioinformático

El preprocesamiento de las lecturas se realizó utilizando fastp ²⁵⁶ para escanear las lecturas y recortar los extremos de baja calidad con una calidad media de ventana <20. A continuación, utilizamos Kraken ^{166,256} para clasificar taxonómicamente las lecturas mediante una base de datos personalizada, manteniendo únicamente las secuencias MTBC para evitar falsas variantes debidas al ADN contaminante. Las lecturas filtradas se fueron mapeadas con BWA ^{166,168} usando como referencia la secuencia del ancestro de MTBC ¹⁶⁷ empleando parámetros por defecto, y se procesaron usando Samtools y Picard ^{257,258}. Por último, se identificaron y eliminaron duplicados ópticos y de PCR para reducir el número de variantes (SNPs) artificiales en intervalos de baja frecuencia.

Dado que la identificación de variantes (SNPs) de baja frecuencia es clave en nuestro estudio, pero también es un reto distinguir estas variantes de los falsos positivos, utilizamos tres programas de llamado de variantes diferentes (VarScan2 ²⁵⁹, GATK HaplotypeCaller ²⁶⁰, LoFreq48 ²⁶¹).

- VarScan2 fue ejecutado con los siguientes parámetros “pileup2snp sample.pileup—min-coverage 20—min-reads2 4—min-avg-qual 20—min-var-freq 0.01—min-freq-for-hom 0.9—strand-filter 1”.
- GATK fue ejecutado con los parámetros “-T HaplotypeCaller -R ref.fasta -I sample.bam -o sample.vcf—min-base-quality-score 20 -ploidy 100”
- LoFreq fue ejecutado con los parámetros “call-parallel—pp-threads 12 -f ref.fasta -o sample.vcf sample.bam” and “filter -i sample.vcf -v 20 -A 0.01 -Q 20 -o filtered.vcf”.

Para establecer un umbral mínimo para la llamada de variantes (SNPs), realizamos una simulación de lectura sintética. Utilizando el paquete de software ART ²⁶², obtuvimos la distribución de calidad y los perfiles de error de nuestras muestras (utilizando art_profiler_illumina con parámetros por defecto), y simulamos 100 carreras de secuenciación con los siguientes parámetros:

- *art_illumina -p -1 profile.txt -2 profile.txt -na -iref.fasta -l 150 -f 610 -m 770 -s 500 -o out.fas*

Analizando las simulaciones con el mismo *pipeline*, definimos un umbral de frecuencia mínima del 2% para validar una variante en las muestras con los tres programas (VarScan2²⁵⁹, GATK's HaplotypeCaller²⁶⁰, LoFreq48²⁶¹).

Una vez obtenida la lista de variantes (SNPs) identificadas, se aplicó un filtro de mapeo¹⁵ con el objetivo de descartar las variantes presentes en regiones genómicas repetitivas, como las familias PE/PPE y fagos²⁶³. Estas regiones genómicas son conocidas por ser difíciles de mapear y pueden generar variantes falsas en los resultados. Además, se amplió el filtro de mapeo a nuevas regiones que mostraban una alta frecuencia de SNPs en las simulaciones. De esta manera, se aseguró la precisión y confiabilidad de las variantes identificadas en el estudio.

Finalmente, construimos la lista completa de SNPs para cada muestra, incluyendo aquellas posiciones que fueron identificadas en, al menos, dos de los programas de llamado de variantes (SNPs). Se descartaron aquellas variantes con frecuencia $\leq 2\%$. Posteriormente, generamos una lista de posiciones comunes a las 42 muestras. Las posiciones comunes fueron “rescatadas” en aquellas muestras donde estaban presentes a una frecuencia $< 2\%$.

Reconstrucción filogenética

Se generó un archivo multifasta, concatenando 4194 SNPs de las muestras (21 pares de transmisión = 42 aislados). A partir de este archivo, se reconstruyó una filogenia de máxima verosimilitud (ML, *maximum likelihood*); empleando el programa IQ-TREE 2¹⁷³ con el modelo GTR y 1.000 réplicas *bootstrap*. Además, se calcularon las distancias genéticas de SNPs entre los pares de transmisión utilizando el método descrito en el capítulo tres.

Validación de variantes (SNPs)

En el presente estudio, se identificaron un total de 1826 posiciones en los 21 pares de transmisión después de la implementación de nuestro análisis informático. Entre ellas, se encontraron 1093 posiciones únicas, es decir, identificadas en un solo aislado, y 733 posiciones compartidas en dos o más aislados, perteneciente a distintos pares de transmisión.

Utilizando la lista de variantes compartidas (SNPs) y sus respectivas frecuencias, se procedió a calcular la amplitud de rango de frecuencias para cada una de las posiciones. Por ejemplo, al identificar una posición, en los aislados A, B y C con frecuencias de 0.02, 0.04 y 0.05 respectivamente, se determinó que la amplitud de rango es de 0.03. Los valores de amplitud de rango fueron graficados (ver Figura 5.4).

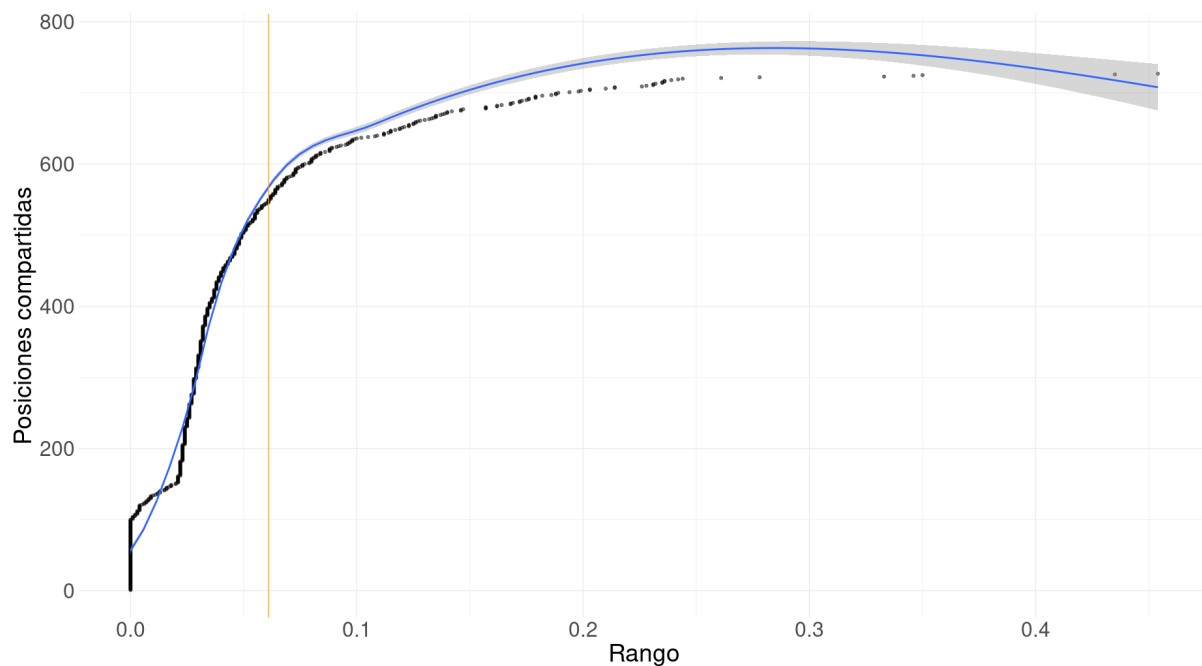


Figura 5.4. Posiciones compartidas de las 42 muestras ordenadas por amplitud de rango
Se muestra el percentil a 0.75 (amplitud de rango 0.061) en línea amarilla.

Al analizar el gráfico, se pudo identificar una relación y concentración de posiciones compartidas que presentaban una amplitud de rango por debajo de 0.1. Para evaluar esta distribución, se calcularon los percentiles 25, 50 y 75. Después de considerar estos valores, se seleccionó el percentil 75, que corresponde a una amplitud de rango de 0.061, como punto de corte. Esta elección se basó en la observación de que a partir de este valor se evidenciaba una pérdida de la relación previamente observada (ver Figura 5.4).

Al aplicar el criterio mencionado, se procedió a descartar aquellas posiciones compartidas entre los 42 aislados que presentaban una amplitud de rango inferior a 0.061. Como resultado de esta evaluación, se conservaron un total de 227 posiciones compartidas que cumplieran con una amplitud de rango igual o superior a 0.061.

Por otro lado, se excluyeron del análisis un total de 506 posiciones compartidas que no cumplían con dicho criterio. Este proceso de exclusión permitió reducir el número de posiciones a analizar de 1826 a 1320, compuesto por 1093 posiciones únicas y 227 posiciones compartidas. De esta manera, al aplicar este procedimiento de selección, se logró enfocar el análisis en un conjunto más específico de posiciones genómicas relevantes.

Estimación del tamaño del cuello de botella

Nuestro análisis para estimar los tamaños de los cuellos de botella de transmisión para cada par donante-receptor se basó en el método de muestreo beta-binomial ²³⁹. El modelo se basa en comparar las frecuencias de cada variante (SNP) entre donante y receptor, teniendo en cuenta que las frecuencias de las variantes (SNPs) en el receptor pueden cambiar entre el momento de la infección y el de muestreo, debido al proceso de replicación aleatorio de las bacterias. En aquellos casos en que la variante no es detectada en el receptor, el modelo evalúa la probabilidad de que se deba, o bien a una ausencia real, o a un error en el llamado de esa variante.

Para un dado valor del tamaño del cuello de botella, el modelo estima la verosimilitud (*likelihood*) para cada variante (SNP). La verosimilitud total, para cada valor de tamaño de cuello de botella, es el producto de las verosimilitudes estimadas para cada variante. Finalmente, el valor inferido del tamaño de cuello de botella, es aquel que presenta la máxima verosimilitud (MLE).

Se han desarrollado dos versiones del método de muestreo beta-binomial: una versión aproximada que asume una profundidad de lectura infinita y una versión exacta que tiene en cuenta la profundidad de lectura para compensar el ruido y la variación que surge en un experimento real. Se espera que la estimación de máxima verosimilitud (MLE) de la población fundadora obtenida mediante ambas versiones -aproximada y exacta- coincida. Sin embargo, esto ocurre cuando la profundidad de secuenciación es alta (>200X); en caso contrario, el método aproximado es el más adecuado. La versión aproximada tiene la ventaja de que, cuando es apropiada, es computacionalmente menos intensiva que la versión exacta que corrige las frecuencias considerando la profundidad de secuenciación de las posiciones de interés.

El método de muestreo beta-binomial se aplicó a las variantes (SNPs) en el intervalo de frecuencia de 0.02 a 0.98, para cada par de transmisión. Este intervalo elimina los posibles errores de secuenciación en la llamada de variantes de baja frecuencia (<0.02 , según el umbral definido previamente) y las posiciones fijadas comunes entre el donante y el receptor a alta frecuencia (>0.98).

Utilizamos la versión aproximada y exacta del método beta-binomial, con los mismos parámetros:

- Bottleneck_size_estimation_ **aprox** was run in the two version with parameters “--plot_bool TRUE --var_calling_threshold 0.02 --Nb_min 1 --Nb_max 350”.
- Bottleneck_size_estimation_ **exact** was run in the two version with parameters “--plot_bool TRUE --var_calling_threshold 0.02 --Nb_min 1 --Nb_max 350”.

Comparación de resultados en otra población

Para comparar nuestros resultados en un contexto epidemiológico distinto, utilizamos un segundo conjunto de datos de secuenciación independiente proveniente de un estudio realizado en la ciudad de Vitória, Brasil ²⁶⁴. En este estudio, se observó una incidencia de tuberculosis de 38/100,000 habitantes, y la población presentaba una prevalencia relativamente baja de comorbilidades, así como un consumo moderado de tabaco y alcohol. En un estudio posterior, se identificaron 24 pares de transmisión ⁵⁶ (PRJNA475130) donde las muestras presentaban una cobertura promedio fue de 498X, y el número de diferencias genéticas entre los pares de transmisión varió de 0 a 13 SNPs con una media de 2,25 SNPs. La duración del intervalo serial (diagnóstico) osciló entre 0 y 63 meses, con una media de 28 meses y una mediana de 27 ⁵⁶.

Para llevar a cabo nuestros análisis, seleccionamos 16 pares de transmisión de los 24 presentes en el conjunto de datos que cumplían con todos nuestros criterios establecidos previamente. El análisis bioinformático de estos genomas se llevó a cabo utilizando la misma metodología descrita en este capítulo.

Resultados

Reconstrucción filogenética

La secuenciación de alta profundidad de las muestras proporcionó una cobertura media del 98% y una profundidad media de 590X (intervalo: 252-883X), con una mediana de 614X (intervalo: 264-876X) (ver Tabla 5.1). La reconstrucción del árbol filogenético, utilizando los 21 pares de transmisión donante-receptor validados, confirmó la relación existente entre los aislados de *M. tuberculosis*. Se identificaron diversos linajes y sublinajes, incluyendo el 2.2.10, 4.1.2, 4.1.2.1, 4.3.2, 4.3.3, 4.3.4.2, 4.6.1.1 y 4.10, siendo el linaje más común el 4.1.2.1 (8 pares). Además, se observaron variaciones de 0 a 2 SNP al calcular las distancias genéticas, en concordancia con los resultados obtenidos en el análisis epidemiológico previo (Figura 5.3).

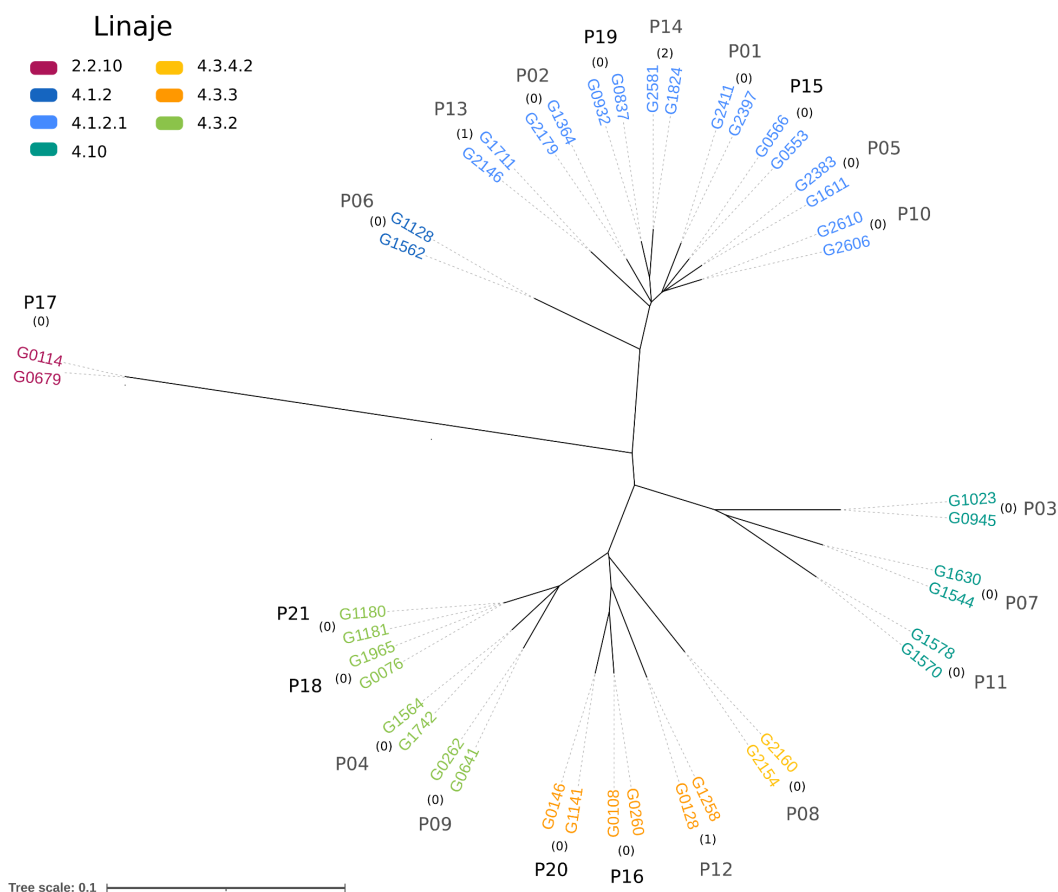


Figura 5.3. Relación filogenética entre los aislados de *M. tuberculosis* incluidos en el estudio.

En la filogenia, se incluyen los 21 pares de transmisión seleccionados, indicando entre paréntesis el número de diferencias de SNPs, basadas en variantes fijadas (frecuencia > 0.90), entre el donante y el receptor. Los pares de transmisión identificados previamente por nuestro grupo se muestran en negrita.

Diversidad genética en los pares de transmisión

Durante la exploración de la diversidad genética de los pares de transmisión, se observaron las variantes (SNPs) con frecuencias entre 0.02 y 0.98 (ver Figura 5.5 A). En promedio, se identificaron 137 SNPs totales en los 21 pares de transmisión, con un intervalo de 65 (par P08) a 504 (par P02) SNPs. En relación con la diversidad de los pares de transmisión, se encontró que en promedio existían 68 SNPs compartidos (37 a 96) entre el donante y el receptor. Además, se observaron promedios de 27 (5 a 113) SNPs únicos en el donante y 43 (4 a 406) SNPs únicos en el receptor (Tabla 5.1).

Además, considerando el conjunto de pares de transmisión, se encontró una correlación positiva, de grado medio, entre las frecuencias de los SNPs compartidos del donante y las del receptor, (correlación de Spearman de 0.77, valor $p < 2.2 \times 10^{-16}$) (Figura 5.5 A). También se calculó la correlación entre los SNPs para cada uno de los pares, encontrando una correlación de Spearman en el rango de 0.66 a 0.89 (grado medio a alto de correlación), con un valor $p < 0.0001$ en todos los casos (ver Figura suplementaria 1).

La gran mayoría de los SNPs se encontraron en frecuencias inferiores a 0.5, concentrándose principalmente en el intervalo de 0.02-0.05. Dentro de este margen, la mayoría de los SNPs son específicos ya sea del receptor o del donante, siendo más frecuentes los SNPs del receptor que los del donante. Sin embargo, a medida que la frecuencia aumenta, se observa un incremento en el número de SNPs compartidos en los pares de transmisión (ver Figura 5.5 B).

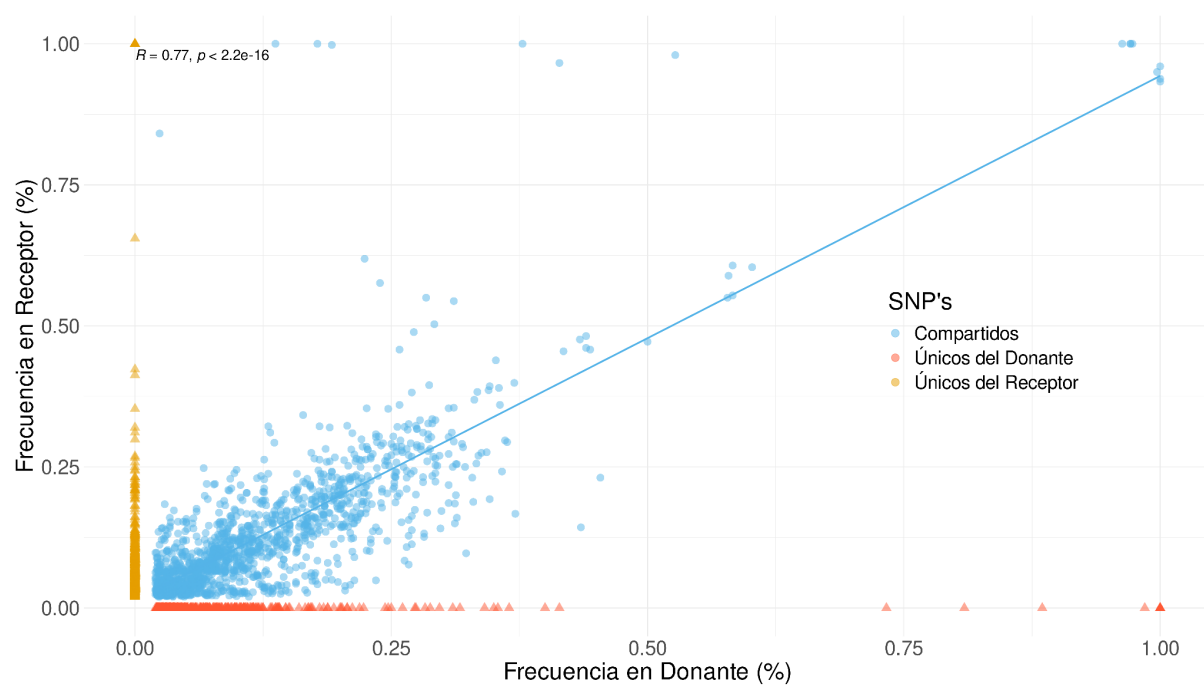
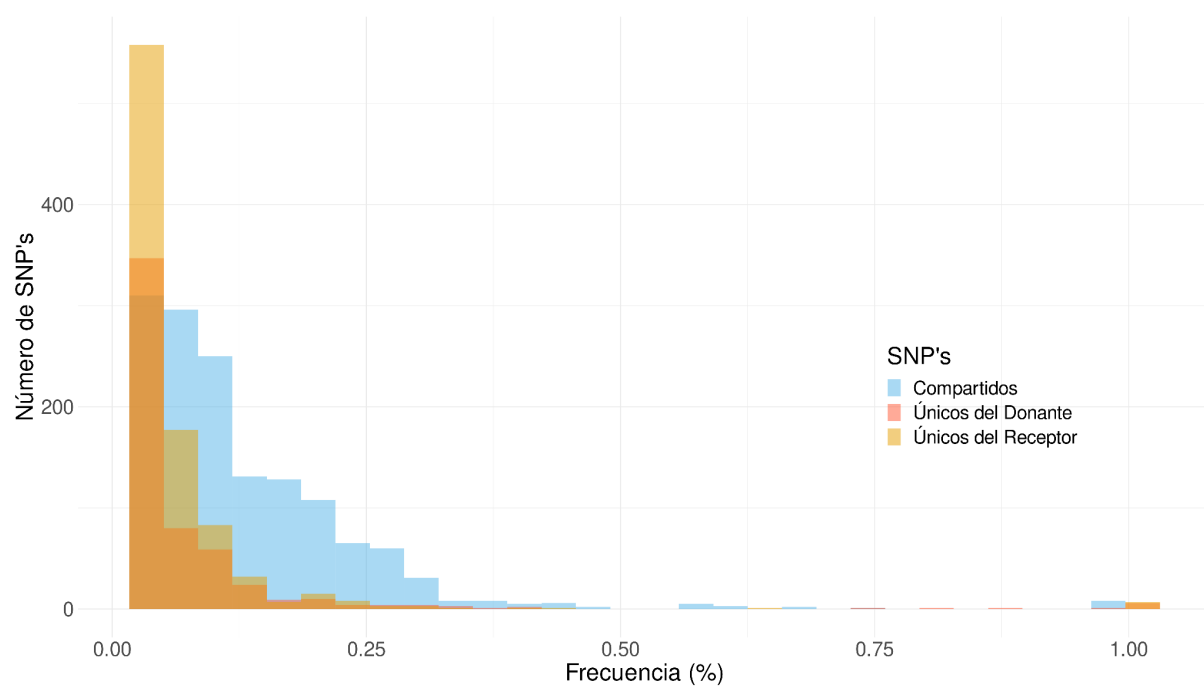
A**B**

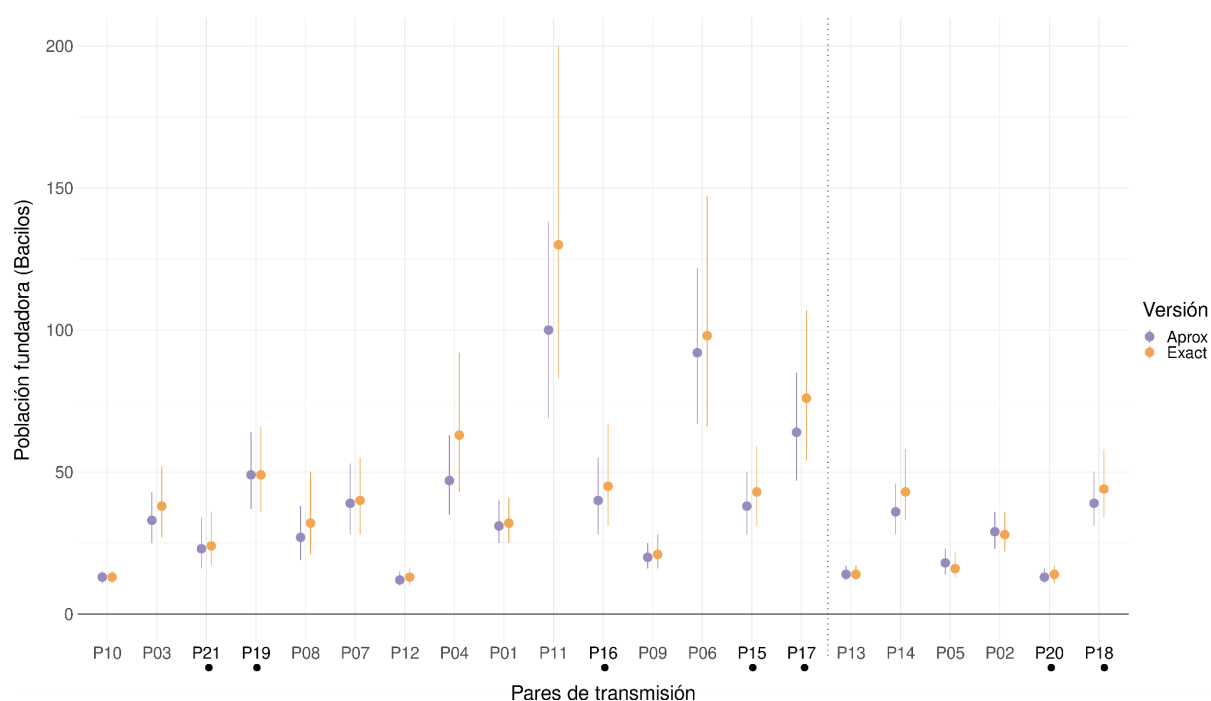
Figura 5.5. SNPs compartidos, únicos del donante y únicos del receptor en los 21 pares de transmisión (frecuencia de SNPs : 0.02-0.98).

Estimación del tamaño del cuello de botella de la transmisión

El número estimado de bacilos fundadores de la infección presentó un rango de 12 a 100 bacilos, con una media de 37 en la versión aproximada, y entre 13 a 130 bacilos, con una media de 42 en la versión exacta (ver Figura 5.6 A).

Utilizando la información epidemiológica disponible, (fechas de inicio de los síntomas, diagnóstico e inicio del tratamiento), se calculó el intervalo serial de los pares de transmisión y el periodo de TB activa de los receptores. El intervalo serial se definió como el período de tiempo que transcurre entre la fecha de inicio de síntomas del donante y la del receptor, correspondiente al período en que ocurrió la transmisión. El intervalo obtenido fue de 22 (en el par P21) a 832 (en el par P18) días, con una media de 261 días, es decir, aproximadamente 8.6 meses. En dos de los pares de transmisión (pares P03 y P10), no se pudo determinar el intervalo serial debido a la falta de información sobre la fecha de inicio de síntomas de los receptores. Además, se calculó el periodo de TB activa en los receptores, definido como la diferencia entre la fecha de inicio de síntomas y la fecha de diagnóstico. Los resultados indicaron un periodo de TB activa de 0 (P09 y P011) a 284 (P02) días, con una media de 53 días (ver Figura 5.6 - B).

A



B

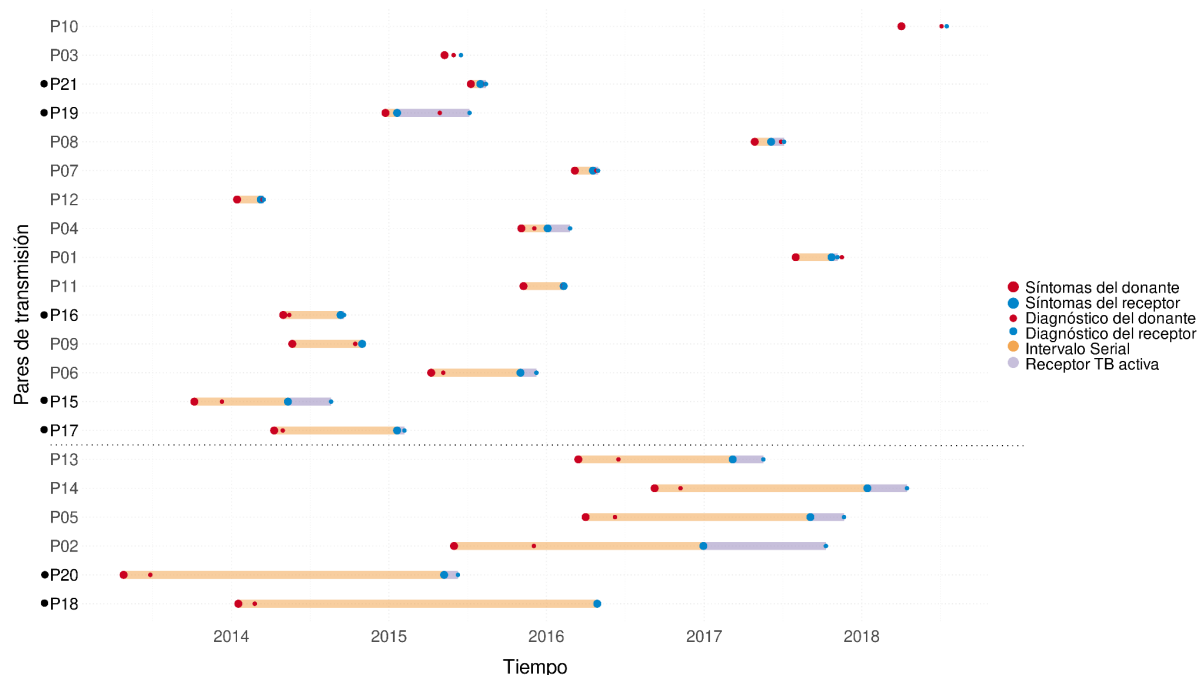


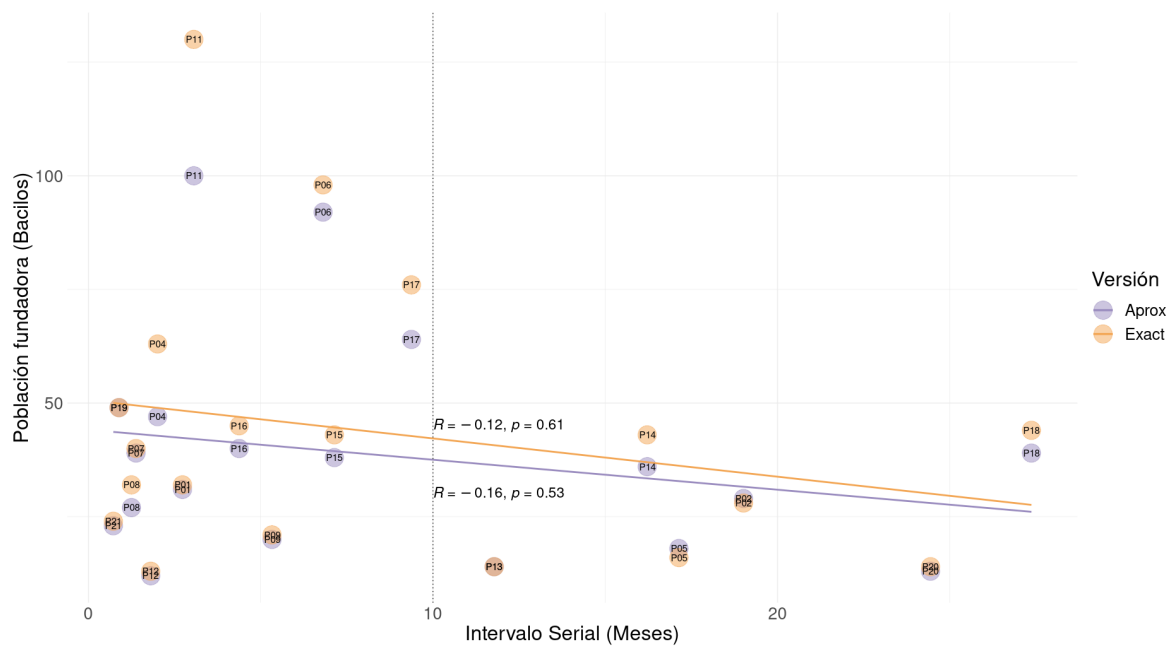
Figura 5.6. (A) Tamaños de los cuellos de botella (número de bacilos). Estimación con versiones aproximada (violeta) y exacta (naranja). (B) Se presenta la cronología de los pares de transmisión. Las fechas de inicio de los síntomas y de diagnóstico se representan mediante puntos grandes y pequeños, respectivamente (rojo=donante, azul=receptor). La línea de tiempo entre las fechas de inicio de los síntomas del donante y el receptor (intervalo serial) se representa mediante una línea naranja. La línea de tiempo entre el inicio de los síntomas y el diagnóstico (TB activa) se representa mediante una línea violeta. Los pares de transmisión identificados previamente por nuestro grupo se muestran marcados por puntos negros.

Por último, al graficar el intervalo serial (tiempo entre los síntomas del donante y del receptor) frente al tamaño del cuello de botella (número de bacilos), se observó una relación negativa débil no significativa entre ambas variables (correlación de Spearman -0.12, valor $p=0.61$).

Tras un análisis minucioso de la relación observada en el gráfico y teniendo en cuenta la variabilidad del intervalo serial de la tuberculosis en diferentes poblaciones, que oscila entre 6 y 24 meses²⁶⁵, establecimos un punto de corte de 10 meses para nuestro conjunto de datos (ver Figura 5.7 A). Mediante esta selección, buscamos evaluar la posible asociación entre el tamaño del cuello de botella y los intervalos seriales.

Los resultados revelaron una relación no significativa al comparar el tamaño del cuello de botella entre los grupos con intervalos seriales de ≤ 10 meses y >10 meses (evaluado mediante la prueba de suma de rangos de Wilcoxon con corrección de continuidad, $p=0.095$) (ver Figura 5.7 B).

A



B

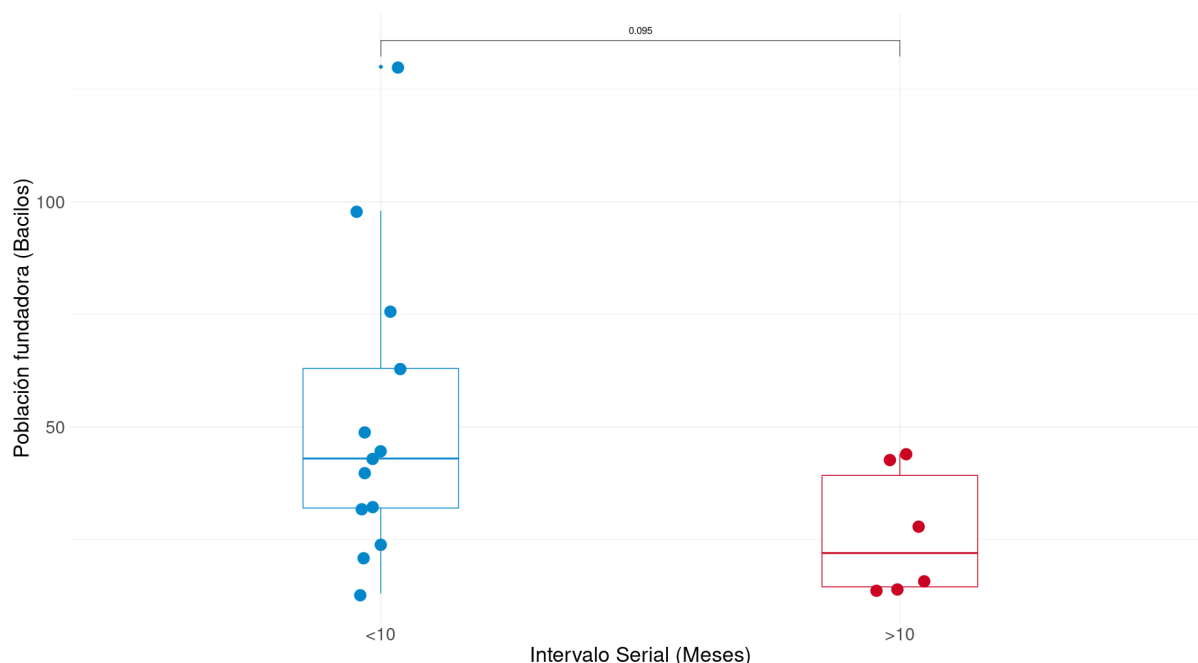


Figura 5.7. (A) Intervalo serial en meses frente al tamaño del cuello de botella (número de bacilos), entre los pares de transmisión. **(B)** Comparación del tamaño del cuello de botella entre pares de transmisión con intervalo serial menor (azul) y mayor (rojo) a 10 meses. Los datos corresponden a la estimación de tamaño de cuello de botella utilizando la versión exacta del método beta-binomial. (Prueba de suma de rangos de Wilcoxon con corrección de continuidad, valores $p=0.095$).

Comparación de resultados en otra población

Al realizar la estimación del tamaño del cuello de botella a partir de datos previamente publicados, se obtuvo un intervalo de 5 a 90 bacilos, con una media de 25 bacilos en la versión aproximada, y de 5 a 125 bacilos, con una media de 30 bacilos en la versión exacta (ver Figura 5.8). Estos resultados son similares a los obtenidos en nuestros 21 pares de transmisión de la Comunidad Valenciana, donde la media fue de 37 bacilos en la versión aproximada y 42 en la versión exacta.

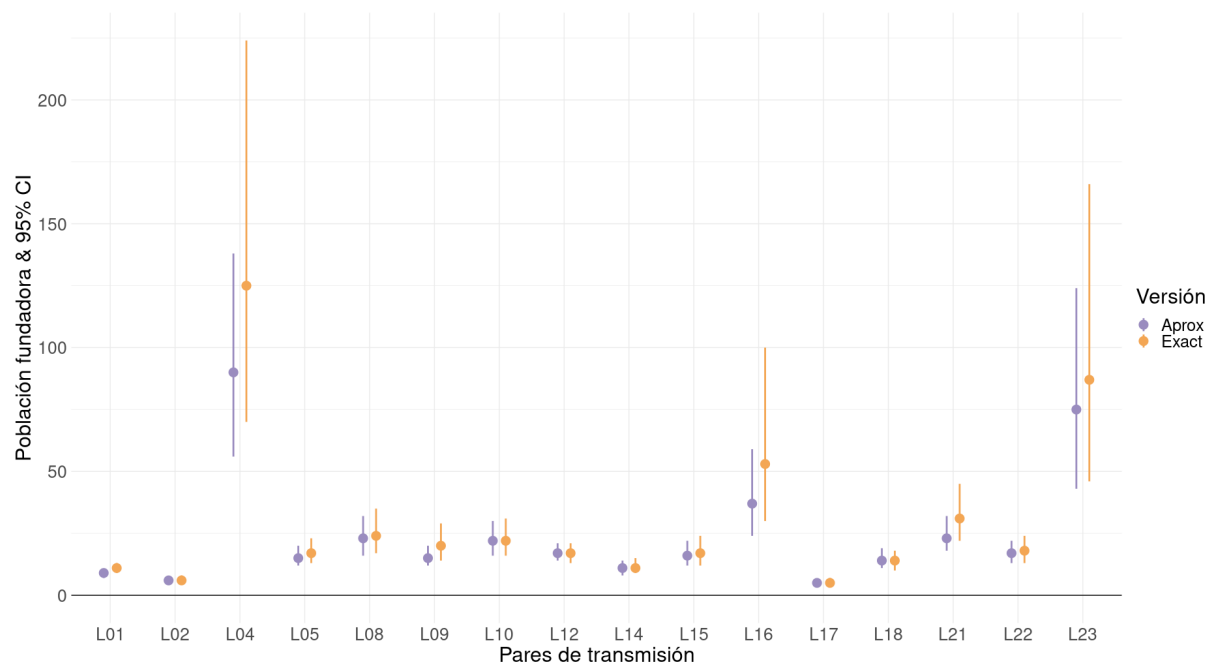


Figura 5.8. Tamaños de los cuellos de botella (número de bacilos). Estimación con versiones aproximada (violeta) y exacta (naranja).

Discusión

Este estudio introduce una metodología que fusiona datos genómicos y sociodemográficos para identificar y seleccionar los pares de transmisión. La determinación de estos pares donante-receptor ha sido un campo subexplorado en la investigación científica, principalmente debido a la complejidad de establecer la dirección de la transmisión en brotes de tuberculosis y la incertidumbre en la correcta identificación de los pares de transmisión directa.

Este desafío se vuelve aún más relevante en áreas con altas tasas de incidencia de la enfermedad, donde se ha comprobado que solo el 40% de los pares de transmisión en entornos domésticos pueden ser confirmados mediante la epidemiología molecular ⁵⁶. No obstante, en regiones de baja incidencia de tuberculosis, como en la Comunidad Valenciana, resulta más viable identificar cadenas de transmisión. Al combinar los resultados de la epidemiología genómica con los datos provenientes de los sistemas de vigilancia, se minimiza la probabilidad de error en la identificación de estas cadenas de transmisión.

En el análisis de nuestra colección, previamente discutido en el capítulo tres, identificamos un 49% de casos en transmisión. Sin embargo, para estimar el cuello de botella, solo logramos seleccionar 21 pares de transmisión donante-receptor, con alta probabilidad de contagio directo. Estos resultados subrayan la complejidad que implica la identificación de la transmisión de persona a persona, ya sea a través de enfoques epidemiológicos y/o genómicos. Si bien, la secuenciación completa del genoma ha demostrado su eficacia para estimar la transmisión en ciertos contextos y se han desarrollado metodologías específicas para ello ²⁶⁶, su aplicación a nivel poblacional para comprender plenamente la dinámica de la transmisión sigue siendo limitada ¹⁸³ casos de transmisión asociados a tuberculosis latente ²⁶⁷. El desarrollo de nuevas tecnologías, como la reconstrucción de árboles de transmisión ¹⁵⁴, permitirá obtener una visión más completa y dinámica de la transmisión de diversos patógenos.

Nuestro trabajo también resalta la importancia de los datos de secuenciación profunda. En nuestro estudio, hemos analizado SNPs de baja frecuencia, ya que el método de muestreo beta-binomial que utilizamos depende de la detección de estas variantes (SNPs) para obtener

estimaciones precisas. Aunque se han estudiado las variantes de baja frecuencia en la farmacoresistencia de la tuberculosis humana ²⁶⁸, su posible implicación en la evolución de la bacteria dentro de una cadena de transmisión o de un individuo ha sido poco explorada hasta ahora ³⁷.

En cuanto a la selección y filtrado de SNPs, empleamos una metodología rigurosa que nos proporcionó una lista confiable de variantes de baja frecuencia. Además, aplicamos un criterio de selección adicional a posiciones compartidas que aparecen con frecuencia similar en dos o más aislados, asumiendo que podrían ser errores de secuenciación u homoplasias. Aunque esta selección no afecta significativamente nuestros resultados y la consideramos una metodología adecuada, presenta la limitación potencial de reducir la diversidad real de las muestras.

En relación con la distribución de los SNPs en los 21 pares de transmisión en un intervalo de frecuencias del 0.02 al 0.98, se observó una clara falta de SNPs con una frecuencia superior al 50%. Este patrón se apreció tanto en los SNPs compartidos como en los SNPs únicos a nivel global (ver Figura 5.5) y en cada par de transmisión (ver Figura 5.8). Estos resultados concuerdan con la baja tasa de sustitución reportada para *M. tuberculosis*.

En contraparte, nuestros resultados indican una gran cantidad de SNPs de baja frecuencia, especialmente en el intervalo de 0.02-0.05. Este aumento de la diversidad a baja frecuencia en SNPs exclusivos del donante o del receptor sugiere una adaptación de la bacteria al hospedador. Los SNPs de baja frecuencia son difíciles de estudiar debido a su escasa fiabilidad y a limitaciones técnicas, pero son clave para calcular con precisión parámetros importantes como las mutaciones de interés y/o las tasas de mutación y sustitución ^{31,32,269}. Por lo tanto, consideramos que el estudio en profundidad de la diversidad a baja frecuencia será esencial en un futuro próximo para comprender los mecanismos evolutivos específicos de *M. tuberculosis*, aunque las metodologías para su estudio aún pueden estar en desarrollo.

Respecto a la dosis inicial de bacilos. Según los experimentos llevados a cabo por Findel, se ha observado que una dosis mínima en bioaerosol de 20-40 bacilos es suficiente para producir la infección en cobayas ⁶². Esta cifra se encuentra en concordancia con el intervalo de bacilos fundadores que hemos estimado: 42 (13-130), mediante la versión exacta del método beta-binomial.

La estimación mencionada concuerda con los hallazgos del modelo macaco de tuberculosis²⁷⁰, en el cual se observa la propagación de la infección in vivo mediante el uso de PET (tomografía de emisión de positrones) y etiquetas genómicas distintivas y únicas para cada bacilo de *M. tuberculosis*. Este modelo utiliza un inóculo inferior a 20 unidades formadoras de colonias (UFC). También coincide con la infección a bajas dosis con bioaerosol (1 UFC) de *M. bovis* en ganado bovino²⁷¹. Es importante destacar que esta cifra también se encuentra dentro del intervalo propuesto por Rich²⁴⁹ quien sugiere la presencia de 1-300 bacilos en una gota esférica de 0,25 mm de diámetro, basándose en los recuentos realizados por Ziesché²⁷². Estos descubrimientos apuntan a que incluso una sola microgota de bioaerosol puede ser suficiente para iniciar una infección de tuberculosis en seres humanos. Esto podría explicar la capacidad infecciosa de un individuo en eventos de superdispersión²⁷³.

Nuestras estimaciones sustentan la idea de que las infecciones naturales de tuberculosis en humanos rara vez se originan a partir de un solo bacilo. Por un lado, el valor mínimo de tamaño de cuello de botella estimado es 13, y por otro, la diversidad observada en las muestras biológicas sugiere que la infección es producida por una población heterogénea de bacterias. Aunque se ha demostrado la presencia de partículas aerotransportadas que contienen 1 o 2 bacilos, nuestros análisis no han identificado una población fundadora compuesta por un solo bacilo en ninguno de los 21 pares de transmisión estudiados. Sin embargo, nuestras estimaciones sugieren que una dosis baja de 13 bacilos podría ser suficiente para iniciar la infección.

Un argumento adicional que respalda nuestras estimaciones, es que una infección lenta y progresiva, como ocurre en la mayoría de los individuos, puede explicarse mediante una infección con una baja dosis de bacilos. Esto se ha demostrado en modelos animales susceptibles, donde la infección con un gran número de bacilos causa lesiones generalizadas y la muerte de los individuos en pocas semanas, mientras que la infección con un número reducido de bacilos progresa lentamente²⁴⁹.

Al considerar las limitaciones y aspectos relevantes en la estimación del número medio de bacilos utilizando el método de muestreo beta-binomial, es importante tener en cuenta los siguientes factores: el método estima el número de bacilos de la población fundadora en una infección activa, es decir, aquellos bacilos viables que fueron transportados a los bronquiolos

terminales de un individuo receptor y lograron reproducirse. Por lo tanto, mediante este método no es posible estimar el número de bacilos no viables o aquellos que, siendo parte de la población transmitida, no lograron establecer la infección. Sin embargo, podemos suponer la presencia de bacilos no viables que podrían acompañar a la población fundadora.

Nuestras estimaciones sugieren que las infecciones por *M. tuberculosis* pueden establecerse a partir de un número limitado de bacilos, planteando que una pequeña proporción de bacilos exhalados por un individuo infectado sería suficiente para infectar a uno o más individuos sanos. Sin embargo, y dado que nuestro análisis no permite estimar el número total de bacilos exhalados por el donante, ya que este valor es probable que varíe en función de diversos factores; no es posible definir cuál es la carga bacilar mínima del donante, necesaria para que se produzca una infección.

En cuanto a la relación entre el intervalo serial y el tamaño del cuello de botella, se ha observado una relación inversa entre ambas variables. Esta relación, aunque no significativa (-0.12 , $p = 0.61$, ver Figura 5.7 A), se mantiene al dividir los datos en dos grupos: aquellos con intervalos seriales ≤ 10 meses y aquellos con intervalos seriales > 10 meses, según el test de suma de rangos de Wilcoxon con corrección de continuidad ($p = 0.095$, ver Figura 5.7B). Sin embargo, es relevante tener en cuenta que la falta de robustez estadística en nuestros resultados puede atribuirse a la limitación en el número de muestras y a los intervalos de tiempo acotados de nuestro estudio.

En este contexto, es crucial destacar que el método de muestreo beta-binomial ha sido ampliamente utilizado en el estudio de infecciones virales agudas transmitidas por el aire, como la gripe y el SARS-CoV-2. Sin embargo, su aplicabilidad en infecciones crónicas puede plantear algunas limitaciones²³⁹. En el caso de infecciones crónicas, el muestreo del huésped receptor puede ocurrir meses o años después del inicio de la infección, lo que podría resultar en diferencias significativas entre la población bacteriana en el momento del muestreo y la población bacteriana fundadora. Esto podría explicar la tendencia que observamos de valores más bajos en la población fundadora de bacilos a medida que avanza el tiempo. Sin embargo, las estimaciones del número de bacilos (menos de 130) para iniciar una infección parecen sólidas en el conjunto global de datos, y esto se corrobora en un segundo conjunto de datos.

En relación con las limitaciones del segundo conjunto de datos, es importante destacar dos aspectos fundamentales. En primer lugar, no se proporcionan las fechas de inicio de síntomas, lo que impide la determinación de un intervalo serial equivalente al que hemos empleado en nuestra investigación. En este conjunto de datos, el intervalo serial se calcula entre las fechas de diagnóstico, expresado en meses, sin registros precisos de fechas exactas. Esta variación en la metodología de cálculo dificulta la comparación de resultados con nuestra investigación. Además, es relevante señalar que el enfoque del segundo conjunto de datos se centra principalmente en la investigación de largos periodos de latencia de la tuberculosis. Por lo tanto, la media de los intervalos seriales de 28 meses difiere significativamente de nuestros datos, que es de tan solo 8.5 meses.

Por último, hay que considerar que en el contexto de la transmisión de la tuberculosis humana, el número de bacilos no es el único factor determinante. Otros elementos como la resistencia innata y adquirida, la hipersensibilidad y la virulencia de los bacilos también tienen un papel significativo en este proceso. Por consiguiente, no se puede predecir que un número específico de bacilos siempre tenga los mismos efectos. Sin embargo, no podemos subestimar su importancia en la transmisión y desarrollo de la enfermedad. La intención de resaltar estos aspectos es demostrar que los cuellos de botella genéticos son mucho más complejos de lo que se presenta en la literatura, donde a menudo se utilizan como una explicación rápida para la baja diversidad ²⁷. Por ende, se necesita investigar más a fondo este fenómeno en patógenos bacterianos.

En resumen, en este estudio revisamos la afirmación de que un solo bacilo puede iniciar una infección tuberculosa natural. Aunque esto puede ser cierto en casos de infecciones inducidas, en nuestra cohorte cuidadosamente seleccionada de casos de donante-receptor las dosis infecciosas oscilan entre 13 y 130 bacilos, con una media de 42. En este sentido, si bien una población pequeña puede causar una infección de manera natural, parece muy poco probable que ésta se produzca por un único bacilo. Estudios futuros deberán trabajar con una cohorte más apropiada que incluya las diversas variables observadas en los entornos de la transmisión de la TB.

6. Discusión General

En esta discusión global, nos enfocaremos en la aplicación de la secuenciación completa del genoma en la epidemiología y la comprensión de la transmisión de la tuberculosis. Cada capítulo previo ha explorado detenidamente aspectos específicos, y ahora examinaremos temas clave relacionados con esta tecnología en el contexto de la TB. También discutiremos las limitaciones identificadas durante el estudio y destacaremos las perspectivas futuras en la investigación de la tuberculosis.

Uso de la WGS en el estudio de la tuberculosis

La secuenciación completa del genoma ha emergido como un recurso esencial en la investigación de la tuberculosis. Particularmente, en la Comunidad Valenciana, esta tecnología nos ha permitido profundizar en el estudio de la transmisión de la enfermedad y evaluar la diversidad genética entre las cepas de *Mycobacterium tuberculosis*.

La integración de datos provenientes de esta herramienta con información sociodemográfica ha enriquecido nuestro entendimiento acerca de las variables y factores de riesgo relacionados con la propagación de la enfermedad en nuestra comunidad. Además, hemos empleado la WGS para identificar resistencia a medicamentos. Los datos recopilados en este estudio nos brindarán la base necesaria para implementar estrategias eficaces en el control y la prevención de la tuberculosis, así como para prevenir el surgimiento de cepas multidrogoresistentes.

No obstante, a pesar de que la secuenciación completa del genoma (WGS) está ganando importancia en la investigación y control de la tuberculosis, es esencial reconocer las limitaciones inherentes a esta técnica. En este contexto, se destacan los siguientes desafíos y restricciones identificados en nuestro estudio:

1. **Complejidad en el análisis de datos:** El análisis de datos de WGS en la tuberculosis es intrincado y demanda habilidades y recursos especializados. En nuestro trabajo, hemos propuesto metodologías para optimizar la utilización de estos datos en la investigación epidemiológica y de transmisión de la TB.

2. **Limitaciones en el análisis temporal y geográfico:** A pesar de que la WGS proporciona información valiosa sobre el genoma de *Mycobacterium tuberculosis*, la integración de datos temporales y geográficos puede presentar desafíos. La secuenciación no siempre permite distinguir entre infecciones que ocurrieron en diferentes momentos o lugares sin datos adicionales.
3. **Limitaciones en la interpretación de los resultados:** La interpretación de los resultados de análisis genómicos de tuberculosis puede ser complicada debido a la complejidad genética de *M. tuberculosis*, especialmente en variantes genéticas poco comunes. Además, existe una falta de información sobre la función y la interacción de muchos genes y variantes genéticas.

Además de las limitaciones previamente mencionadas, es importante destacar una limitación adicional de gran relevancia en este estudio: la falta de datos completos sobre variables epidemiológicas en nuestra base de datos. Esta carencia de información integral en los aislados ha restringido nuestra capacidad para obtener una visión precisa de la epidemiología de la tuberculosis en la comunidad. Mejorar la recopilación de datos de este tipo sería un avance sustancial y permitiría futuras investigaciones con una base más sólida y una comprensión más profunda de la situación.

Al reflexionar sobre esta limitación y considerar la abundancia de datos presentados en este capítulo y en estudios similares a nivel global, especialmente dado el impacto significativo de la tuberculosis en la salud pública, se plantea una perspectiva intrigante. Sería pertinente enfocar esfuerzos en el desarrollo de plataformas digitales interconectadas con los sistemas de vigilancia. Estas plataformas podrían automatizar la organización de datos y facilitar la comunicación entre centros de salud, sistemas de vigilancia y laboratorios especializados en genómica. Este avance permitiría realizar análisis epidemiológicos de manera ágil, ofreciendo una visión más rápida y precisa de la situación. La implementación de esta iniciativa podría ser esencial para mejorar la rapidez y eficacia en la respuesta ante desafíos de salud pública, incluyendo eventos como la reciente pandemia de COVID-19. En nuestro grupo de investigación, hemos dado pasos en esta dirección al organizar las secuencias genómicas y los metadatos de nuestros aislados en una única base de datos desarrollada como sistema de gestión, lo que facilita el acceso a la información para nuestro equipo.

En síntesis, la epidemiología molecular, en particular el análisis de datos de secuenciación de genomas completos (WGS), ha demostrado ser una herramienta invaluable en el estudio de la tuberculosis en la Comunidad Valenciana. Facilita el rastreo de la propagación de la enfermedad, la detección de brotes y la vigilancia de la resistencia a los medicamentos. A pesar de ciertas restricciones, el WGS continuará desempeñando un papel fundamental en el control y la prevención de la tuberculosis. Se impone la necesidad de priorizar la investigación y desarrollo de estrategias y metodologías que permitan superar las limitaciones actuales, maximizando así el potencial de esta herramienta en el futuro.

Limitaciones y utilidades del catálogo de resistencias a fármacos de la OMS para tuberculosis

El catálogo de resistencias a fármacos de la OMS destinado a la tuberculosis, es un recurso invaluable para abordar la TB resistente a medicamentos. Este compendio detalla tanto los medicamentos de primera como de segunda línea empleados en el tratamiento de la TB, así como las mutaciones genéticas asociadas, y no asociadas, con la resistencia a estos fármacos.

Dicho catálogo suministra información de gran utilidad tanto para profesionales clínicos como para programas de salud pública, al facilitar la selección de tratamientos más eficaces y adecuados para pacientes con TB resistente a medicamentos. Asimismo, desempeña un papel fundamental en la vigilancia global de la resistencia a medicamentos en la TB. En esta investigación, hemos evaluado los niveles uno y dos de este catálogo con resultados positivos.

Sin embargo, es imperativo reconocer que este catálogo no abarca todas las posibles mutaciones genéticas vinculadas con la resistencia a fármacos en diversas poblaciones; asimismo muchas de las mutaciones incluidas figuran como “inciertas”, es decir no se dispone de evidencia suficiente para determinar si está asociada, o no, con la resistencia a un antibiótico. Este ejemplo ilustra la mutación 1674481 Ser94Ala del gen *inhA*. En nuestra investigación, se detectó un significativo número de aislados que portan esta mutación, los cuales demostraron resistencia fenotípica a isoniazida. Además, en nuestro grupo, confirmamos la relevancia de esta mutación mediante análisis de las Concentraciones Mínimas Inhibitorias (CMI) en estos aislados ²⁷⁴.

Es importante señalar que este catálogo exhibe una alta sensibilidad y especificidad para antibióticos de primera línea, como isoniazida y rifampicina. Sin embargo, presenta limitaciones significativas para las drogas de los grupos A-C, utilizadas en el tratamiento de tuberculosis resistente a múltiples fármacos (RR/MDR/XDR-TB), principalmente debido a la escasez de datos fenotípicos.

En resumen, el catálogo de resistencias a fármacos de la OMS para la TB es un instrumento crucial para tratar y controlar la TB resistente a medicamentos, pero con ciertas limitaciones. Es esencial que esta herramienta continúe su desarrollo y actualización para reflejar los avances más recientes en el ámbito de la resistencia a medicamentos, con el fin de optimizar la eficacia de los programas de control de la TB. Además, es fundamental emplear esta información en conjunto con otras herramientas y enfoques clínicos para determinar el tratamiento más adecuado para cada paciente.

Uso WGS para la identificación de transmisión en tuberculosis

La aplicación de la secuenciación de genoma completo (WGS) en la identificación de la transmisión de tuberculosis se ha revelado como una herramienta esencial en la gestión y prevención de esta enfermedad. Este enfoque permite identificar patrones genéticos compartidos entre las bacterias causantes de la tuberculosis, posibilitando el rastreo de posibles cadenas de transmisión.

Una de las principales ventajas de emplear WGS para detectar clústeres de transmisión de tuberculosis radica en su alta resolución, dado que interroga el 90% del genoma de la bacteria, esta técnica es capaz de tipificar la bacteria a nivel de sublinaje e identificar entre *links* epidemiológicos recientes o más antiguos, según las diferencias genéticas entre los aislados.

Sin embargo, se presentan desafíos significativos al implementar WGS para la identificación de transmisión en tuberculosis. Uno de ellos es la necesidad de contar con infraestructura y recursos técnicos adecuados para llevar a cabo la secuenciación y el análisis de datos, además de crear bases de datos que combinen información genómica y sociodemográfica. Por otro lado, para estimar la transmisión de manera más precisa, será necesario superar las

limitaciones de cultivo, para ampliar el número de casos secuenciados. La identificación de cadenas de transmisión en tiempo real y su seguimiento por parte de los sistemas de salud pública, es también imprescindible.

Otro desafío radica en la interpretación de los resultados en el contexto epidemiológico. A pesar de que WGS puede revelar patrones genéticos compartidos entre bacterias, esto no siempre implica una cadena de transmisión activa. Existe la posibilidad de que las bacterias se hayan adquirido de una fuente común o provengan de un individuo no diagnosticado.

En resumen, el empleo de la secuenciación de genoma completo para la identificación de clústeres de transmisión en tuberculosis es una perspectiva prometedora en el control y prevención de esta enfermedad. No obstante, su aplicación con éxito demanda una infraestructura adecuada, recursos técnicos y experiencia en genómica y epidemiología para una interpretación efectiva de los resultados.

Identificación y monitoreo de genotipos endémicos de tuberculosis en la Comunidad Valenciana.

La adaptación de los genotipos endémicos de tuberculosis (TB) a la población humana puede estar influenciada por diversos factores, como la genética humana, la geografía, la demografía, los patrones de migración y la exposición a la enfermedad. En algunas regiones, *M. tuberculosis* ha evolucionado para ser más virulento y resistente a los tratamientos ²⁷⁵, lo que se traduce en mayores tasas de mortalidad y desafíos en el control de la enfermedad.

Investigar la coevolución entre genotipos endémicos y la población local puede proporcionar conocimientos fundamentales sobre los mecanismos moleculares y celulares involucrados en la infección por tuberculosis, con implicancias importantes en el diseño eficiente de estrategias de control de la enfermedad.

En la Comunidad Valenciana hemos identificado posibles genotipos endémicos. Estudios futuros podrían revelar adaptaciones específicas de la población humana a estos genotipos, similar a lo observado con patógenos como la malaria y el VIH ^{276,277}. La exposición continua a ciertos genotipos de tuberculosis puede conducir a la selección natural de variantes genéticas que confieren mayor resistencia o susceptibilidad a la enfermedad.

Identificación de pares de transmisión donante-receptor: desafíos y recomendaciones

En este estudio, hemos desarrollado una metodología que combina datos genómicos y sociodemográficos para identificar pares de transmisión donante-receptor en estudios de base poblacional de tuberculosis (TB). Sin embargo, es importante resaltar la complejidad de determinar la dirección de la transmisión y la correcta identificación de estos pares donante-receptor.

Nuestro estudio ha revelado desafíos similares a los encontrados en otras investigaciones que emplean la WGS para analizar la transmisión ²⁷⁸, destacando la importancia de los datos de secuenciación profunda. Hemos identificado variantes genéticas de baja frecuencia que son cruciales para obtener una estimación precisa de la diversidad de *M. tuberculosis* en el huésped humano. Cabe mencionar que estudiar estas variantes de baja frecuencia plantea desafíos debido a su escasa fiabilidad y a limitaciones técnicas, pero son esenciales para comprender los mecanismos evolutivos específicos de esta bacteria en el contexto humano.

En resumen, para identificar de manera efectiva los pares de transmisión donante-receptor en la tuberculosis humana, es imperativo considerar conjuntamente datos clínicos, epidemiológicos y moleculares. Dependiendo exclusivamente de uno de estos aspectos, existe el riesgo de obtener una representación incompleta de este fenómeno complejo.

¿Cuántos bacilos son necesarios para iniciar una infección natural de tuberculosis? Discusión de las estimaciones y su importancia para el control de la enfermedad.

La discusión sobre la dosis inicial de bacilos necesaria para iniciar una infección natural de *M. tuberculosis* plantea cuestiones cruciales. Nuestras estimaciones sugieren que la probabilidad de que un solo bacilo inicie una infección natural es muy baja, y es probable que se necesiten aproximadamente 42 bacilos fundadores, en línea con investigaciones previas. La confirmación de que una carga bacilar tan baja puede generar una infección plantea numerosas interrogantes que han sido previamente exploradas y que, con el uso de tecnologías avanzadas, pueden arrojar resultados diferentes o validarse.

Por ejemplo, se ha sugerido que *M. tuberculosis* es un patógeno altamente especializado capaz de infectar a los humanos con una carga bacilar extremadamente baja, y nuestras observaciones respaldan esta idea. Sin embargo, también es plausible argumentar que el sistema inmunológico humano es altamente eficiente en la eliminación de cargas bacilares de *M. tuberculosis*, lo que podría explicar el reducido número de bacilos fundadores. Independientemente de nuestras estimaciones sobre el número de bacilos capaces de causar una nueva infección, lo cual es de gran relevancia, es esencial cambiar nuestra perspectiva al abordar el problema de la transmisión de la tuberculosis humana.

Debemos considerarla como una dinámica compleja de variables interconectadas en lugar de un simple fenómeno de causa y efecto. Con el avance de tecnologías emergentes, tenemos la capacidad de desarrollar algoritmos que puedan tener en cuenta el gran número de variables involucradas en la transmisión de esta enfermedad ^{279,280}. Al comprender mejor esta dinámica, podríamos incluso anticipar su comportamiento y diseñar estrategias más efectivas para prevenir y controlar su propagación.

7. Conclusiones

- La transmisión de la TB en la Comunidad Valenciana está asociada a la edad, es mayor entre individuo de sexo masculino, con tinción de esputo positiva y cuando la localización de la enfermedad es pulmonar. Notablemente, existe una asociación con la condición de haber nacido en España; es decir en nuestro entorno la mayor transmisión ocurre entre población local.
- Los factores de riesgo identificados para la transmisión de la TB en esta región son el alcoholismo y la privación de libertad.
- La WGS supera en eficacia a los estudios de contactos tradicionales en la detección de transmisión en la Comunidad Valenciana, aunque ambas aproximaciones resultan complementarias.
- Los análisis de sensibilidad y especificidad de la WGS para la detección de resistencias utilizando el catálogo de la OMS, considerando sus dos niveles principales, respaldan la fiabilidad de esta herramienta para drogas como isoniazida y rifampicina, en la Comunidad Valenciana.
- La información adicional obtenida para la mutación 1674481 Ser94Ala del gen *inhA*, sugiere que la misma debería ser reclasificada como “asociada a resistencia” en la nueva versión del catálogo de la OMS.
- La identificación de genotipos endémicos implica el desarrollo de nuevas estrategias y metodologías, en esta tesis hemos propuesto una de ellas identificando probables genotipos endémicos de *M. tuberculosis* en la Comunidad Valenciana. Estos genotipos estarían asociados a transmisión de la población local, probablemente una transmisión sostenida en el tiempo, siendo potencialmente responsables del mantenimiento de la incidencia de TB en nuestro entorno.
- La identificación y selección correcta de pares de transmisión donante-receptor en TB humana es posible mediante la complementación de datos epidemiológicos clásicos y genómicos.
- Nuestros resultados muestran la importancia del empleo de la secuenciación profunda para la detección de SNPs a baja frecuencia para los estudios de la diversidad genética en *M. tuberculosis*.

- Nuestros resultados sugieren que es poco probable que las infecciones naturales de tuberculosis sean causadas por un único bacilo de *M. tuberculosis*. En cambio, nuestra estimación sugiere que pocos bacilos serían suficientes para iniciar una infección natural activa de TB en humanos.

8. Bibliografía

1. CDCTB. Tuberculosis. *Centers for Disease Control and Prevention*
<https://www.cdc.gov/tb/default.htm> (2022).
2. Moule, M. G. & Cirillo, J. D. Dissemination Plays a Critical Role in Pathogenesis.
Front. Cell. Infect. Microbiol. **10**, 65 (2020).
3. Frascella, B. *et al.* Subclinical Tuberculosis Disease-A Review and Analysis of
Prevalence Surveys to Inform Definitions, Burden, Associations, and Screening
Methodology. *Clin. Infect. Dis.* **73**, e830–e841 (2021).
4. Kendall, E. A., Shrestha, S. & Dowdy, D. W. The Epidemiological Importance of
Subclinical Tuberculosis. A Critical Reappraisal. *Am. J. Respir. Crit. Care Med.* (2021)
doi:10.1164/rccm.202006-2394PP.
5. CDCTB. Tuberculosis (TB) - Questions and Answers About TB. *Centers for Disease
Control and Prevention* <https://www.cdc.gov/tb/publications/faqs/default.htm> (2022).
6. Zong, Z. *et al.* Relapse Versus Reinfection of Recurrent Tuberculosis Patients in a
National Tuberculosis Specialized Hospital in Beijing, China. *Front. Microbiol.* **0**,
(2018).
7. Lekko, Y. M. *et al.* complex in wildlife: Review of current applications of antemortem
and postmortem diagnosis. *Vet World* **13**, 1822–1836 (2020).
8. Estrategias innovadoras para el diagnóstico y seguimiento de los pacientes tuberculosos.
Archivos de Bronconeumología **43**, 225–232 (2007).
9. Saxena, S., Spaink, H. P. & Forn-Cuní, G. Drug Resistance in Nontuberculous
Mycobacteria: Mechanisms and Models. *Biology* **10**, (2021).
10. Tortoli, E. Microbiological features and clinical relevance of new species of the genus
Mycobacterium. *Clin. Microbiol. Rev.* **27**, (2014).

11. Simner, P. J., Woods, G. L. & Wengenack, N. L. Mycobacteria. *Microbiol Spectr* **4**, (2016).
12. Boshoff, H. I. M. & Barry, C. E., 3rd. Tuberculosis - metabolism and respiration in the absence of growth. *Nat. Rev. Microbiol.* **3**, 70–80 (2005).
13. Forrellad, M. A. *et al.* Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence* **4**, (2013).
14. Converse, S. E. *et al.* MmpL8 is required for sulfolipid-1 biosynthesis and *Mycobacterium tuberculosis* virulence. *Proc. Natl. Acad. Sci. U. S. A.* **100**, (2003).
15. Cole, S. T. *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544 (1998).
16. Coscolla, M. *et al.* Phylogenomics of reveals a new lineage and a complex evolutionary history. *Microb Genom* **7**, (2021).
17. Firdessa, R. *et al.* Mycobacterial lineages causing pulmonary and extrapulmonary tuberculosis, Ethiopia. *Emerg. Infect. Dis.* **19**, 460–463 (2013).
18. Ngabonziza, J. C. S. *et al.* A sister lineage of the *Mycobacterium tuberculosis* complex discovered in the African Great Lakes region. *Nat. Commun.* **11**, 2917 (2020).
19. Coscolla, M. *et al.* Phylogenomics of *Mycobacterium africanum* reveals a new lineage and a complex evolutionary history. *bioRxiv* (2020) doi:10.1101/2020.06.10.141788.
20. Brosch, R. *et al.* A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 3684–3689 (2002).
21. Smith, N. H., Hewinson, R. G., Kremer, K., Brosch, R. & Gordon, S. V. Myths and misconceptions: the origin and evolution of *Mycobacterium tuberculosis*. *Nat. Rev. Microbiol.* **7**, 537–544 (2009).
22. Galagan, J. E. Genomic insights into tuberculosis. *Nat. Rev. Genet.* **15**, 307–320 (2014).
23. Coscolla, M. & Gagneux, S. Consequences of genomic diversity in *Mycobacterium*

- tuberculosis*. *Semin. Immunol.* **26**, 431–444 (2014).
24. Asare, P. *et al.* Reduced transmission of *Mycobacterium africanum* compared to *Mycobacterium tuberculosis* in urban West Africa. *Int. J. Infect. Dis.* **73**, 30–42 (2018).
 25. Gagneux, S. Host-pathogen coevolution in human tuberculosis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 850–859 (2012).
 26. Zwyrer, M. *et al.* A new nomenclature for the livestock-associated *Mycobacterium tuberculosis* complex based on phylogenomics. *Open Res. Eur.* **1**, 100 (2021).
 27. Gagneux, S. & Stritt, C. How do monomorphic bacteria evolve? The **Mycobacterium tuberculosis** complex and the awkward population genetics of extreme clonality. (2022).
 28. Gagneux, S. Ecology and evolution of *Mycobacterium tuberculosis*. *Nat. Rev. Microbiol.* **16**, 202–213 (2018).
 29. Coll, F. *et al.* A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat. Commun.* **5**, 4812 (2014).
 30. Comas, I. *et al.* Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat. Genet.* **45**, 1176–1182 (2013).
 31. Ford, C. B. *et al.* *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat. Genet.* **45**, 784–790 (2013).
 32. Ford, C. B. *et al.* Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. *Nat. Genet.* **43**, 482–486 (2011).
 33. Guerra-Assunção, J. A. *et al.* Large-scale whole genome sequencing of *M. tuberculosis* provides insights into transmission in a high prevalence area. *Elife* **4**, (2015).
 34. Folkvardsen, D. B. *et al.* Genomic Epidemiology of a Major *Mycobacterium tuberculosis* Outbreak: Retrospective Cohort Study in a Low-Incidence Setting Using

- Sparse Time-Series Sampling. *J. Infect. Dis.* **216**, 366–374 (2017).
35. Roetzer, A. *et al.* Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. *PLoS Med.* **10**, e1001387 (2013).
 36. Gagneux, S. Genetic diversity in *Mycobacterium tuberculosis*. *Curr. Top. Microbiol. Immunol.* **374**, 1–25 (2013).
 37. Moreno-Molina, M. *et al.* Genomic analyses of *Mycobacterium tuberculosis* from human lung resections reveal a high frequency of polyclonal infections. *Nat. Commun.* **12**, 2716 (2021).
 38. Coscolla, M. & Gagneux, S. Does *M. tuberculosis* genomic diversity explain disease diversity? *Drug Discov. Today Dis. Mech.* **7**, e43–e59 (2010).
 39. Achtman, M. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu. Rev. Microbiol.* **62**, 53–70 (2008).
 40. Hailu, E. *et al.* Lack of methoxy-mycolates characterizes the geographically restricted lineage 7 of complex. *Microb Genom* **9**, (2023).
 41. Mitchison, D. A. *et al.* A comparison of the virulence in guinea-pigs of South Indian and British tubercle bacilli. *Tubercle* **41**, 1–22 (1960).
 42. Brudey, K. *et al.* *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol.* **6**, 23 (2006).
 43. Gagneux, S. *et al.* Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 2869–2873 (2006).
 44. Gutierrez, M. C. *et al.* Predominance of ancestral lineages of *Mycobacterium tuberculosis* in India. *Emerg. Infect. Dis.* **12**, 1367–1374 (2006).
 45. Glynn, J. R., Whiteley, J., Bifani, P. J., Kremer, K. & van Soolingen, D. Worldwide

- occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: A systematic review. *Emerg. Infect. Dis.* **8**, 843–849 (2002).
46. Varghese, B. *et al.* Impact of *Mycobacterium tuberculosis* complex lineages as a determinant of disease phenotypes from an immigrant rich moderate tuberculosis burden country. *Respir. Res.* **19**, 259 (2018).
 47. Rutaihwa, L. K. *et al.* Insights into the genetic diversity of *Mycobacterium tuberculosis* in Tanzania. *PLoS One* **14**, e0206334 (2019).
 48. Hurtado, J. *et al.* Clinical and epidemiological features of tuberculosis isolated from critically ill patients. *Rev. Argent. Microbiol.* **54**, 43–47 (2022).
 49. Ghebremichael, S. *et al.* Drug resistant *Mycobacterium tuberculosis* of the Beijing genotype does not spread in Sweden. *PLoS One* **5**, e10893 (2010).
 50. Harishankar, M., Selvaraj, P. & Bethunaickan, R. Influence of Genetic Polymorphism Towards Pulmonary Tuberculosis Susceptibility. *Front. Med.* **5**, 213 (2018).
 51. Cai, L. *et al.* The Research Progress of Host Genes and Tuberculosis Susceptibility. *Oxid. Med. Cell. Longev.* **2019**, 9273056 (2019).
 52. Liu, P. T. *et al.* Toll-like receptor triggering of a vitamin D-mediated human antimicrobial response. *Science* **311**, 1770–1773 (2006).
 53. Morales-Arce, A. Y., Sabin, S. J., Stone, A. C. & Jensen, J. D. The population genomics of within-host *Mycobacterium tuberculosis*. *Heredity* **126**, 1–9 (2021).
 54. Pérez-Lago, L. *et al.* Whole genome sequencing analysis of inpatient microevolution in *Mycobacterium tuberculosis*: potential impact on the inference of tuberculosis transmission. *J. Infect. Dis.* **209**, 98–108 (2014).
 55. Bespyatykh, J. *et al.* Metabolic Changes of during the Anti-Tuberculosis Therapy. *Pathogens* **9**, (2020).
 56. Colangeli, R. *et al.* *Mycobacterium tuberculosis* progresses through two phases of latent

- infection in humans. *Nat. Commun.* **11**, 4870 (2020).
57. Flügge, C. Die mikroorganismen. Mit besonderer berücksichtigung der ätiologie der infektionskrankheiten. Preprint at <https://doi.org/10.5962/bhl.title.53117> (1896).
 58. Flügge, C. & Royal College of Physicians of Edinburgh. Die mikroorganismen : Mit besonderer berücksichtigung der ätiologie der infektionskrankheiten /. Preprint at <https://doi.org/10.5962/bhl.title.110056> (1896).
 59. Ferguson, R. M. W. *et al.* Size fractionation of bioaerosol emissions from green-waste composting. *Environ. Int.* **147**, 106327 (2021).
 60. Williams, C. M. *et al.* Exhaled *Mycobacterium tuberculosis* output and detection of subclinical disease by face-mask sampling: prospective observational studies. *Lancet Infect. Dis.* **20**, 607–617 (2020).
 61. Dinkele, R. *et al.* Aerosolization of by Tidal Breathing. *Am. J. Respir. Crit. Care Med.* (2022) doi:10.1164/rccm.202110-2378OC.
 62. Findel, H. Vergleichende Untersuchungen über Inhalations- und Fütterungstuberkulose. *Zeitschrift für Hygiene und Infektionskrankheiten* vol. 57 104–153 Preprint at <https://doi.org/10.1007/bf02140512> (1907).
 63. Plumlee, C. R. *et al.* Ultra-low Dose Aerosol Infection of Mice with *Mycobacterium tuberculosis* More Closely Models Human Tuberculosis. *Cell Host Microbe* **29**, 68–82.e5 (2021).
 64. Medzhitov, R. & Janeway, C. A., Jr. Innate immunity: impact on the adaptive immune response. *Curr. Opin. Immunol.* **9**, 4–9 (1997).
 65. Diamond, G., Legarda, D. & Ryan, L. K. The innate immune response of the respiratory epithelium. *Immunol. Rev.* **173**, 27–38 (2000).
 66. Andersen, A. A. New sampler for the collection, sizing, and enumeration of viable airborne particles. *J. Bacteriol.* **76**, 471–484 (1958).

67. Pieters, J. & Gatfield, J. Hijacking the host: survival of pathogenic mycobacteria inside macrophages. *Trends Microbiol.* **10**, (2002).
68. Kaufmann, S. H. Recent findings in immunology give tuberculosis vaccines a new boost. *Trends Immunol.* **26**, (2005).
69. Flynn, J. L. & Chan, J. Immunology of tuberculosis. *Annu. Rev. Immunol.* **19**, (2001).
70. Pai, M. *et al.* Tuberculosis. *Nat Rev Dis Primers* **2**, 16076 (2016).
71. Smith, J. P. *et al.* Characterizing tuberculosis transmission dynamics in high-burden urban and rural settings. *Sci. Rep.* **12**, 6780 (2022).
72. Bunyasi, E. W. *et al.* Molecular Detection of Airborne in South African High Schools. *Am. J. Respir. Crit. Care Med.* **205**, 350–356 (2022).
73. Hermans, S., Boulle, A., Caldwell, J., Pienaar, D. & Wood, R. Temporal trends in TB notification rates during ART scale-up in Cape Town: an ecological analysis. *J. Int. AIDS Soc.* **18**, 20240 (2015).
74. Broekmans, J. F. *et al.* European framework for tuberculosis control and elimination in countries with a low incidence. Recommendations of the World Health Organization (WHO), International Union Against Tuberculosis and Lung Disease (IUATLD) and Royal Netherlands Tuberculosis Association (KNCV) Working Group. *Eur. Respir. J.* **19**, 765–775 (2002).
75. Mathema, B. *et al.* Drivers of Tuberculosis Transmission. *J. Infect. Dis.* **216**, S644–S653 (2017).
76. Auld, S. C. *et al.* Research Roadmap for Tuberculosis Transmission Science: Where Do We Go From Here and How Will We Know When We’re There? *J. Infect. Dis.* **216**, S662–S668 (2017).
77. van Embden, J. D. *et al.* Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J. Clin. Microbiol.* **31**,

- 406–409 (1993).
78. Small, P. M. *et al.* The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. *N. Engl. J. Med.* **330**, 1703–1709 (1994).
 79. Kasaie, P. *et al.* A Novel Tool Improves Existing Estimates of Recent Tuberculosis Transmission in Settings of Sparse Data Collection. *PLoS One* **10**, e0144137 (2015).
 80. France, A. M., Grant, J., Kammerer, J. S. & Navin, T. R. A field-validated approach using surveillance and genotyping data to estimate tuberculosis attributable to recent transmission in the United States. *Am. J. Epidemiol.* **182**, 799–807 (2015).
 81. Yuen, C. M., Kurbatova, E. V., Click, E. S., Cavanaugh, J. S. & Cegielski, J. P. Association between *Mycobacterium tuberculosis* complex phylogenetic lineage and acquired drug resistance. *PLoS One* **8**, e83006 (2013).
 82. Wobudeya, E. *et al.* Epidemiology of tuberculosis in children in Kampala district, Uganda, 2009-2010; a retrospective cross-sectional study. *BMC Public Health* **15**, 967 (2015).
 83. Bamrah, S. *et al.* Tuberculosis among the homeless, United States, 1994-2010. *Int. J. Tuberc. Lung Dis.* **17**, 1414–1419 (2013).
 84. Valença, M. S. *et al.* Prevalence of tuberculosis in prisons: risk factors and molecular epidemiology. *Int. J. Tuberc. Lung Dis.* **19**, 1182–1187 (2015).
 85. Meehan, C. J. *et al.* The relationship between transmission time and clustering methods in *Mycobacterium tuberculosis* epidemiology. *EBioMedicine* **37**, 410–416 (2018).
 86. Mathema, B., Kurepina, N. E., Bifani, P. J. & Kreiswirth, B. N. Molecular epidemiology of tuberculosis: current insights. *Clin. Microbiol. Rev.* **19**, 658–685 (2006).
 87. Chamie, G. *et al.* Identifying locations of recent TB transmission in rural Uganda: a multidisciplinary approach. *Trop. Med. Int. Health* **20**, 537–545 (2015).

88. Dowdy, D. W., Golub, J. E., Chaisson, R. E. & Saraceni, V. Heterogeneity in tuberculosis transmission and the role of geographic hotspots in propagating epidemics. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 9557–9562 (2012).
89. Zelner, J. L. *et al.* Identifying Hotspots of Multidrug-Resistant Tuberculosis Transmission Using Spatial and Molecular Genetic Data. *J. Infect. Dis.* **213**, 287–294 (2016).
90. Global Tuberculosis Report 2022.
<https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022>.
91. Bobrik, A., Danishevski, K., Eroshina, K. & McKee, M. Prison health in Russia: the larger picture. *J. Public Health Policy* **26**, 30–59 (2005).
92. Menzies, D., Joshi, R. & Pai, M. Risk of tuberculosis infection and disease associated with work in health care settings. *Int. J. Tuberc. Lung Dis.* **11**, 593–605 (2007).
93. Hamada, Y., Getahun, H., Tadesse, B. T. & Ford, N. HIV-associated tuberculosis. *Int. J. STD AIDS* **32**, 780–790 (2021).
94. Machuca, I., Vidal, E., de la Torre-Cisneros, J. & Rivero-Román, A. Tuberculosis in immunosuppressed patients. *Enferm. Infecc. Microbiol. Clin.* **36**, 366–374 (2018).
95. Tellez-Navarrete, N. A., Ramon-Luing, L. A., Muñoz-Torrico, M., Osuna-Padilla, I. A. & Chavez-Galan, L. Malnutrition and tuberculosis: the gap between basic research and clinical trials. *J. Infect. Dev. Ctries.* **15**, 310–319 (2021).
96. Lönnroth, K., Jaramillo, E., Williams, B. G., Dye, C. & Raviglione, M. Drivers of tuberculosis epidemics: the role of risk factors and social determinants. *Soc. Sci. Med.* **68**, 2240–2246 (2009).
97. Programme, G. T. Global tuberculosis report 2021.
<https://www.who.int/publications/i/item/9789240037021> (2021).

98. TB profile.
https://worldhealthorg.shinyapps.io/tb_profiles/?_inputs_&entity_type=%22country%22&lan=%22EN%22&iso2=%22ES%22.
99. Generalitat: Gabinet de Comunicació.
https://www.gva.es/portal-gva-portlet/htdocs/area_de_prensa/versionImprimiblePrensa.jsp?id_nota=932676&idioma=ES.
100. Nava Paz, O., Hassanhi, M. & Prieto, L. Evaluation of Baciloscapy, Cultivation and Polimerasa Chain Reaction for the Diagnostic of Lung Tuberculosis. *Kasmera* **33**, 119–131 (2005).
101. *WHO operational handbook on tuberculosis. Module 3: diagnosis - rapid diagnostics for tuberculosis detection, 2021 update*. (World Health Organization, 2021).
102. Williamson, D. A. *et al.* An evaluation of the Xpert MTB/RIF assay and detection of false-positive rifampicin resistance in *Mycobacterium tuberculosis*. *Diagn. Microbiol. Infect. Dis.* **74**, (2012).
103. Van Rie, A. *et al.* False-positive rifampicin resistance on Xpert® MTB/RIF: case report and clinical implications. *Int. J. Tuberc. Lung Dis.* **16**, (2012).
104. Enter a New World of Tuberculosis Diagnostics.
https://info.cepheid.com/2023-EMEA-EN-CHILD-PPC-SEARCH-CID-MTBXDR?utm_campaign=2023-emea-en-child-ppc-search-cid-mtbxdr&utm_medium=cpc&utm_source=google&utm_term=xpert%20mtb%20fxdr&LeadCampaignID=7011T000001Yvz3QAC&gclid=CjwKCAjwvdajBhBEEiwAeMh1U8FuswQEI8r4O7I_1BdwH3Ig_YWFMNg_i4PoKlbQq3jjxSUQZC7TLqhoCqrEQAvD_BwE.
105. Dong, B. *et al.* Improved Conventional and New Approaches in the Diagnosis of Tuberculosis. *Front. Microbiol.* **13**, 924410 (2022).
106. Butler, W. R. & Guthertz, L. S. Mycolic acid analysis by high-performance liquid

- chromatography for identification of *Mycobacterium* species. *Clin. Microbiol. Rev.* **14**, (2001).
107. Ali, A. *et al.* Whole genome sequencing based characterization of extensively drug-resistant *Mycobacterium tuberculosis* isolates from Pakistan. *PLoS One* **10**, (2015).
 108. Programme, G. T. WHO operational handbook on tuberculosis: module 4: treatment: drug-susceptible tuberculosis treatment.
<https://www.who.int/publications/i/item/9789240050761> (2022).
 109. Nahid, P. *et al.* Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America Clinical Practice Guidelines: Treatment of Drug-Susceptible Tuberculosis. *Clin. Infect. Dis.* **63**, (2016).
 110. Dorman, S. E. *et al.* High-dose rifapentine with or without moxifloxacin for shortening treatment of pulmonary tuberculosis: Study protocol for TBTC study 31/ACTG A5349 phase 3 clinical trial. *Contemp. Clin. Trials* **90**, 105938 (2020).
 111. Chabala, C. *et al.* Shorter treatment for minimal tuberculosis (TB) in children (SHINE): a study protocol for a randomised controlled trial. *Trials* **19**, 237 (2018).
 112. World Health Organization. *Companion Handbook to the 2011 Who Guidelines for the Programmatic Management of Multidrug-Resistant Tuberculosis*. (WHO Regional Office for the Western Pacific, 2014).
 113. Caminero, J. A. *et al.* Drug Resistant Tuberculosis: New WHO Definitions and Their Implication in the SEPAR Guideline. *Arch. Bronconeumol.* **58**, 87–89 (2022).
 114. Eldholm, V. *et al.* Evolution of extensively drug-resistant *Mycobacterium tuberculosis* from a susceptible ancestor in a single patient. *Genome Biol.* **15**, (2014).
 115. Merker, M. *et al.* Whole Genome Sequencing Reveals Complex Evolution Patterns of Multidrug-Resistant *Mycobacterium tuberculosis* Beijing Strains in Patients. *PLoS One* **8**, e82551 (2013).

116. Sun, G. *et al.* Dynamic population changes in *Mycobacterium tuberculosis* during acquisition and fixation of drug resistance in patients. *J. Infect. Dis.* **206**, (2012).
117. Eldholm, V. *et al.* Four decades of transmission of a multidrug-resistant *Mycobacterium tuberculosis* outbreak strain. *Nat. Commun.* **6**, 1–9 (2015).
118. Cohen, T. & Murray, M. Modeling epidemics of multidrug-resistant *M. tuberculosis* of heterogeneous fitness. *Nat. Med.* **10**, 1117–1121 (2004).
119. WHO announces updated definitions of extensively drug-resistant tuberculosis.
<https://www.who.int/news/item/27-01-2021-who-announces-updated-definitions-of-extensively-drug-resistant-tuberculosis>.
120. Katale, B. Z. *et al.* Whole genome sequencing of *Mycobacterium tuberculosis* isolates and clinical outcomes of patients treated for multidrug-resistant tuberculosis in Tanzania. *BMC Genomics* **21**, 174 (2020).
121. Bespyatykh, D., Bespyatykh, J., Mokrousov, I. & Shitikov, E. A Comprehensive Map of *Mycobacterium tuberculosis* Complex Regions of Difference. *mSphere* **6**, e0053521 (2021).
122. Goig, G. A. *et al.* Towards next-generation diagnostics for tuberculosis: identification of novel molecular targets by large-scale comparative genomics. *Bioinformatics* **36**, 985–989 (2020).
123. Comas, I. *et al.* Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat. Genet.* **44**, 106–110 (2011).
124. Wollenberg, K. R. *et al.* Whole-Genome Sequencing of *Mycobacterium tuberculosis* Provides Insight into the Evolution and Genetic Composition of Drug-Resistant Tuberculosis in Belarus. *J. Clin. Microbiol.* **55**, (2017).
125. Cohen, K. A. *et al.* Evolution of Extensively Drug-Resistant Tuberculosis over Four

- Decades: Whole Genome Sequencing and Dating Analysis of *Mycobacterium tuberculosis* Isolates from KwaZulu-Natal. *PLoS Med.* **12**, (2015).
126. Casali, N. *et al.* Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nat. Genet.* **46**, 279–286 (2014).
127. Yang, C. *et al.* Transmission of multidrug-resistant *Mycobacterium tuberculosis* in Shanghai, China: a retrospective observational study using whole-genome sequencing and epidemiological investigation. *Lancet Infect. Dis.* **17**, 275–284 (2017).
128. Schürch, A. C. *et al.* The tempo and mode of molecular evolution of *Mycobacterium tuberculosis* at patient-to-patient scale. *Infect. Genet. Evol.* **10**, (2010).
129. Meehan, C. J. *et al.* Whole genome sequencing of *Mycobacterium tuberculosis*: current standards and open issues. *Nat. Rev. Microbiol.* **17**, 533–545 (2019).
130. Walker, T. M. *et al.* Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect. Dis.* **15**, 1193–1202 (2015).
131. Whole-genome sequencing of *Mycobacterium tuberculosis* for rapid diagnostics and beyond. *The Lancet Respiratory Medicine* **4**, 6–8 (2016).
132. Papaventsis, D. *et al.* Whole genome sequencing of *Mycobacterium tuberculosis* for detection of drug resistance: a systematic review. *Clin. Microbiol. Infect.* **23**, 61–68 (2017).
133. Yang, C. & Gao, Q. Recent transmission of *Mycobacterium tuberculosis* in China: the implication of molecular epidemiology for tuberculosis control. *Front. Med.* **12**, 76–83 (2018).
134. Nimmo, C. *et al.* Rapid identification of a *Mycobacterium tuberculosis* full genetic drug resistance profile through whole genome sequencing directly from sputum. *Int. J. Infect. Dis.* **62**, (2017).

135. Nilgiriwala, K. *et al.* Genomic Sequencing from Sputum for Tuberculosis Disease Diagnosis, Lineage Determination, and Drug Susceptibility Prediction. *J. Clin. Microbiol.* **61**, e0157822 (2023).
136. Public Health England. England world leaders in the use of whole genome sequencing to diagnose TB. *GOV.UK*
<https://www.gov.uk/government/news/england-world-leaders-in-the-use-of-whole-genome-sequencing-to-diagnose-tb> (2017).
137. Park, M., Lalvani, A., Satta, G. & Kon, O. M. Evaluating the clinical impact of routine whole genome sequencing in tuberculosis treatment decisions and the issue of isoniazid mono-resistance. *BMC Infect. Dis.* **22**, 349 (2022).
138. [No title].
https://www.cdc.gov/tb/programs/genotyping/Tuberculosis_WGS_Training_Module.pdf
.
139. Starks, A. M. *et al.* Collaborative Effort for a Centralized Worldwide Tuberculosis Relational Sequencing Data Platform. *Clin. Infect. Dis.* **61Suppl 3**, S141–6 (2015).
140. TB Portals. <https://tbportals.niaid.nih.gov/>.
141. Reuters. IMF sees cost of COVID pandemic rising beyond \$12.5 trillion estimate. *Reuters* (2022).
142. López, M. G. *et al.* The first wave of the COVID-19 epidemic in Spain was associated with early introductions and fast spread of a dominating genetic variant. *Nat. Genet.* **53**, 1405–1414 (2021).
143. Kizny Gordon, A., Marais, B., Walker, T. M. & Sintchenko, V. Clinical and public health utility of *Mycobacterium tuberculosis* whole genome sequencing. *Int. J. Infect. Dis.* **113 Suppl 1**, S40–S42 (2021).
144. Supply, P. *et al.* Proposal for standardization of optimized mycobacterial interspersed

- repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **44**, 4498–4510 (2006).
145. Supply, P. *et al.* Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. *Mol. Microbiol.* **36**, 762–771 (2000).
 146. Kamerbeek, J. *et al.* Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol.* **35**, 907–914 (1997).
 147. Armstrong, G. L. *et al.* Pathogen Genomics in Public Health. *N. Engl. J. Med.* **381**, 2569–2580 (2019).
 148. Nelson, K. N. *et al.* Mutation of and Implications for Using Whole-Genome Sequencing for Investigating Recent Tuberculosis Transmission. *Front Public Health* **9**, 790544 (2021).
 149. Comas, I. & Gagneux, S. A role for systems epidemiology in tuberculosis research. *Trends Microbiol.* **19**, 492–500 (2011).
 150. Cox, H. *et al.* Whole-Genome Sequencing Has the Potential To Improve Treatment for Rifampicin-Resistant Tuberculosis in High-Burden Settings: a Retrospective Cohort Study. *J. Clin. Microbiol.* **60**, e0236221 (2022).
 151. Wang, L. *et al.* Whole-genome sequencing of for prediction of drug resistance. *Epidemiol. Infect.* **150**, e22 (2022).
 152. Walker, T. M. *et al.* Whole genome sequencing for M/XDR tuberculosis surveillance and for resistance testing. *Clin. Microbiol. Infect.* **23**, 161–166 (2017).
 153. Borrell, S. *et al.* Reference set of *Mycobacterium tuberculosis* clinical strains: A tool for research and product development. *PLoS One* **14**, e0214088 (2019).
 154. Xu, Y. *et al.* High-resolution mapping of tuberculosis transmission: Whole genome sequencing and phylogenetic modelling of a cohort from Valencia Region, Spain. *PLoS Med.* **16**, e1002961 (2019).

155. Jajou, R. *et al.* Towards standardisation: comparison of five whole genome sequencing (WGS) analysis pipelines for detection of epidemiologically linked tuberculosis cases. *Euro Surveill.* **24**, (2019).
156. Kohl, T. A. *et al.* Harmonized Genome Wide Typing of Tubercle Bacilli Using a Web-Based Gene-By-Gene Nomenclature System. *EBioMedicine* **34**, 131–138 (2018).
157. Walker, T. M. *et al.* Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect. Dis.* **13**, 137–146 (2013).
158. Jiménez-Ruano, A. C. *et al.* Whole genomic sequencing based genotyping reveals a specific X3 sublineage restricted to Mexico and related with multidrug resistance. *Sci. Rep.* **11**, 1870 (2021).
159. Suppli, C. H. *et al.* First outbreak of multidrug-resistant tuberculosis (MDR-TB) in Denmark involving six Danish-born cases. *Int. J. Infect. Dis.* **117**, 258–263 (2022).
160. Yang, C. *et al.* Internal migration and transmission dynamics of tuberculosis in Shanghai, China: an epidemiological, spatial, genomic analysis. *Lancet Infect. Dis.* **18**, 788–795 (2018).
161. Alaridah, N. *et al.* Transmission dynamics study of tuberculosis isolates with whole genome sequencing in southern Sweden. *Sci. Rep.* **9**, 4931 (2019).
162. Smith, N. H. *et al.* Ecotypes of the *Mycobacterium tuberculosis* complex. *J. Theor. Biol.* **239**, 220–225 (2006).
163. TB profile.
https://worldhealthorg.shinyapps.io/tb_profiles/?_inputs_&lan=%22ES%22&entity_type=%22group%22&group_code=%22global%22.
164. Muñoz, I., Vanaclocha, H., Martín-Sierra, M. & González, F. [Microbiological Surveillance Network in the Valencian community]. *Enferm. Infecc. Microbiol. Clin.* **26**, 77–81 (2008).

165. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
166. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).
167. Comas, I. *et al.* Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat. Genet.* **42**, 498–503 (2010).
168. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
169. Picard. <https://broadinstitute.github.io/picard/>.
170. bedtools: a powerful toolset for genome arithmetic — bedtools 2.30.0 documentation. <https://bedtools.readthedocs.io/en/latest/>.
171. VarScan. <http://varscan.sourceforge.net/>.
172. Website. <https://gatk.broadinstitute.org/hc/en-us>.
173. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
174. Stucki, D. *et al.* *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. *Nat. Genet.* **48**, 1535–1543 (2016).
175. The 2021 WHO catalogue of *Mycobacterium tuberculosis* complex mutations associated with drug resistance: a genotypic analysis. *The Lancet Microbe* **3**, e265–e273 (2022).
176. Feuerriegel, S. *et al.* PhyResSE: a Web Tool Delineating *Mycobacterium tuberculosis* Antibiotic Resistance and Lineage from Whole-Genome Sequencing Data. *J. Clin. Microbiol.* **53**, 1908–1914 (2015).
177. *Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance.* (World Health Organization, 2021).
178. Wickham, H. *et al.* Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).

179. Lempens, P. *et al.* Isoniazid resistance levels of *Mycobacterium tuberculosis* can largely be predicted by high-confidence resistance-conferring mutations. *Sci. Rep.* **8**, 3246 (2018).
180. Brossier, F., Veziris, N., Truffot-Pernot, C., Jarlier, V. & Sougakoff, W. Performance of the genotype MTBDR line probe assay for detection of resistance to rifampin and isoniazid in strains of *Mycobacterium tuberculosis* with low- and high-level resistance. *J. Clin. Microbiol.* **44**, 3659–3664 (2006).
181. TB profile.
https://worldhealthorg.shinyapps.io/tb_profiles/?_inputs_&entity_type=%22country%22&lan=%22EN%22&iso2=%22ES%22.
182. Nebenzahl-Guimaraes, H. *et al.* Genomic characterization of lineage 7 and a proposed name: ‘Aethiops vetus’. *Microb Genom* **2**, e000063 (2016).
183. Cancino-Muñoz, I. *et al.* Population-based sequencing of reveals how current population dynamics are shaped by past epidemics. *Elife* **11**, (2022).
184. Jajou, R. *et al.* Epidemiological links between tuberculosis cases identified twice as efficiently by whole genome sequencing than conventional molecular typing: A population-based study. *PLoS One* **13**, e0195413 (2018).
185. Walker, T. M. *et al.* Assessment of *Mycobacterium tuberculosis* transmission in Oxfordshire, UK, 2007-12, with whole pathogen genome sequences: an observational study. *Lancet Respir Med* **2**, 285–292 (2014).
186. Fox, G. J. *et al.* Household-Contact Investigation for Detection of Tuberculosis in Vietnam. *N. Engl. J. Med.* **378**, 221–229 (2018).
187. Foster, I. *et al.* The role of counselling in tuberculosis diagnostic evaluation and contact tracing: scoping review and stakeholder consultation of knowledge and research gaps. *BMC Public Health* **22**, 1–18 (2022).

188. Diel, R. *et al.* Accuracy of whole-genome sequencing to determine recent tuberculosis transmission: an 11-year population-based study in Hamburg, Germany. *Eur. Respir. J.* **54**, (2019).
189. Yates, T. A. *et al.* The transmission of *Mycobacterium tuberculosis* in high burden settings. *Lancet Infect. Dis.* **16**, 227–238 (2016).
190. Auld, S. C. *et al.* Extensively drug-resistant tuberculosis in South Africa: genomic evidence supporting transmission in communities. *Eur. Respir. J.* **52**, (2018).
191. Middelkoop, K. *et al.* Transmission of tuberculosis in a South African community with a high prevalence of HIV infection. *J. Infect. Dis.* **211**, 53–61 (2015).
192. Verver, S. *et al.* Proportion of tuberculosis transmission that takes place in households in a high-incidence area. *Lancet* **363**, 212–214 (2004).
193. McCreesh, N. & White, R. G. An explanation for the low proportion of tuberculosis that results from transmission between household and known social contacts. *Sci. Rep.* **8**, 5382 (2018).
194. Surie, D. *et al.* Molecular, Spatial, and Field Epidemiology Suggesting TB Transmission in Community, Not Hospital, Gaborone, Botswana. *Emerg. Infect. Dis.* **23**, 487–490 (2017).
195. Andrews, J. R., Morrow, C., Walensky, R. P. & Wood, R. Integrating social contact and environmental data in evaluating tuberculosis transmission in a South African township. *J. Infect. Dis.* **210**, 597–603 (2014).
196. A model on the influence of age on immunity to infection with *Mycobacterium tuberculosis*. *Exp. Gerontol.* **43**, 275–285 (2008).
197. Ramos, L., Lunney, J. K. & Gonzalez-Juarrero, M. Neonatal and infant immunity for tuberculosis vaccine development: importance of age-matched animal models. *Dis. Model. Mech.* **13**, dmm045740 (2020).

198. Vesosky, B. & Turner, J. The influence of age on immunity to infection with *Mycobacterium tuberculosis*. *Immunol. Rev.* **205**, 229–243 (2005).
199. Genetic and hormonal mechanisms underlying sex-specific immune responses in tuberculosis. *Trends Immunol.* **43**, 640–656 (2022).
200. Min, J. *et al.* Differential effects of sex on tuberculosis location and severity across the lifespan. *Sci. Rep.* **13**, 6023 (2023).
201. The role of sex differences in the prevalence and transmission of tuberculosis. *Kekkaku* **93**, 104–107 (2013).
202. Walker, T. M. *et al.* transmission in Birmingham, UK, 2009-19: An observational study. *Lancet Reg Health Eur* **17**, 100361 (2022).
203. Condenados según sexo, edad y nacionalidad. *INE* <https://www.ine.es/>.
204. Portal Plan Nacional sobre Drogas - Observatorio Español de las Drogas y las Adicciones (OEDA). <https://pnsd.sanidad.gob.es/profesionales/sistemasInformacion/home.htm>.
205. Madrazo-Moya, C. F. *et al.* Whole genomic sequencing as a tool for diagnosis of drug and multidrug-resistance tuberculosis in an endemic region in Mexico. *PLoS One* **14**, e0213046 (2019).
206. Coll, F. *et al.* Genome-wide analysis of multi- and extensively drug-resistant *Mycobacterium tuberculosis*. *Nat. Genet.* **50**, 307–316 (2018).
207. Shea, J. *et al.* Comprehensive Whole-Genome Sequencing and Reporting of Drug Resistance Profiles on Clinical Cases of *Mycobacterium tuberculosis* in New York State. *J. Clin. Microbiol.* **55**, 1871–1882 (2017).
208. The CRyPTIC Consortium. Genome-wide association studies of global *Mycobacterium tuberculosis* resistance to 13 antimicrobials in 10,228 genomes identify new resistance mechanisms. *PLoS Biol.* **20**, e3001755 (2022).

209. Morlock, G. P., Metchock, B., Sikes, D., Crawford, J. T. & Cooksey, R. C. *ethA*, *inhA*, and *katG* loci of ethionamide-resistant clinical *Mycobacterium tuberculosis* isolates. *Antimicrob. Agents Chemother.* **47**, 3799–3805 (2003).
210. Population-based resistance of *Mycobacterium tuberculosis* isolates to pyrazinamide and fluoroquinolones: results from a multicountry surveillance project. *Lancet Infect. Dis.* **16**, 1185–1192 (2016).
211. Tong, X. *et al.* Association between the mannose-binding lectin (MBL)-2 gene variants and serum MBL with pulmonary tuberculosis: An update meta-analysis and systematic review. *Microb. Pathog.* **132**, 374–380 (2019).
212. van Tong, H., Velavan, T. P., Thye, T. & Meyer, C. G. Human genetic factors in tuberculosis: an update. *Trop. Med. Int. Health* **22**, 1063–1071 (2017).
213. Mashabela, G. T., de Wet, T. J. & Warner, D. F. *Mycobacterium tuberculosis* metabolism. *Microbiol. Spectr.* **7**, (2019).
214. Ziskind, B. & Halioua, B. [Tuberculosis in ancient Egypt]. *Rev. Mal. Respir.* **24**, 1277–1283 (2007).
215. Allison, M. J., Mendoza, D. & Pezzia, A. Documentation of a case of tuberculosis in Pre-Columbian America. *Am. Rev. Respir. Dis.* **107**, 985–991 (1973).
216. Mokrousov, I. *et al.* Early ancient sublineages of *Mycobacterium tuberculosis* Beijing genotype: unexpected clues from phylogenomics of the pathogen and human history. *Clin. Microbiol. Infect.* **25**, 1039.e1–1039.e6 (2019).
217. Shuaib, Y. A. *et al.* Origin and Global Expansion of Complex Lineage 3. *Genes* **13**, (2022).
218. Reed, M. B. *et al.* Major *Mycobacterium tuberculosis* lineages associate with patient country of origin. *J. Clin. Microbiol.* **47**, 1119–1128 (2009).
219. Singh, J. *et al.* A comparative analysis of molecular genotypes of isolates from

- HIV-positive and HIV-negative patients. *Front. Cell. Infect. Microbiol.* **12**, 953443 (2022).
220. Aguilar, D. *et al.* *Mycobacterium tuberculosis* strains with the Beijing genotype demonstrate variability in virulence associated with transmission. *Tuberculosis* **90**, 319–325 (2010).
221. Ates, L. S. *et al.* Mutations in ppe38 block PE_PGRS secretion and increase virulence of *Mycobacterium tuberculosis*. *Nat Microbiol* **3**, 181–188 (2018).
222. Lee, R. S. *et al.* Population genomics of *Mycobacterium tuberculosis* in the Inuit. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13609–13614 (2015).
223. Menardo, F. *et al.* Local adaptation in populations of endemic to the Indian Ocean Rim. *F1000Res.* **10**, 60 (2021).
224. Phelan, J. *et al.* Genome-wide host-pathogen analyses reveal genetic interaction points in tuberculosis disease. *Nat. Commun.* **14**, 549 (2023).
225. Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D. W. & Corander, J. Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Res.* **47**, 5539–5549 (2019).
226. Ruis, C. *et al.* Dissemination of *Mycobacterium abscessus* via global transmission networks. *Nat Microbiol* **6**, 1279–1288 (2021).
227. Reid, C. J. *et al.* A role for ColV plasmids in the evolution of pathogenic *Escherichia coli* ST58. *Nat. Commun.* **13**, 683 (2022).
228. Argimón, S. *et al.* Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb Genom* **2**, e000093 (2016).
229. Cui, N., Chen, T., Liao, B., Xu, J. & Li, X. The biology of medicinal resource substitution in *Salvia*. *Chin. Med.* **16**, 141 (2021).
230. Zhou, S.-Y.-D. *et al.* Conurbation size drives antibiotic resistance along the river. *Sci.*

- Total Environ.* **823**, 153822 (2022).
231. GitHub - lh3/seqtk: Toolkit for processing sequences in FASTA/Q formats. *GitHub*
<https://github.com/lh3/seqtk>.
232. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021).
233. Yu, Y., Blair, C. & He, X. RASP 4: Ancestral State Reconstruction Tool for Multiple Genes and Characters. *Mol. Biol. Evol.* **37**, 604–606 (2020).
234. Yu, Y., Harris, A. J., Blair, C. & He, X. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. *Mol. Phylogenet. Evol.* **87**, 46–49 (2015).
235. Said, H. M. *et al.* Molecular characterization and second-line antituberculosis drug resistance patterns of multidrug-resistant *Mycobacterium tuberculosis* isolates from the northern region of South Africa. *J. Clin. Microbiol.* **50**, 2857–2862 (2012).
236. Liu, Q. *et al.* Local adaptation of on the Tibetan Plateau. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
237. Zwyrer, M. *et al.* Back-to-Africa introductions of *Mycobacterium tuberculosis* as the main cause of tuberculosis in Dar es Salaam, Tanzania. *PLoS Pathog.* **19**, e1010893 (2023).
238. Gutiérrez, S., Michalakis, Y. & Blanc, S. Virus population bottlenecks during within-host progression and host-to-host transmission. *Curr. Opin. Virol.* **2**, 546–555 (2012).
239. Sobel Leonard, A., Weissman, D. B., Greenbaum, B., Ghedin, E. & Koelle, K. Transmission Bottleneck Size Estimation from Pathogen Deep-Sequencing Data, with an Application to Human Influenza A Virus. *J. Virol.* **91**, (2017).
240. Sakula, A. Robert Koch: centenary of the discovery of the tubercle bacillus, 1882.

- Thorax* **37**, 246–251 (1982).
241. Donald, P. R. *et al.* Droplets, dust and guinea pigs: an historical review of tuberculosis transmission research, 1878-1940. *Int. J. Tuberc. Lung Dis.* **22**, 972–982 (2018).
 242. Heymann, B. Ueber die Ausstreuung infectiöser Tröpfchen beim Husten der Phthisiker. *Zeitschrift für Hygiene und Infektionskrankheiten* vol. 30 139–162 Preprint at <https://doi.org/10.1007/bf02198685> (1899).
 243. Hippke, E. Neue Versuche über die Bedeutung der Tröpfcheninfektion für die Ausbreitung der Lungenschwindsucht. *Zeitschrift für Hygiene und Infektionskrankheiten* vol. 93 122–146 Preprint at <https://doi.org/10.1007/bf02175466> (1921).
 244. Saini, D. *et al.* Ultra-low dose of *Mycobacterium tuberculosis* aerosol creates partial infection in mice. *Tuberculosis* **92**, 160–165 (2012).
 245. Dinkele, R. *et al.* Capture and visualization of live *Mycobacterium tuberculosis* bacilli from tuberculosis patient bioaerosols. *PLoS Pathog.* **17**, e1009262 (2021).
 246. Theron, G. *et al.* Bacterial and host determinants of cough aerosol culture positivity in patients with drug-resistant versus drug-susceptible tuberculosis. *Nat. Med.* **26**, 1435–1443 (2020).
 247. Dinkele, R. *et al.* Aerosolization of by Tidal Breathing. *Am. J. Respir. Crit. Care Med.* (2022) doi:10.1164/rccm.202110-2378OC.
 248. Cadena, A. M., Flynn, J. L. & Fortune, S. M. The Importance of First Impressions: Early Events in *Mycobacterium tuberculosis* Infection Influence Outcome. *MBio* **7**, e00342–16 (2016).
 249. The Pathogenesis of Tuberculosis The Pathogenesis of Tuberculosis. By RichArnold R., M.D., Associate Professor of Pathology, The Johns Hopkins University School of Medicine, Baltimore. A volume of 909 pages; 35-page index; 89 figures; 20 tables; 1417 references; 4 charts. Published by Charles C. Thomas, Springfield, Ill. Price \$10.50

- postpaid. *Radiology* **43**, 394–395 (1944).
250. Hall, J. S., French, R., Hein, G. L., Morris, T. J. & Stenger, D. C. Three distinct mechanisms facilitate genetic isolation of sympatric wheat streak mosaic virus lineages. *Virology* **282**, 230–236 (2001).
 251. Sacristán, S., Díaz, M., Fraile, A. & García-Arenal, F. Contact transmission of Tobacco mosaic virus: a quantitative analysis of parameters relevant for virus evolution. *J. Virol.* **85**, 4974–4981 (2011).
 252. Poon, L. L. M. *et al.* Quantifying influenza virus diversity and transmission in humans. *Nat. Genet.* **48**, 195–200 (2016).
 253. Emmett, K. J., Lee, A., Khiabani, H. & Rabadan, R. High-resolution Genomic Surveillance of 2014 Ebolavirus Using Shared Subclonal Variants. *PLoS Curr.* **7**, (2015).
 254. Braun, K. M. *et al.* Acute SARS-CoV-2 infections harbor limited within-host diversity and transmit via tight transmission bottlenecks. *PLoS Pathog.* **17**, e1009849 (2021).
 255. Sobel Leonard, A. *et al.* Deep Sequencing of Influenza A Virus from a Human Challenge Study Reveals a Selective Bottleneck and Only Limited Intra-host Genetic Diversification. *J. Virol.* **90**, 11247–11258 (2016).
 256. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
 257. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
 258. Picard. <http://broadinstitute.github.io/picard/>.
 259. Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
 260. Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**,

- 11.10.1–11.10.33 (2013).
261. Wilm, A. *et al.* LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res.* **40**, 11189–11201 (2012).
262. Huang, W., Li, L., Myers, J. R. & Marth, G. T. ART: a next-generation sequencing read simulator. *Bioinformatics* **28**, 593–594 (2012).
263. Marin, M. *et al.* Benchmarking the empirical accuracy of short-read sequencing across the *M. tuberculosis* genome. *Bioinformatics* **38**, 1781–1787 (2022).
264. Jones-López, E. C. *et al.* Incident *Mycobacterium tuberculosis* infection in household contacts of infectious tuberculosis patients in Brazil. *BMC Infect. Dis.* **17**, 576 (2017).
265. Ma, Y. *et al.* Using Cure Models to Estimate the Serial Interval of Tuberculosis With Limited Follow-up. *Am. J. Epidemiol.* **189**, 1421–1426 (2020).
266. Rossi, G. *et al.* Identifying likely transmissions in *Mycobacterium bovis* infected populations of cattle and badgers using the Kolmogorov Forward Equations. *Sci. Rep.* **10**, 21980 (2020).
267. López, M. G. *et al.* Tuberculosis in Liberia: high multidrug-resistance burden, transmission and diversity modelled by multiple importation events. *Microb Genom* **6**, (2020).
268. Vázquez-Chacón, C. A. *et al.* Intra-host genetic population diversity: Role in emergence and persistence of drug resistance among *Mycobacterium tuberculosis* complex minor variants. *Infect. Genet. Evol.* **101**, 105288 (2022).
269. Comas, I. Genomic Epidemiology of Tuberculosis. *Adv. Exp. Med. Biol.* **1019**, 79–93 (2017).
270. Martin, C. J. *et al.* Digitally Barcoding Reveals Infection Dynamics in the Macaque Model of Tuberculosis. *MBio* **8**, (2017).

271. Dean, G. S. *et al.* Minimum infective dose of *Mycobacterium bovis* in cattle. *Infect. Immun.* **73**, 6467–6471 (2005).
272. Ziesché, H. Über die quantitativen Verhältnisse der Tröpfchenausstreung durch hustende Phthisiker. *Z. Hyg. Infektionskr.* **57**, 50–82 (1907).
273. Melsew, Y. A. *et al.* The role of super-spreading events in *Mycobacterium tuberculosis* transmission: evidence from contact tracing. *BMC Infect. Dis.* **19**, 244 (2019).
274. García-Marín, A. M. *et al.* Role of the first WHO mutation catalogue in the diagnosis of antibiotic resistance in *Mycobacterium tuberculosis* in the Valencia Region, Spain: a retrospective genomic analysis. *Lancet Microbe* (2023)
doi:10.1016/S2666-5247(23)00252-5.
275. Pasechnik, O. *et al.* Major genotype families and epidemic clones of *Mycobacterium tuberculosis* in Omsk region, Western Siberia, Russia, marked by a high burden of tuberculosis-HIV coinfection. *Tuberculosis* **108**, 163–168 (2018).
276. Haldane, J. B. S. The rate of mutation of human genes. *Hereditas* **35**, 267–273 (2010).
277. Tamalet, C., Devaux, C., Dubourg, G. & Colson, P. Resistance to human immunodeficiency virus infection: a rare but neglected state. *Ann. N. Y. Acad. Sci.* **1485**, 22–42 (2021).
278. Hatherell, H.-A. *et al.* Interpreting whole genome sequencing for investigating tuberculosis transmission: a systematic review. *BMC Med.* **14**, 21 (2016).
279. Dodd, P. J. *et al.* Transmission modeling to infer tuberculosis incidence prevalence and mortality in settings with generalized HIV epidemics. *Nat. Commun.* **14**, 1639 (2023).
280. Winglee, K. *et al.* Logically Inferred Tuberculosis Transmission (LITT): A Data Integration Algorithm to Rank Potential Source Cases. *Front Public Health* **9**, 667337 (2021).

Abreviaturas

TB tuberculosis

OMS organización mundial de la salud

WGS Whole Genome Sequencing (Secuenciación del genoma completo)

DNA Deoxyribonucleic acid (Ácido desoxirribonucleico)

MTBC *Mycobacterium tuberculosis* complex (Complejo *Mycobacterium tuberculosis*)

VIH virus de la inmunodeficiencia humana

PPL personas privadas de su libertad

SNP single Nucleotide Polymorphism (Polimorfismo de un solo nucleótido)

fSNP fixed SNP (Polimorfismo de un solo nucleótido, fijado)

vSNP variable SNP (Polimorfismo de un solo nucleótido, variable)

SM estreptomicina

INH isoniazida

RIF rifampicina

EMB etambutol

PZA pirazinamida

KM kanamicina

AM amikacina

CM capreomicina

OFX ofloxacino

LFX levofloxacino

ETO etionamida

LZD linezolid

PAS ácido p-amino salicílico

DST (Drug susceptibility testing) test de susceptibilidad a fármacos

RR rifampicina resistente

MDR multidrogoresistente

XDR extremodrogorresistente

MTB anc common ancestor of MTBC (ancestro común del complejo *Mycobacterium tuberculosis*)

PDU parenteral drug user (usuario de drogas por vía parenteral)

SI Serial interval (intervalo serial)

MLE (maximum likelihood estimation) máxima verosimilitud

NTM Nontuberculous mycobacteria

Suplementario

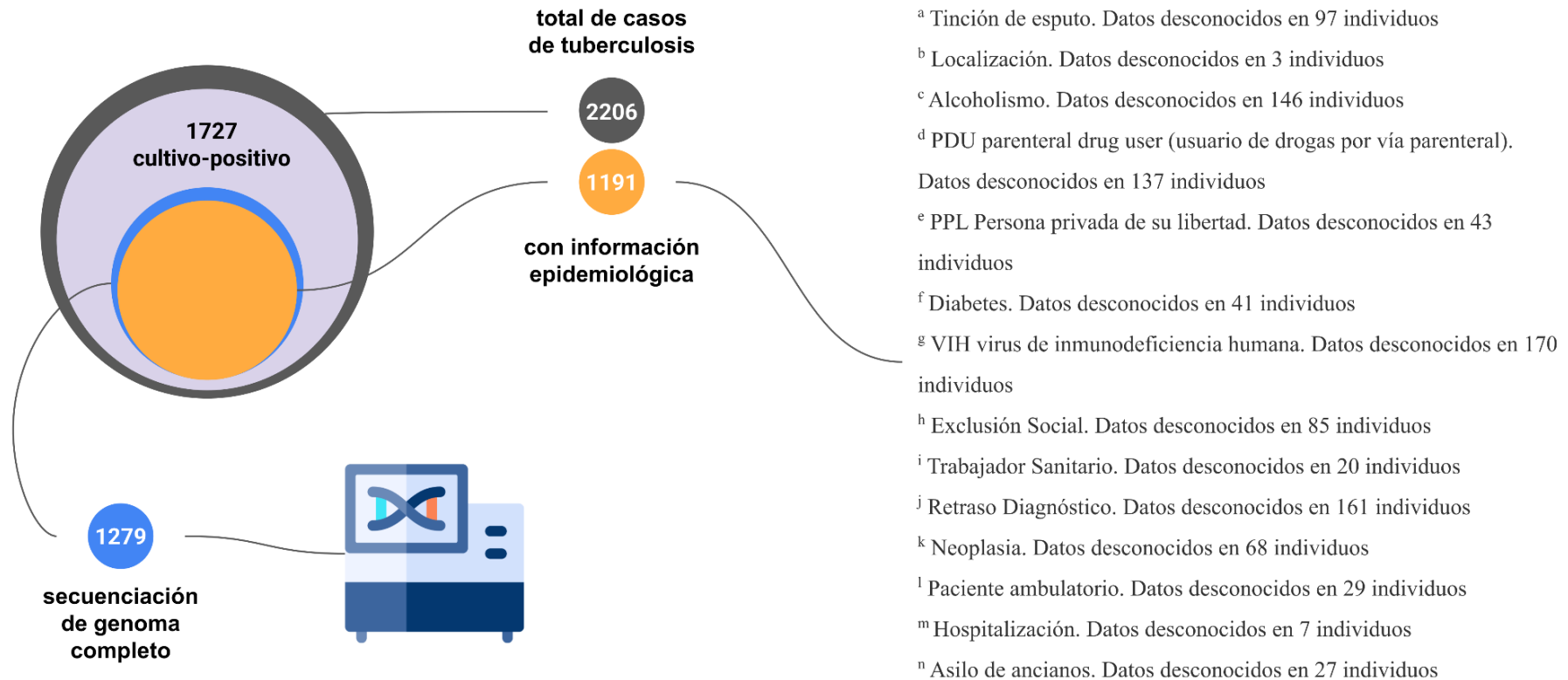


Figura suplementaria 3.0. Diagrama de inclusión y exclusión de casos de tuberculosis en la Comunidad Valenciana entre 2014 y 2018. De 2206 casos registrados (gris), 1727 fueron cultivo-positivos (beige), 1279 secuenciados (azul) y 1191 contaron con datos epidemiológicos (amarillo). Los superíndices A-N indican datos faltantes por variable dentro de las 1191 muestras.

Tabla 3.10. Genes y fármacos incluidos en la predicción de resistencias del catálogo de la Organización Mundial de la Salud (OMS)

Gen / Índice Genómico	Sentido	Inicio de la región	Inicio de codón	Final de la región	Longitud (pb)	Antibiótico	Nivel OMS
Primera línea							
<i>katG</i> (corriente arriba incluye todo <i>furA</i>)	R	2156652	2156111	2153889	2764	Isoniazida	1
<i>ahpC</i>	F	2726053	2726193	2726780	728	Isoniazida	1
<i>inhA</i> (corriente arriba incluye todo <i>fabG1</i>)	F	1673336	1674202	1675011	1676	Isoniazida, Etionamida	1
<i>mshA</i>	F	574670	575348	576790	2121	Isoniazida, Etionamida	2
<i>ndh</i>	R	2103147	2103042	2101651	1497	Isoniazida, Etionamida	2
<i>Rv1258c</i>	R	1407407	1407340	1406081	1327	Isoniazida, Pirazinamida, Estreptomicina	1 (Estreptomicina); 2 (Isoniazida, Pirazinamida)
<i>Rv2752c</i>	R	3066221	3066191	3064515	1707	Isoniazida, Rifampicina	2
<i>Rv2752c</i> (promotor corriente arriba de <i>dapA</i>)	R	3067184		3067125	60	Isoniazida, Rifampicina	2
<i>rpoB-rpoC</i>	F	759420	759807	767320	7901	Rifampicina	1 (<i>rpoB</i>); 2 (<i>rpoC</i>)
<i>rpoA</i>	R	3879052	3878507	3877464	1589	Rifampicina	2
<i>pncA</i>	R	2289301	2289241	2288681	621	Pirazinamida	1
<i>panD</i> (promotor corriente arriba de <i>Rv3603c</i>)	R	4046239		4046119	121	Pirazinamida	1
<i>panD</i>	R		4044281	4043862	420	Pirazinamida	1
<i>clpC1</i>	R	4040819	4040704	4038158	2662	Pirazinamida	1
<i>PPE35</i>	R	2170743	2170612	2167649	3095	Pirazinamida	2
<i>Rv3236c</i>	R	3613663	3613116	3611959	1705	Pirazinamida	2
<i>embC-embB</i> (corriente arriba de <i>Rv3789</i>)	F	4235250		4235373	124	Etambutol	1
<i>embC-embB</i> (corriente arriba de <i>dprE1</i>)	F	4235718		4235778	61	Etambutol	1
<i>embC-embB</i> (corriente arriba incluye todo <i>aftA</i>)	F	4237932	4239863	4249810	11879	Etambutol	1
<i>embR</i>	R	1417624	1417347	1416181	1444	Etambutol	2

<i>ubiA</i>	R	4269839	4269833	4268925	915	Etambutol	2
<i>rpsL</i>	F	781300	781560	781934	635	Estreptomicina	1
<i>gid</i>	R	4408322	4408202	4407528	795	Estreptomicina	1

Gen / Índice Genómico	Sentido	Inicio de la región	Inicio de codón	Final de la región	Longitud (pb)	Antibiótico	Nivel OMS
Segunda línea							
<i>gyrB-gyrA</i>	F	5069	5240	9818	4750	Moxifloxacino, Levofloxacino	1
<i>ccsA</i>	F	619691	619891	620865	1175	Amikacina, Capreomicina	2
<i>fprA</i>	F	3473843	3474007	3475377	1535	Amikacina, Capreomicina	2
<i>aftB</i>	R	4268974	4268836	4266953	2022	Amikacina, Capreomicina	2
<i>eis</i>	R	2715394	2715332	2714124	1271	Amikacina, Kanamicina	1
<i>rrs</i>	F	1471520	1471846	1473382	1863	Amikacina, Estreptomicina, Capreomicina, Kanamicina	1
<i>whiB6</i>	R	4338714	4338521	4338171	544	Amikacina, Estreptomicina, Capreomicina	2
<i>whiB7</i>	R	3569092	3568679	3568401	692	Amikacina, Estreptomicina, Kanamicina	1
<i>tlyA</i>	F	1917695	1917940	1918746	1052	Capreomicina	1
<i>atpE</i>	F	1460985	1461045	1461290	306	Bedaquilina	1
<i>pepQ</i>	R	2860478	2860418	2859300	1179	Bedaquilina, Clofazimina	1
<i>Rv1979c</i>	R	2223643	2223164	2221719	1925	Bedaquilina, Clofazimina	2
<i>mmpL5-Rv0678</i> (incluida toda la región intergénica)	R & F	775586		779487	3902	Bedaquilina, Clofazimina	1
<i>ddn</i>	F	3986682	3986844	3987299	618	Delamanida	1

<i>fbiA-fbiB</i>	F	3640385	3640543	3642881	2497	Delamanida	1
<i>fbiC</i>	F	1302795	1302931	1305501	2707	Delamanida	1
<i>Rv2983</i>	F	3339058	3339118	3339762	705	Delamanida	1
<i>fgdI</i>	F	490723	490783	491793	1071	Delamanida	1
<i>Rv3083</i>	F	3448444	3448504	3449991	1548	Etionamida	2
<i>ethA-ethR</i> (incluida toda la región intergénica)	R & F	4326004		4328199	2196	Etionamida	1 (<i>ethA</i>); 2 (<i>ethR</i>)
<i>rplC</i>	F	800297	800809	801462	1166	Linezolida	1
<i>rrl</i>	F	1473383	1473658	1476795	3413	Linezolida	1
Para Linezolid: <i>rrl</i> más región intergénica							
<i>rrs-rrl</i> más promotor del operón que está corriente arriba de <i>rrs</i> , pero no de <i>rrs</i> en sí mismo	-	-	-	-	-	Linezolida	1

Adaptado de la tabla original de genes y regiones de predicción de resistencia empleada por la OMS ¹⁷⁵

Tabla 3.11. Comparación del porcentaje de valores de concordancia para la predicción de farmacorresistencia obtenidos por WGS en este y otros estudios.

Fármaco	Estudio actual, 2022 Valencia (n = 1279)		Zenteno et al, 2018 México (n = 81)		Walker et al, 2015 Inglaterra (n = 701)		Coll et al, 2018 Global (n = 6465)		Shea et al, 2017 New York (n = 608)		Cryptic Consortium, 2018 (n = 10,209)	
	Sen	Esp	Sen	Esp	Sen	Esp	Sen	Esp	Sen	Esp	Sen	Esp
Isoniazida	96	100	84	100	85	98	90	97	87	99	97	99
Rifampicina	69	100	96	94	92	99	93	98	100	97	97	99
Etambutol	50	99	71	90	82	95	90	92	79	99	94	94
Pirazinamida	51	100	75	90	24	100	65	92	86	98	91	96

Sen = Sensibilidad (%). Esp = Especificidad (%).

Tabla 3.12. Panel de tablas tetracóricas para el cálculo de sensibilidad y especificidad de resistencia a fármacos

		DST			
		INH			
		Resistente	Sensible		
WGS	Resistente	65	3	68	Sensibilidad 0.96
	Sensible	3	1097	1100	Especificidad 1.00
		68	1100	1168	

		DST			
		RIF			
		Resistente	Sensible		
WGS	Resistente	20	0	20	Sensibilidad 0.69
	Sensible	9	1141	1150	Especificidad 1.00
		29	1141	1170	

		DST			
		EMB			
		Resistente	Sensible		
WGS	Resistente	10	6	16	Sensibilidad 0.50
	Sensible	10	1136	1146	Especificidad 0.99
		20	1142	1162	

		DST			
		PZA			
		Resistente	Sensible		
WGS	Resistente	19	1	20	Sensibilidad 0.51
	Sensible	18	1117	1135	Especificidad 1.00
		37	1118	1155	

* Abreviaturas: DST, prueba de susceptibilidad a los medicamentos; WGS (Whole genome sequencing): secuenciación del genoma.

Tabla 4.3. Resultados individuales del test de Mantel de los 75 genotipos

Genotipo	r_m	p	Genotipo	r_m	p	Genotipo	r_m	p	Genotipo	r_m	p
1	-0.280	0.914	21	-0.527	0.892	41	0.045	0.405	61	0.060	0.196
2	-0.010	0.508	22	-	-	42	-0.014	0.474	62	0.051	0.227
3	-0.083	0.702	23	-	-	43	0.237	0.350	63	0.006	0.427
4	0.200	0.250	24	-0.371	0.708	44	-	-	64	-0.190	0.926
5	-0.031	0.618	25	-0.096	0.718	45	-	-	65	-0.011	0.560
6	0.188	0.206	26	-0.063	0.611	46	0.500	0.500	66	0.771	0.250
7	-0.133	0.894	27	-	-	47	0.239	0.200	67	-	-
8	-1.000	1.000	28	0.133	0.211	48	0.917	0.050	68	-	-
9	-0.206	0.813	29	0.062	0.383	49	0.053	0.144	69	0.500	0.500
10	-0.029	0.547	30	0.000	0.667	50	0.258	0.200	70	-0.039	0.488
11	-0.013	0.583	31	-0.426	0.800	51	-0.064	0.941	71	-	-
12	-0.097	0.881	32	-	-	52	-0.127	0.957	72	-1.000	1.000
13	-0.124	0.819	33	-	-	53	0.140	0.240	73	0.331	0.118
14	-0.025	0.542	34	-0.582	1.000	54	0.324	0.333	74	0.004	0.410
15	-0.112	0.785	35	-	-	55	0.394	0.250	75	-	-
16	-	-	36	-	-	56	-0.085	0.847			
17	-	-	37	-	-	57	-	-			
18	-	-	38	0.314	0.417	58	0.019	0.409			
19	-	-	39	-	-	59	0.105	0.203			
20	-	-	40	-0.101	0.667	60	0.105	0.192			

r_m : Coeficiente de correlación de Mantel; p : Valor p asociado con la correlación ; “-”No se pudo realizar la prueba debido a menos de 3 aislados por genotipo.

Tabla 5.1. Variables sociodemográficas y genómicas de los 21 pares de transmisión donante-recepto

Pair	Donor	Sample	SNPs	Symtom date	SI (Days)	Bottleneck size		Donor	%	Shared and unique SNPs				Total (0.02 -0.98)	Run stats		Lineage	Anatomical location	Sample	Age	Sex	Relation
						Aprox	Exact			Recipient	%	Share	%		Mean depth	Median deep						
1	☒	G2411	0	1/08/2017	83	31	32	32	0,25	14	0,11	82	0,64	128	569	601	4.1.2.1	Pulmonar	Sputum	31 años	H	Agricultural worker
	☐	G2397	0	23/10/2017											531	558	4.1.2.1	Pulmonar	Sputum	40 años	H	Agricultural worker
2	☒	G1364	0	1/06/2015	578	29	28	16	0,03	406	0,81	82	0,16	504	550	582	4.1.2.1	Pulmonar	Sputum	36 años	H	Brothers
	☐	G2179	0	30/12/2016											632	600	4.1.2.1	Pulmonar	Sputum	40 años	M	Brothers
3	☒	G0945	0	10/05/2015	NA	33	38	24	0,26	4	0,04	65	0,70	93	783	824	4.1.0	Pulmonar	Sputum	27 años	M	Aunt and niece
	☐	G1023	0	NA											654	690	4.1.0	Pulmonar	Gastric secretion	4 años	M	Aunt and niece
4	☒	G1564	0	4/11/2015	61	47	63	15	0,15	16	0,16	69	0,69	100	771	817	4.3.2	Pulmonar	Sputum	79 años	H	Husband and wife
	☐	G1742	0	4/01/2016											688	720	4.3.2	Pulmonar	Sputum	39 años	M	Husband and wife
5	☒	G1611	0	1/04/2016	521	18	16	28	0,24	32	0,27	58	0,49	118	582	615	4.1.2.1	Pulmonar	Sputum	22 años	H	Cohabitants
	☐	G2383	0	4/09/2017											438	456	4.1.2.1	Pulmonar	Sputum	18 años	M	Cohabitants
6	☒	G1128	0	9/04/2015	207	92	98	11	0,12	6	0,06	77	0,82	94	585	617	4.1.2	Pulmonar	Sputum	25 años	M	Relatives
	☐	G1562	0	2/11/2015											537	568	4.1.2	Pulmonar	Sputum	48 años	M	Relatives
7	☒	G1544	0	7/03/2016	42	39	40	22	0,18	36	0,30	62	0,52	120	428	452	4.1.0	Pulmonar	Sputum	49 años	H	Grandfather and grandson
	☐	G1630	0	18/04/2016											704	732	4.1.0	Pulmonar	Strain	9 años	H	Grandfather and grandson
8	☒	G2154	0	28/04/2017	38	27	32	18	0,28	10	0,15	37	0,57	65	657	690	4.3.4.2	Pulmonar	Sputum	38 años	H	Bar cohabitants
	☐	G2160	0	5/06/2017											408	432	4.3.4.2	Pulmonar	Sputum	43 años	H	Bar cohabitants
9	☒	G0262	0	22/05/2014	162	20	21	17	0,18	17	0,18	60	0,64	94	566	594	4.3.2	Pulmonar	NA	51 años	H	Father and child
	☐	G0641	0	31/10/2014											519	546	4.3.2	Pulmonar	Sputum	14 años	H	Father and child
10	☒	G2606	0	3/04/2018	NA	13	13	39	0,31	18	0,14	69	0,55	126	252	264	4.1.2.1	Pulmonar	Sputum	46 años	M	Cohabitants
	☐	G2610	0	NA											409	427	4.1.2.1	Pulmonar	Sputum	17 años	M	Cohabitants
11	☒	G1570	0	9/11/2015	93	100	130	8	0,09	17	0,20	61	0,71	86	521	548	4.1.0	Pulmonar	Sputum	32 años	H	
	☐	G1578	0	10/02/2016											647	685	4.1.0	Pulmonar	Gastric secretion	18 meses	H	
12	☒	G1258	1	14/01/2014	55	12	13	21	0,25	13	0,16	49	0,59	83	506	532	4.3.3	Pulmonar	Sputum	29 años	M	Mother and child
	☐	G0128	1	10/03/2014											704	742	4.3.3	Pulmonar	Gastric secretion	4 años	H	Mother and child
13	☒	G1711	1	15/03/2016	358	14	14	46	0,28	21	0,13	95	0,59	162	739	766	4.1.2.1	Pulmonar	Sputum	27 años	M	Work cohabitants
	☐	G2146	1	8/03/2017											510	533	4.1.2.1	Pulmonar	Sputum	51 años	M	Work cohabitants
14	☒	G1824	1	8/09/2016	493	36	43	35	0,19	52	0,28	96	0,52	183	596	623	4.1.2.1	Pulmonar	Sputum	46 años	H	Bar cohabitants
	☐	G2581	1	14/01/2018											451	375	4.1.2.1	Pulmonar	Sputum	65 años	H	Bar cohabitants
15	☒	G0566	1	7/10/2013	217	38	43	38	0,30	12	0,10	76	0,60	126	685	717	4.1.2.1	Pulmonar	Sputum	40 años	H	
	☐	G0553	1	12/05/2014											567	598	4.1.2.1	Pulmonar	Sputum	77 años	M	
16	☒	G0108	0	1/05/2014	133	40	45	19	0,17	37	0,34	54	0,49	110	586	615	4.6.1.1	Pulmonar	Sputum	43 años	H	
	☐	G0260	0	11/09/2014											634	666	4.6.1.1	Pulmonar	NA	77 años	H	
17	☒	G0114	0	10/04/2014	285	64	76	18	0,14	38	0,29	75	0,57	131	544	573	2.2.10	Pulmonar	Bronchial suction sample	37 años	M	
	☐	G0679	0	20/01/2015											662	695	2.2.10	Pulmonar	Sputum	49 años	M	
18	☒	G0076	1	17/01/2014	832	39	44	11	0,06	92	0,49	83	0,45	186	567	593	4.3.2	Pulmonar	Sputum	27 años	M	Cohabitants
	☐	G1965	1	28/04/2016											567	586	4.3.2	Pulmonar	Gastric secretion	47 años	M	Cohabitants
19	☒	G0837	0	24/12/2014	27	49	49	5	0,04	37	0,32	74	0,64	116	653	687	4.1.2.1	Pulmonar	Bronchoalveolar lavage	49 años	H	
	☐	G0932	0	20/01/2015											751	792	4.1.2.1	Pulmonar	Bronchial suction sample	36 años	H	
20	☒	G0146	1	26/04/2013	743	13	14	21	0,22	8	0,09	65	0,69	94	833	876	4.3.3	Pulmonar	Sputum	38 años	H	
	☐	G1141	1	9/05/2015											591	620	4.3.3	Pulmonar	Sputum	47 años	H	
21	☒	G1180	0	10/07/2015	22	23	24	113	0,70	11	0,07	38	0,23	162	690	714	4.3.2	Pulmonar	Bronchoalveolar lavage	45 años	M	Father and child
	☐	G1181	0	1/08/2015											495	506	4.3.2	Pulmonar	Sputum	21 años	H	Father and child

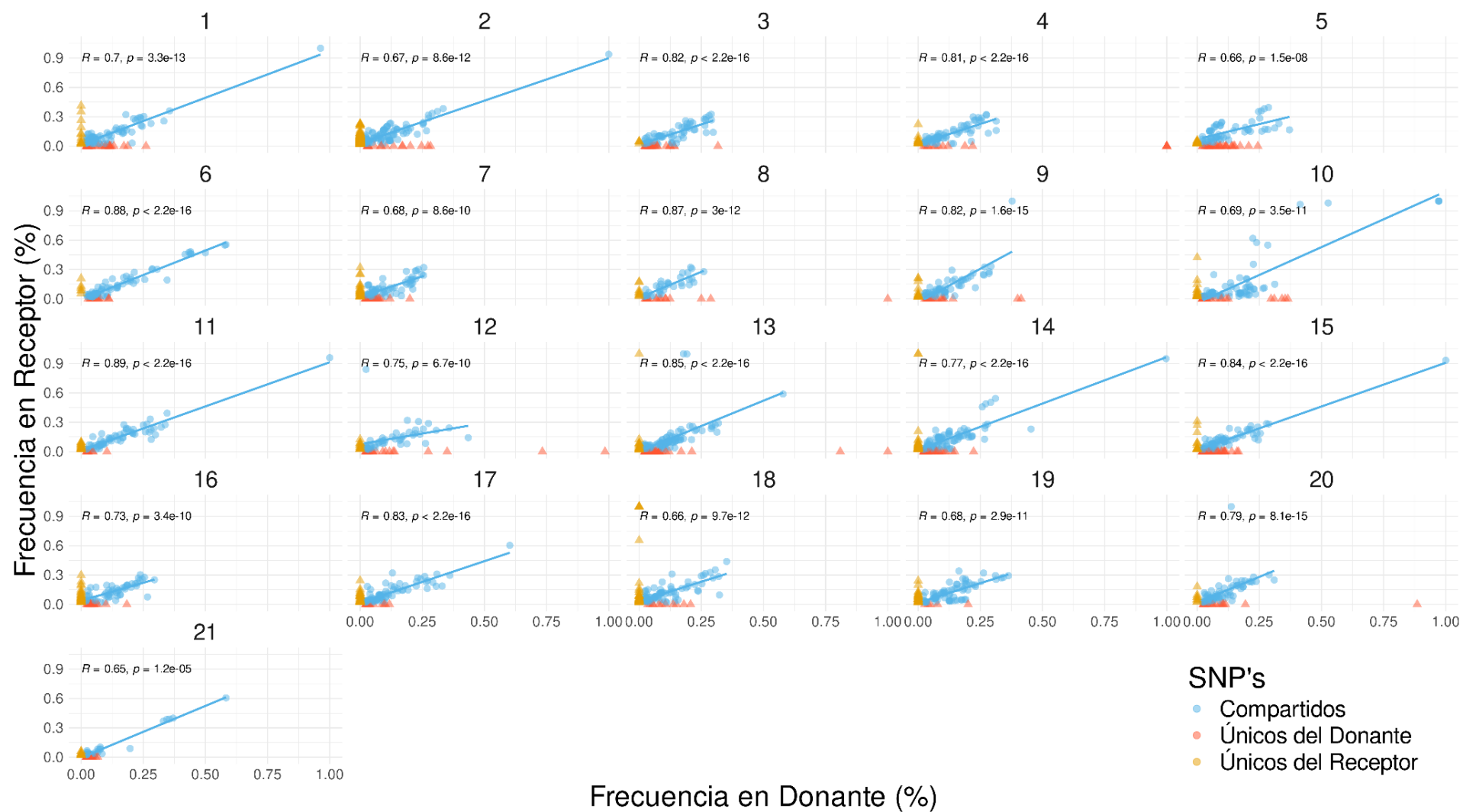


Figura suplementaria 5.1. SNPs compartidos y únicos en los 21 pares de transmisión (0.02-0.98) Calculamos la correlación entre los SNPs de los pares. (Correlación de Spearman, intervalo 0.66-0.89, valor $p = 1.2 \text{ e-}05$ a $2.2 \text{ e-}16$).

Anexo uno: Modelo de regresión logística

Evaluación y Resultados del Modelo de Regresión Logística

```
[1] "Tabla de Comparación de Predicciones y Valores Reales"
```

```
ajuste.train.pred
```

```
0_Surv 1_Surv
```

```
No      70      24
```

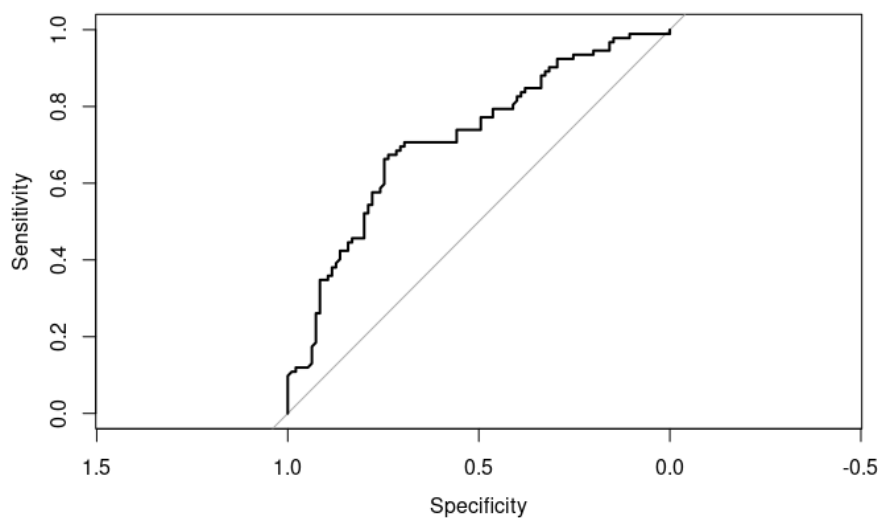
```
Yes     28     61
```

```
[1] "Curva ROC y Área bajo la Curva (AUC) para el Modelo de Regresión Logística"
```

```
Setting levels: control = No, case = Yes
```

```
Setting direction: controls < cases
```

```
Area under the curve: 0.7168
```



Evaluación de Modelos de Regresión Logística: Prueba Hosmer-Lemeshow y Validación Cruzada

```
[1] "Prueba de bondad de ajuste de Hosmer-Lemeshow para el Modelo de Regresión Logística"
```

```
Hosmer and Lemeshow goodness of fit (GOF) test
```

```
data: train1$inCluster_10, fitted(Reg_Model3)
```

```
X-squared = 13.083, df = 8, p-value = 0.109
```

```
[1] "Resultados de la Validación Cruzada para el Modelo de Regresión Logística (cv.Modelo3)"
```

```
[1] 0.2208100 0.2208037
```

Análisis y Resumen del Modelo de Regresión Logística (Reg_Model3)

[1] "Resumen del Modelo de Regresión Logística (Reg_Model3)"

Call:

```
glm(formula = inCluster_10 ~ N°.Edad + Sexo + Extranjero + Tinción +
    Localización + Alcoholismo + Prisiones + Sexo:Localización +
    Tinción:Localización + Extranjero:Alcoholismo + Sexo:Prisiones +
    Extranjero:Prisiones + Alcoholismo:Prisiones + Extranjero:Alcoholismo:Prisiones,
    family = binomial(link = logit), data = train1)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.1187	-0.9928	-0.4935	1.0050	2.3065

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.359510	0.400304	3.396	0.000683	***
N°.Edad	-0.033747	0.004766	-7.081	1.43e-12	***
SexoMUJER	0.501110	0.398540	1.257	0.208622	
ExtranjeroSi	-1.415841	0.209763	-6.750	1.48e-11	***
TinciónPositivo	-0.602536	0.457138	-1.318	0.187483	
LocalizaciónPulmonar	0.790659	0.328373	2.408	0.016049	*
AlcoholismoSi	0.365680	0.267581	1.367	0.171747	
PrisionesSi	0.165992	0.562059	0.295	0.767743	
SexoMUJER:LocalizaciónPulmonar	-0.733907	0.438677	-1.673	0.094327	.
TinciónPositivo:LocalizaciónPulmonar	0.558835	0.494305	1.131	0.258246	
ExtranjeroSi:AlcoholismoSi	0.179878	0.421830	0.426	0.669800	
SexoMUJER:PrisionesSi	-2.092820	1.448007	-1.445	0.148371	
ExtranjeroSi:PrisionesSi	2.391160	1.260967	1.896	0.057922	.
AlcoholismoSi:PrisionesSi	1.282995	1.213844	1.057	0.290526	
ExtranjeroSi:AlcoholismoSi:PrisionesSi	-4.589442	1.903518	-2.411	0.015907	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1039.53 on 749 degrees of freedom
 Residual deviance: 915.63 on 735 degrees of freedom
 AIC: 945.63

Number of Fisher Scoring iterations: 4