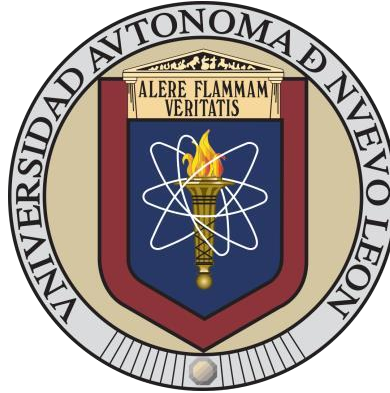


UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS BIOLÓGICAS



EXPRESIÓN DIFERENCIAL DE FACTORES DE
VIRULENCIA POR ANÁLISIS TRANSCRIPTÓMICO EN
Mycobacterium tuberculosis BEIJING CLÁSICO 391 Y
TIPO BEIJING 323.

POR:

DAVID HUMBERTO SUÁREZ GARZA

COMO REQUISITO PARA OBTENER EL GRADO DE
MAESTRO EN CIENCIAS CON ORIENTACIÓN EN
INMUNOBIOLOGÍA

SAN NICOLÁS DE LOS GARZA, N.L., 2025

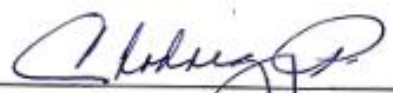
EXPRESIÓN DIFERENCIAL DE FACTORES DE VIRULENCIA
POR ANÁLISIS TRANSCRIPTÓMICO EN *Mycobacterium*
tuberculosis BEIJING CLÁSICO 391 Y TIPO BEIJING 323.

Comité de Tesis



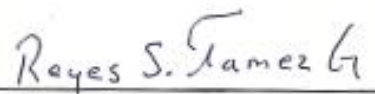
Dra. Lydia Guadalupe Rivera Morales

Presidenta



Dra. María Cristina Rodríguez Padilla

Secretario



Dr. Reyes Silvestre Tamez Guerra

Vocal 1



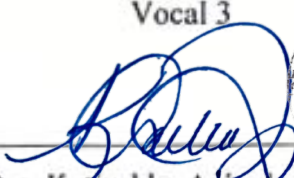
Dr. Patricio Adrián Zapata Morín

Vocal 2



Dr. Jorge Enrique Castro Garza

Vocal 3



Dra. Katiushka Arévalo Niño

Subdirectora de posgrado

DIRECCIÓN
DE POSGRADO

EXPRESIÓN DIFERENCIAL DE FACTORES DE VIRULENCIA
POR ANÁLISIS TRANSCRIPTÓMICO EN *Mycobacterium*
tuberculosis BEIJING CLÁSICO 391 Y TIPO BEIJING 323.

Dirección de Tesis



Dra. Lydia Guadalupe Rivera Morales

Director



Dr. Rogelio E. Hernández Pando
Asesor Externo

DERECHOS RESERVADOS©

PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta Tesis está protegido, el uso de imágenes, fragmentos de videos, y demás material contenido que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde se obtuvo mencionando al autor o autores

Agradecimientos

Doy gracias al Conacyt actualmente Secihti por concederme el apoyo financiero otorgado a través de becas destinadas a investigación. Sin el apoyo financiero, la destinación de recursos humanos orientados al desarrollo de tecnología en análisis del código genético y el sostener un ambiente apropiado para el aprendizaje y capacitación en inmunobiología no habría tenido lugar.

Doy gracias al Dr. Patricio Adrián Zapata Morín por haberme tratado con respeto, aprecio y haberme guiado a través de los avances en la realización de mi tesis. Particularmente fue de bastante dicha para mi como celebró el éxito en comunicar la primera forma de presentación oral pública de mi tesis. Esos sentimientos forman parte de lo mejor de uno mismo.

Agradezco también a la Dra. Lydia Guadalupe Rivera Morales, simplemente no es una persona que falla en tratar cálidamente a las personas, lleva consigo su moral y vela por mantener la cercanía de sus estudiantes organizando los convivios y esfuerzos necesarios para cumplir con cada evento de divulgación científica y de búsqueda de patrocinios. También, no es sino a través de su trayectoria profesional que se establecen los lazos de colaboración científica necesarios para que las investigaciones existentes tengan lugar, ciertamente agradezco haber tenido el privilegio de haber realizado un proyecto de naturaleza in silico y de amplio potencial en trascendencia como lo es la transcriptómica de cepas agresivas tuberculosas.

A mis compañeros de laboratorio y de generación les agradezco el haber sido un grupo de personas con intenciones claras y amigables. Muchos momentos fueron disfrutados. Particularmente Ana Espinosa y Gustavo Garza, gracias por la cantidad infinita de risas y conversaciones que sostuvimos, los aprecio bastante y eso no cambiará.

A mi futura esposa, Angela González, gracias por ser un motivante asombroso para motivar el que hiciera los esfuerzos en tiempo y forma. No solo académicos. De no haber pasado los días contigo me parece imposible que hubiera sostenido un balance saludable entre el esfuerzo académico y vida personal.

A mi familia, gracias por el apoyo y por la fe espero honrar sus esfuerzos. Les amo.

Dedicatoria

A Dios, al romance que existe en prestar atención a su creación y al potencial. Que no tarde en materializarse.

ÍNDICE

1.	RESUMEN.....	11
2.	INTRODUCCIÓN	13-14
3.	ANTECEDENTES.....	15-30
3.1.	MICOBACTERIAS.....	15
3.2.	MORFOLOGÍA BACTERIANA E IDENTIFICACIÓN.....	15-16
3.3.	TRATAMIENTO TRADICIONAL	16-17
3.4.	PATOGENIA	18-19
3.5.	INTERACCIONES ENTRE EL SISTEMA INMUNE Y LA MICOBACTERIA	20-23
3.6.	METABOLISMO DE <i>MYCOBACTERIUM TUBERCULOSIS</i>	24-25
3.7.	FACTORES DE VIRULENCIA Y ADAPTABILIDAD.....	26-28
3.8.	TRANSCRIPTÓMICA DE <i>MYCOBACTERIUM TUBERCULOSIS</i>	28-30
4.	JUSTIFICACIÓN	30-31
5.	HIPÓTESIS	31
6.	OBJETIVO DEL TRABAJO	31
7.	MATERIAL Y MÉTODOS	32-40
7.1.	BIBLIOTECAS DE SECUENCIADO DE ÁCIDOS RIBONUCLEICOS. COMPOSICIÓN Y CALIDAD	32-33
7.2.	TRANSFORMACIÓN DE ARCHIVOS DESDE FASTQ A .BAM	33-35
7.3.	TABLA DE CONTEO	36
7.4.	NORMALIZACIÓN.....	37
7.5.	PROCESADO DE DATOS	38-39
7.6.	AJUSTADO DE VALORES DE EXPRESIÓN PROMEDIO PARA VISUALIZACIÓN.....	40
8.	RESULTADOS	40-64
9.	DISCUSIÓN	64-69
10.	CONCLUSIÓN	69-70
11.	PERSPECTIVAS	70
12.	BIBLIOGRAFÍA.....	70-71

ÍNDICE DE TABLAS

Tabla	Página
Tabla 1: Inicio de clasificado de los receptores reconocedores de patrones, las moléculas que los estimulan, proceso que desemboca de la estimulación del PRR y si es o no protectorio.	22
Tabla 2: Componentes producidos por la micobacteria con propiedades inmunomoduladoras.	23
Tabla 3. Genes metabólicos micobacterianos asociados a condiciones de estrés.	25
Tabla 4: Principales factores de virulencia de <i>Mycobacterium tuberculosis</i> .	27
Tabla 5: Resumen de alineamientos de archivos por BC391 y BL323 al día 14.	42
Tabla 6: Resumen de alineamientos de archivos por BC391 y BL323 al día 28.	44
Tabla 7: Los 20 genes de mayor expresión diferencial en BC391 al día 14.	49
Tabla 8: Los 20 genes de mayor expresión en BL323 al día 14.	50
Tabla 9: Los 20 genes de mayor expresión en BC391 al día 28 al ser comparado con BL323 al día 28.	51
Tabla 10: Los 20 genes de mayor expresión en BL323 al día 28 al ser comparado con BC391 al día 28.	52
Tabla 11: Descripción de genes metabólicos expresados diferencialmente por M. tuberculosis BC y BL al día 14 y 28.	57
Tabla 12: Genes relacionados con transcripción y traducción.	60
Tabla 13: Genes sobreexpresados en BC391 y BL323 relacionados a estrés.	63

ÍNDICE DE FIGURAS

Figura	Página
Figura 1: Dianas de los tratamientos contra la tuberculosis.	17
Figura 2: Mapa metabólico de <i>M. tuberculosis</i> (Chang y Guan 2021).	20
Figura 3: Factores de virulencia de las micobacterias.	24
Figura 4. Generación de un índice de genoma de referencia y contenedores de archivos de secuenciado.	27
Figura 5. Alineamiento de archivos de secuenciado al genoma de referencia.	34
Figura 6. Conteo de transcritos.	35
Figura 7. Generación de metadatos.	36
Figura 8. Matriz para análisis transcriptómico y normalizado.	36
Figura 9. Uso de funciones para realizar el análisis transcriptómico y generar el gráfico de volcán.	38
Figura 10. Selección de genes de expresión diferencial.	38
Figura 11. Código de asignación de clasificación a los genes.	39
Figura 12. Gráfico de frecuencias de grupos de genes.	39
Figura 13: Gráficos de control de calidad demostrados.	40
Figura 14: Expresión diferencial de BC y BL al día 14 y 28.	46
Figura 15: Gráficos de pie de grupos de proteínas en genes sobreexpresados en BC391 y BL323 al día 14.	53
Figura 16: Gráficos de pie de grupos de proteínas en genes sobreexpresados en BC391 y BL323 al día 14.	53
Figura 17: Frecuencia de grupos de proteínas o genes por cepa y día.	54
Figura 18: Comparación de los promedios geométricos en logaritmo de los 20 genes de mayor expresión relativa para BC391 y BL323 al día 14 y día 28 posterior a remoción de genes hipotéticos.	55
Figura 19: Gráfico de comparación del eje metabólico en genes sobreexpresados por las cepas BC391 y BL323 al día 14 y 28 .	56

Figura 20: Gráfico de comparación del eje transcripción-traducción en genes sobreexpresados por las cepas BC391 y BL323 al día 14 y 28.

Figura 21: Gráfico de comparación del eje adaptabilidad y estrés en genes sobreexpresados por las cepas BC391 y BL323 al día 14 y 28.

LISTA DE SÍMBOLOS Y ABREVIATURAS

Mtb	<i>Mycobacterium tuberculosis</i>
MTBC	Complejo de <i>Mycobacterium tuberculosis</i>
PPR	Receptores reconocedores de patrones
PAMP	Patrones Moleculares Asociados a Patógenos
BC	Beijing Clásica
BL	Tipo Beijing
d	Día

I. RESUMEN

La tuberculosis, ocasionada por *Mycobacterium tuberculosis* (Mtb), es una enfermedad de 2 fases de infección (latente y activa) que se encuentra entre las principales causas de mortalidad e impacto detrimental socioeconómico a nivel mundial. Si bien la posibilidad de pasar de infección latente a activa es reducida, factores agravantes como comorbilidades, edad avanzada y cepas más virulentas alteran el pronóstico de la enfermedad. Para enfrentar la tuberculosis es necesario continuar la caracterización molecular de las cepas agresivas y entender cómo su perfil de expresión incide sobre la patogénesis. El análisis de la expresión genética diferencial entre las cepas con menor o mayor virulencia es útil para identificar potenciales dianas terapéuticas contra las variantes de cepas tuberculosas y para entender qué moléculas son las responsables del aumento de la virulencia.

El objetivo del presente proyecto de investigación es explorar las diferencias en expresión de genes entre la cepa tipo Beijing 323 (BL) y beijing clásica 391 (BC) mediante un análisis transcriptómico de archivos digitales provenientes de secuenciación de ácidos ribonucleicos en un modelo de infección murino.

Se encontraron los 20 genes de mayor expresión diferencial para las cepas BC y BL al día 14 y 28, 80 genes en total (20 por cada cepa en cada día). Los genes sobreexpresados fueron clasificados en a) metabólicos; b) relacionados a transcripción o traducción; y c) de respuesta a estrés. Al día 14 se encontraron diferencias en maquillaje celular externo donde BC sobre expresó Ag85 y policétidos mientras que en BL hay trehalosa, un componente inmunoestimulante no antigénico y acumulación de triacilglicerolos. BL al día 14 tiene mayor adaptación a hipoxia por generación de energía adicional mediante la entrada de metilcitrato al ciclo de krebs reductor y actividad de secuestro de metales hierro y molibdeno los cuales son útiles para generar regulación transcripcional por whiB. Se observó un rezago en la actividad transcripcional, BC al día 28 expresa el mismo set de genes que BL al día 14. BL al día 28 sobreexpresa genes metabólicos y ESAT-6, una proteína asociada a disrupción de membranas, promoción de necrosis tisular y liberación a un ambiente no restrictivo o exterior al granuloma señalando la generación de tuberculosis activa.

I. ABSTRACT

Tuberculosis, caused by *Mycobacterium tuberculosis* (Mtb), is a disease with two phases of infection (latent and active) and is among the leading causes of mortality and socioeconomic burden worldwide. Although the likelihood of progressing from latent to active infection is low, aggravating factors such as comorbidities, advanced age, and highly aggressive strains significantly affect disease prognosis. To combat tuberculosis, it is critical to continue advancing the molecular characterization of aggressive strains and understanding how their expression profiles influence pathogenesis. Identifying genes that are upregulated—or exclusively expressed—in aggressive strains compared with less aggressive strains (differential expression) is valuable for discovering potential therapeutic targets and understanding the molecular basis for increased virulence.

The objective of this research project is to explore differences in gene overexpression between the Beijing-type strains (BL) and the classical Beijing strains (BC), based on transcriptomic analyses of digital files derived from RNA sequencing in a murine infection model.

The top 20 genes with the highest differential expression were identified for strains BC and BL on days 14 and 28, 80 genes in total. The overexpressed genes were classified as metabolic, related to transcription or translation, and stress-response genes. At day 14, differences were observed in the composition of the external cellular envelope: BC overexpressed Ag85 and polyketide genes, whereas BL showed increased trehalose—a less immunostimulatory component—and triacylglycerol accumulation. At the same time point, BL exhibited greater adaptation to hypoxia through the generation of additional energy via the methylcitrate pathway and increased activity of iron and molybdenum sequestration, supporting transcriptional regulation by WhiB. A delay in transcriptional activity was observed, as BC at day 28 expressed the same set of genes as BL at day 14. By day 28, BL overexpressed metabolic genes and ESAT-6, a protein associated with membrane disruption, promotion of tissue necrosis, and release into a less restrictive or extragranulomatous environment, marking the transition toward active tuberculosis.

II. INTRODUCCIÓN

La tuberculosis, causada por *Mycobacterium tuberculosis* (Mtb), es la enfermedad infecciosa líder en impacto socioeconómico a nivel mundial. En los pasados 200 años, se le atribuyen más de 1 billón de defunciones a las cuales se añaden casi 2 millones de defunciones anuales y 10 millones de contagios. En adición, se estima que entre la cuarta y tercera parte de la humanidad se encuentran infectados en fase latente (Bussi y Gutierrez 2019; Chandra y Grigsby 2022; Ferraris et al. 2018; Gagneux 2018).

La tuberculosis se caracteriza por contar con 2 fases de infección: latente y activa. En la fase activa se genera la sintomatología y diseminación de la bacteria a nuevos hospedadores, en la fase latente por otra parte hay ausencia de síntomas y no se es infectivo. Afortunadamente, solo de un 5 a 15% de las personas contagiadas desarrollan tuberculosis activa en los primeros 5 años de infección (Kiazzyk y Ball 2017). Factores agravantes como la diseminación de cepas virulentas, resistencia a fármacos y presencia de comorbilidades en la sociedad como VIH u obesidad facilitan la evolución de la enfermedad tuberculosa latente a infección activa (Getahun et al. 2015).

Mtb ha coevolucionado con el humano desde hace 70000 años (Gagneux 2018), por lo que se encuentra muy bien adaptada para sobrevivir al sistema inmune humano con un amplio arsenal de genes orientados a irrumpir mecanismos protectores o de erradicación de patógenos por parte del hospedante. Algunos de los factores de virulencia más conocidos de Mtb son proteínas transportadoras de minerales y azúcares, enzimas metabolizadoras y sintetizadoras de lípidos, proteínas secretoras, proteínas secretadas y proteínas de modulación genética (Al-Asady y Ali 2023).

Aunque las variantes de Mtb sean altamente parecidas con una similaridad genética mayor al 99.9%, existen diferencias sustanciales en los perfiles de expresión génica o programas transcripcionales. A estas distinciones en actividad de expresión génica se le adjudican cambios en el grado de virulencia reportado en modelos murinos ((Cerezo et al. 2021)).

En un modelo murino, BL provocó la defunción al 100% a los 45 días comparado con la cepa BC produjo el 100% de muerte a los 60 días (Cerezo et al. 2021). La variabilidad en la virulencia entre las cepas puede ser examinada mediante la integración de la información proporcionada por los avances tecnológicos, como el análisis transcriptómico. El cual es un estudio integral de todos los transcritos de RNA presentes en una célula en un momento dado y en condiciones específicas. Este estudio involucra la extracción del RNA, su conversión a cDNA y posterior secuenciación,

para finalmente realizar un análisis bioinformático. Este tipo de estudio, en Mtb, se lleva a cabo para determinar las diferencias en expresión entre cepas con diferente virulencia. Lo cual posibilita la interpretación de las actividades biológicas del patógeno para infectar y sobrevivir, así como las desarrolladas por el hospedante en respuesta a la infección. De esta manera se logran avances significativos en el entendimiento de la enfermedad y el correspondiente potencial para el control de la enfermedad.

Los estudios basados en secuenciación de ácidos ribonucleicos (ARN) resuelven preguntas de ciencia fundamental sobre el metabolismo del patógeno para explorar el potencial de generar terapias como bloqueos génicos de componentes claves para el ciclo patogénico de la bacteria, administración de anticuerpos contra receptores inmunes que no generen una respuesta protectora contra la bacteria, producción de fármacos o modulación metabólica endógena que favorezca al hospedante en el enfrentamiento contra la bacteria.

El objetivo del presente estudio es, basado en la información recabada en los trabajos previos publicados (Cerezo et al. 2021; Cerezo-Cortés et al. 2022), realizar un análisis de la expresión diferencial entre Mtb BC y BL en un modelo murino infectado. El enfoque principal de este estudio es determinar la expresión diferencial de los factores de virulencia entre las dos cepas para dilucidar los factores de virulencia que producen una mortalidad acelerada por la cepa tipo Beijing 323.

III. ANTECEDENTES

3.1 Micobacterias

El género *Mycobacterium* contiene más de 170 especies de micobacterias. Las que se dividen en micobacterias de lento o rápido crecimiento (tiempo de duplicación de 18-24h a diferencia de los 20-30 min de otras bacterias como *Escherichia coli*). Generalmente aquellas que son patógenos del ser humano son de crecimiento lento y pertenecen al complejo de micobacterias tuberculosas (MTBC), *Mycobacterium ulcerans* o *Mycobacterium leprae*. El resto de micobacterias no tuberculosas suelen generar enfermedad en individuos inmunocomprometidos.

El complejo de Mtb incluye a las especies genéticamente cercanas: *M. tuberculosis*, *M. -canettii*, *M. -africanum*, *M. -microti*, *M. -bovis*, *M. caprae* y *M. pinnipedii*. Del complejo, Mtb es la responsable de afectar a alrededor de un 33% de la población global de forma latente y generar amplios estragos socioeconómicos (Gagneux 2018).

3.2 Identificación Bacteriana

Mtb es un bacilo aerobio, grampositivo, no esporulante, resistente tanto a ácidos como a alcoholes (Al-Asady y Ali 2023).

Para identificar *in vitro* la presencia de micobacterias se realiza un cultivo bacteriano a partir de esputo en medio Lowenstein Jensen o Middlebrook (tiempo estimado de 6 a 8 semanas). Se continúa con una tinción Zhiels Neelsen y baciloscopía. Además se emplean sondas específicas para ácidos nucleicos de Mtb y la prueba de inhibición de crecimiento específica para MTB conocida como NAP (siglas en inglés de prueba de crecimiento en Ácido Para-Nitrobenzóico). También se puede corroborar por pruebas bioquímicas mediante reacción a niacina, reducción de nitratos y catalasa. Si se busca reconocer si hay previa exposición a micobacterias *in vivo* se utiliza la prueba de Manteaux y radiografía pectoral (Talbot y Raffa 2015).

Por otra parte, posterior a validar la presencia de micobacterias, la identificación de las familias micobacterianas y cepas se logra mediante la técnica de espoligotipificación. Una técnica que consiste en la amplificación de 43 secuencias únicas en la región de repetidos directos (DR) presente en las micobacterias. Actualmente los resultados se traducen a un formato digital que devuelve un código de reconocimiento SIT (siglas en inglés de tipos internacionales de espoligotipificación) así como el nombre de la familia a la cual pertenecen los valores de la espoligotipificación (Bakula et al. 2023).

3.3 Tratamiento

La tuberculosis es una enfermedad tratable. Sin embargo, es un tratamiento largo (al menos 6 meses) y la emergencia de cepas resistentes compromete la eficiencia del tratamiento. El tratamiento estándar es de una fase de 2 meses con pirazinamida, isoniazida y rifampicina, seguida por 4 meses más con rifampicina e isoniazida. En algunos casos se añade etambutol como el cuarto fármaco de primera línea. (Horsburgh et al. 2015).

Medicamentos de nueva generación son una alternativa para el tratamiento de tuberculosis causada por cepas resistentes a fármacos. Entre ellos están la bedaquilina (una diarylquinolina que bloquea la síntesis de ATP) y la delamanida (derivado del metronidazol) que han mostrado resultados prometedores (Murray et al. 2015).

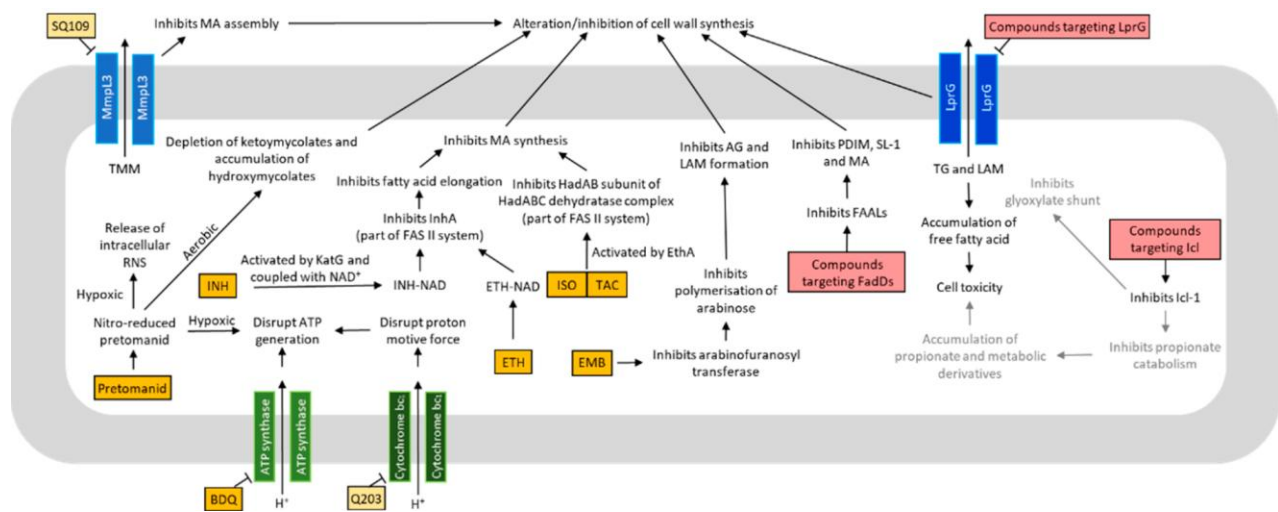


Figura 1: Dianas de los tratamientos contra la tuberculosis. En naranja los fármacos clínicamente aprobados, en amarillo aquellos en proceso de aprobación clínica y en rojo los compuestos en pruebas in vitro (Chang y Guan 2021).

Los fármacos más populares etambutol e isoniazida evitan la producción de componentes de la pared celular como ácidos micólicos y compuestos con arabinosa aumentando la permeabilidad celular. La rifampicina por otra parte es un compuesto que afecta la polimerasa de rna rpoB disminuyendo la transcripción. Algunos fármacos en desarrollo y moléculas pequeñas son capaces de atacar directamente al metabolismo lipídico forzando a la bacteria a permanecer más activa y menos persistente. El itaconato, producido endógenamente, inhibe la enzima isocitrato liasa inhibiendo el procesamiento de ácidos grasos bajo condiciones de estrés hipóxico y SQ10 inhibidor del fijador de trehalosa también puede ser útil para sensibilizar a la bacteria

contra antibióticos debido a que la trehalosa es un azúcar útil como fuente de carbono de reserva y componente de la pared celular asociado a reducciones de permeabilidad y farmacorresistencia (Chang y Guan 2021).

3.4 Patogenia

Aunque la patogenia puede variar entre cepas de Mtb, la patogenia convencional puede ser resumida en 5 pasos:

1. Entrada de Mtb al hospedante en partículas vía nasal u oral.
2. Fagocitosis por células del sistema inmune innato.
3. Liberación al citosol.
4. Aumento de carga bacilar
5. Diseminación y contagio.

3.4.1 Entrada de Mtb en partículas vía nasal u oral.

Pequeñas gotas o aerosoles con carga bacilar de Mtb provenientes de una persona con tuberculosis activa entran vía oral o nasal en un segundo ahora nuevo hospedante (Ferraris et al. 2018). Posterior a ser inhalada, Mtb recorre las vías respiratorias e interactúa con receptores reconocedores de patrones (PRR) en las células del sistema inmune innato (macrófagos, células dendríticas y neutrófilos). Algunos de los receptores de interacción con Mtb son DC-SIGN, Dectin-2 y receptor de manosa; estos 3 interactúan con lipoarabinomano, un componente de la pared celular de Mtb. Mincle y Marco son otros dos PRR que interactúan con el componente dimicolato de trehalosa. El receptor carroñero tipo A y los receptores del complemento también sostienen interacción con Mtb (Ravesloot-Chávez et al. 2021). Aunque sería beneficioso explorar las vías metabólicas exactas estimuladas en el hospedante a partir de las interacciones con Mtb mediante los PRR para fines de integrar una comprensión sistémica de la patogenia el presente escrito no lo aborda.

3.4.2. Fagocitosis por células del sistema inmune innato.

Posterior a la entrada e interacción con PRRs, Mtb es fagocitada por macrófagos residentes de pulmón y neutrófilos. Esta acción induce la producción de las citocinas proinflamatorias interleucina 1 (IL1), factor tumoral necrótico alfa (TNF α) e interferón γ (IFN γ). La IL1 e IFN γ promueven el reclutamiento de monocitos proveyendo de dianas de infección a la micobacteria, lo que contribuye a la replicación y diseminación de la bacteria (Ravesloot-Chávez et al. 2021).

3.4.3. Liberación al citosol.

Las micobacterias fagocitadas tienen la capacidad de sobrevivir y desestabilizar a los autofagolisosomas. Los genes ESX codifican para unos complejos proteicos que se insertan en la membrana autofagolisosomal irrumpiendo la continuidad de la membrana y liberando componentes de membrana, material genético y bacterias completas al citosol. Una vez en el citosol, el material genético y bacterias son procesadas por las vías cGAS-STING (detecta DNA citosólico), LC3-II y AIM2. cGAS-STING produce IFN γ y activa la autofagia selectiva LC3-II que lleva bacterias completas etiquetadas con ubiquitina directo a lisosomas. Simultáneamente, AIM2 desemboca en la activación de inflammasoma (NLRP3) y activa caspasa 1 generando un corte activador sobre IL1B e IL-18. Por este medio también se produce la citocina pronunciante de actividad fagocítica IFN γ (Chandra y Grigsby 2022; Sayes et al. 2016). Por otra parte, Mtb secreta una serie de proteínas fosfatasas que inhiben procesos celulares (codificadas en los genes PtpA y PtpB) y otras se translocan al núcleo del hospedante para ejercer un rol inmunomodulador mediante la metilación o acetilación del ADN. Las proteínas codificadas en los genes Rv1988, y Rv2966c, con actividad de histona metiltransferasa, son un par de ejemplos caracterizados por alterar la expresión genética (Gauba et al. 2021).

3.4.4. Formación del granuloma y aumento de carga bacilar

La inflamación generada por la interacción con los PRR induce el reclutamiento de células inmunes. La acumulación de células inmunes en el sitio de infección se organiza para formar los granulomas. Los granulomas limitan el acceso al oxígeno necesario para sobrellevar un ritmo metabólico elevado por lo que ocasionan una disminución del ciclo de krebs y un aumento del metabolismo lipídico de Mtb. El metabolismo lipídico de Mtb induce la polarización de los macrófagos a macrófagos espumosos facilitando el suministro de ácidos grasos para su aprovechamiento por las micobacterias. La expresión de factores de crecimiento incentiva eventos de angiogénesis que permiten la entrada ocasional de oxígeno favoreciendo la supervivencia y crecimiento bacteriano hasta la formación de granulomas necróticos incapaces de contener la diseminación de la micobacteria (Bussi y Gutierrez 2019; Chang y Guan 2021).

3.4.5. Diseminación y contagio

Mtb accede a los vasos linfáticos y se disemina ocasionando tuberculosis extrapulmonar. Asimismo, concentraciones bacterianas elevadas son expulsadas en forma de aerosol o a través de esputo (Kak et al. 2020).

3.5 Interacción entre el sistema inmune y la micobacteria

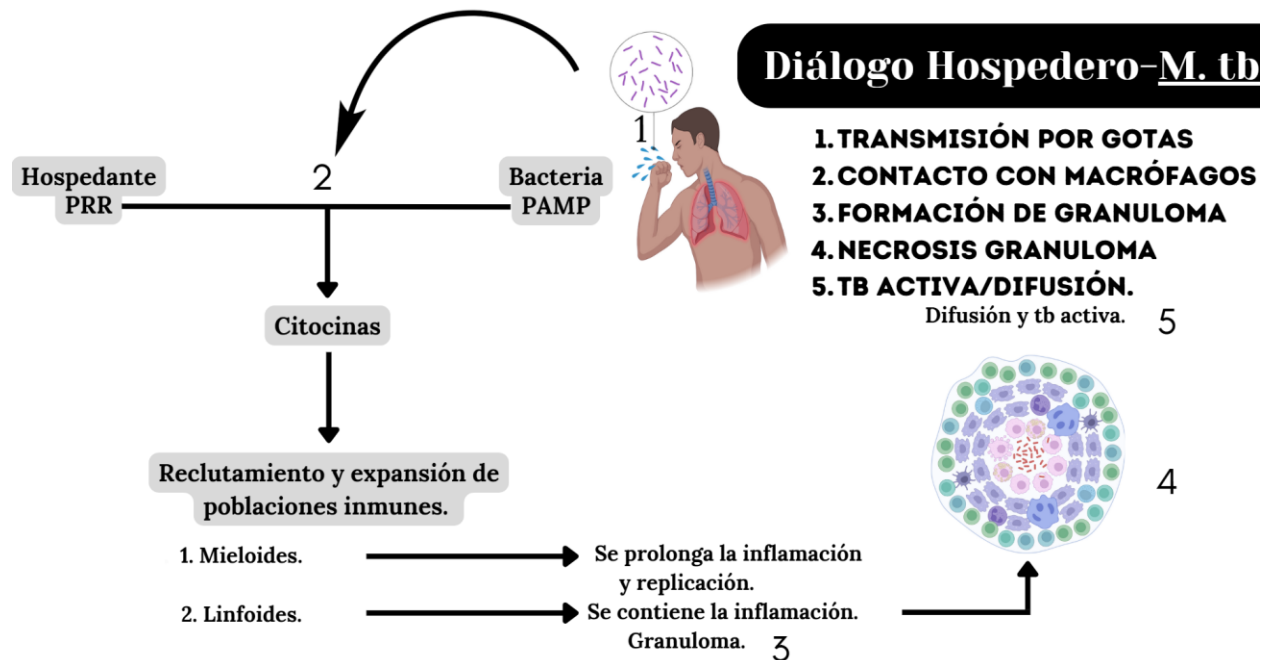


Figura 2: Generalidad de patogénesis. (Imagen generada a partir de la información en Dorhoi et al. 2011).

Gran parte del daño producido durante la infección tuberculosa se debe al proceso inflamatorio generado por la respuesta inmune ante *M. tuberculosis* y sus componentes. Estas interacciones están siendo estudiadas a nivel molecular, celular y tisular. La inducción de los diferentes tipos de respuesta ante la invasión del patógeno dependerá de las moléculas bacterianas y los diferentes receptores celulares involucrados.

Entre algunos de los componentes bacterianos que interactúan con las células están: Glicolípidos: Lipoarabinomanano (LAM), lipoarabinomanano manosilado (MANLAM), fosfatidil inositol manósido (PIM), micolil arabino galactano (MDP). Polisacáridos: arabinomanano (AM). Proteína glicosilada peptidoglicano (PID) (ref). Todas estas moléculas poseen patrones de reconocimiento por parte del sistema inmune. A estos se les conoce como patrones moleculares asociados a patógenos (PAMP) (ref). En tanto, las células inmunes poseen los receptores reconocedores de patrones (PPR) clasificados como receptores tipo toll (TLR), receptores tipo nod (NLR), receptores de lectina tipo c (CLR) y receptores carroñeros (SR) (Dorhoi et al. 2011).

Los PRRs y los PAMPs interactúan generando señalización celular que producen cambios transcripcionales, inducción de expresión de citocinas y modulación de eventos de reclutamiento y expansión de tipos celulares inmunes. Depende de las interacciones entre las moléculas del patógeno y los receptores celulares, la respuesta se dirigirá hacia el reclutamiento de células inmunes innatas que sirven como nichos replicativos de la bacteria y mantienen la infección o de células inmunes adaptativas que son capaces de contener la infección. Conocer los componentes de capas bacterianas y sus proporciones determinado por la expresión génica, así como las interacciones de dichos componentes, proveería información para entender su influencia en la virulencia de diferentes cepas a través de todo el proceso patogénico.

Tabla 1. Interacciones moleculares descritas entre PRRs de Mtb y PAMPs.

PRR	PAMP	Actividad	Efecto	Referencia
LAM, MAN, PIM	TLR2	Producción de catelicidinas, inflamación	Protector	Chang y Guan 2021
ADN intracelular	TLR9	Producción de interferón, inflamación	Protector	Chang y Guan 2021
MDP	NOD2	Formación de granulomas	Protector	Chang y Guan 2021
Ag85 Ácidos micólicos	TCR	Contención de granuloma mediado por respuesta Th1	Protector	Ehrt et al. 2018

ManLim, PIM, AM	LCR3 y MR	Fagocitosis de <i>M. tuberculosis</i> inflamación presentación antígenos	<i>M.</i> sin ni de	No protector	Chang y Guan 2021
	DC SIGN	Presentación antígeno, Disminuye inflamación, Activación macrófagos	de	Protección situacional	Chang y Guan 2021
moléculas con trehalosa	Mincle Dectin	Formación granulomas	de	Protección situacional	Ehrt et al. 2018

Los TLR, NLR y CLR aumentan la inflamación y expansión celular mediante las vías de factor nuclear Kappa B (NFkB), MAPK y producción de interferones. LCR3 y MR no se consideran como protectoras debido a que no generan inflamación mientras que permiten la entrada de Mtb. Además, ausencia de mincle, dectin y nod2 se relaciona con aumento de monocitos polimorfonucleares como los neutrófilos los cuales exacerban la inflamación (Dorhoi et al. 2011).

Tabla 2. Componentes producidos por Mtb con propiedades inmunomoduladoras.

Componente	Función
Policétidos (Ehrt et al. 2018).	Evasión de reclutamiento y reconocimiento inmune.
ESAT6 (Francis et al. 2014).	Necrosis por disrupción de membranas, interferencia en la estimulación de TCR.
LprG (Chang y Guan 2021).	Traslocación de lípidos de membrana.

3.6 Metabolismo de *Mycobacterium tuberculosis*

Las micobacterias cuentan con un metabolismo versátil, una amplia gama de enzimas que les confieren la capacidad de procesar distintos metabolitos como fuente de carbono. En condiciones de estrés como ambiente hipóxico o ácido, la micobacteria aumenta la síntesis de enzimas que producen componentes grasos entre ellos ácido micólico y trehalosa. Esto aumenta la gluconeogénesis mientras que disminuye la entrada de moléculas al ciclo de Krebs o ciclo del ácido tricarboxílico (TCA). Con la consiguiente disminución de su metabolismo. El aumento de producción de componentes grasos le asegura una fuente de carbono a largo plazo para mantener la producción de ATP mediante el ciclo TCA reductor y genera una pared celular más robusta a expensas de reducir el crecimiento bacteriano y permanecer como célula viable no cultivable. Se ha encontrado que las micobacterias resistentes a antibióticos cuentan con niveles elevados de ácidos grasos conjugados con trehalosa (Chang y Guan 2021).

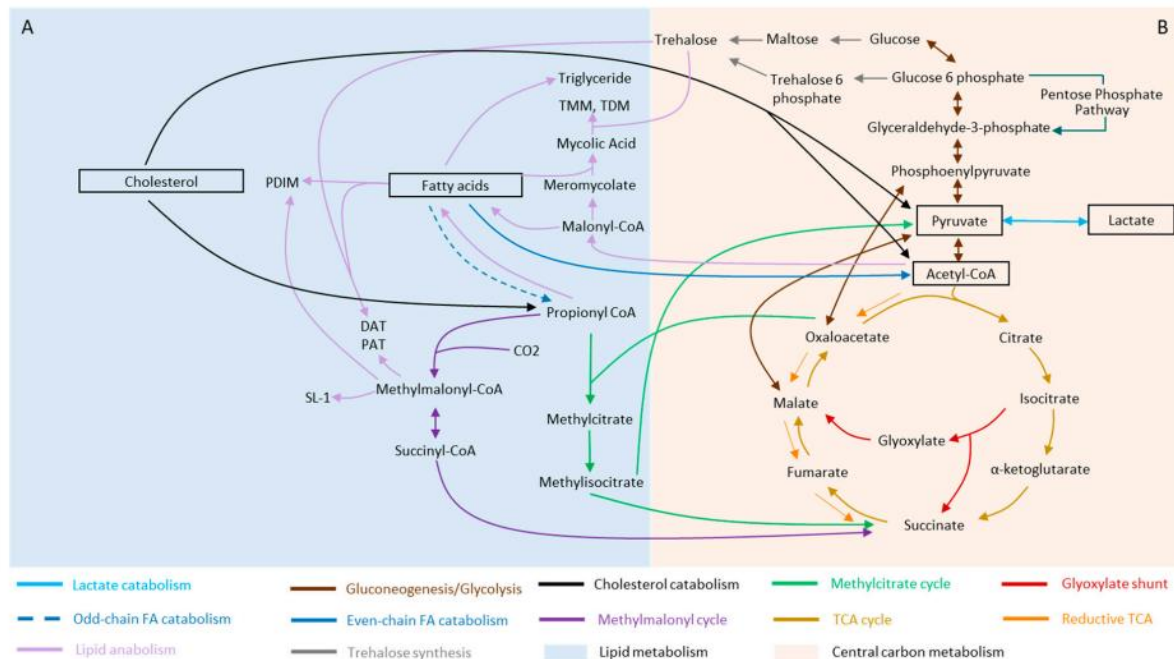


Figura 2: Mapa metabólico de *M. tuberculosis* (Chang y Guan 2021).

Tabla 3. Genes metabólicos de Mtb asociados a condiciones de estrés.

Gen	Proteína	Función
<i>DosR</i>	Regulón en condiciones de estrés	Regulación de expresión genética.
<i>whiB3</i>	WhiB3. Factor de transcripción sensible a redox.	Expresión de proteínas como factores de virulencia.
<i>pckA</i>	PEPCK. Fosfoenol piruvato carboxikinasa.	Promueve gluconeogénesis. Producción de fosfoenolpiruvato a partir de oxaloacetato.
<i>icl</i>	ICL. Isocitrato liasa.	Conversión de isocitrato a glioxilato y succinato.
<i>lprG</i>	LprG. Lipoproteína G.	Biosíntesis de la pared celular y transporte de lípidos. Modula la respuesta inmune
<i>lip</i>	Lip. Lipasas.	Degradación de lípidos para su uso.

A partir de ácidos grasos se produce acetil coA el cual puede continuar el TCA formando glioxilato por isocitrato liasa y producir oxaloacetato. La enzima PEPCK produce fosfoenolpiruvato a partir de oxaloacetato pronunciando la gluconeogénesis hasta la formación de trehalosas. Además, la isocitrato liasa, disminuye el uso de la cadena de electrones fermentando compuestos carbonados

formando ATP en las reacciones sin liberar dióxido de carbono ya que se salta los dos pasos del ciclo de Krebs que producen dióxido de carbono: formación de alfa ceto glutarato a partir de isocitrato y formación de succinato a partir de alfa ceto glutarato cabe destacar que el dióxido de carbono es una molécula que puede contribuir a generar inflamación mediante la ruta NFkB, evitar su producción puede ser un nivel más de evasión inmune. En breve ICL y PEPCK favorecen la producción de componentes de membrana, reducen los intermediarios del ciclo TCA y evitan la formación de dióxido de carbono favoreciendo la producción de ATP sin generar componentes inflamatorios (Chang y Guan 2021).

3.7 Factores de virulencia y adaptabilidad

El cromosoma de Mtb está compuesto de alrededor de 4000 genes. De estos, al menos 877 están asociados a procesos metabólicos o adaptabilidad y 91 genes están asociados a virulencia. Es decir, los genes de virulencia constituyen el 2.3% del genoma (Al-Asady y Ali 2023). Sin embargo, los genes relacionados al metabolismo y adaptabilidad son necesarios para la expresión de los genes de virulencia. En *M. tuberculosis* no hay moléculas como determinantes específicos de virulencia, sino que la combinación de los diferentes factores de virulencia y la respuesta del sistema inmune del hospedante ante la micobacteria son los responsables de generar la patología (Kroesen et al. 2019). Los factores de virulencia se definen como aquellos que facilitan la supervivencia de un patógeno en su hospedante (Al-Asady y Ali 2023). Siguiendo ese concepto, se estima que al menos 400 genes son necesarios para la persistencia de *M. tuberculosis*, incluyendo genes que codifican proteínas involucradas en rutas metabólicas, sistemas de secreción y otras proteínas que le confieren resistencia a condiciones adversas (Kroesen et al. 2019). Existen proteínas de *M. tuberculosis* con un hipotético rol como factor de virulencia y delimitarlo como tal va más allá del enfoque del presente proyecto por lo cual solo exploraremos el rol teórico reportado de las proteínas más expresadas.

Tabla 4: Principales factores de virulencia de *Mycobacterium tuberculosis*. Imagen extraída de (Al-Asady y Ali 2023).

1	Heat-shock protein- HspX, Alpha-crystallin.
2	Magnesium uptake- MgtC, Mg ²⁺ transport protein.
3	Secreted proteins- ESAT-6/CFP-10, antigen secretion targets.
4	Stress protein- AhpC, Alkyl hydroperoxide reductase C, KatG, Catalase-peroxidase, SodA, SodC, Superoxide dismutase.
5	Adherence- HbhA, Heparin-binding hemagglutinin.
6	Unclassified-PE/PE-PGRS, Pro-Glu (PE)/ PGRS (polymorphic GC-rich sequence)
7	Intracellular survival -Erp, Exported repetitive protein.
8	Metabolic adaptation- FadD33, long-chain-fatty-acid-ligase, Isocitrate lyase, LipF, Carboxylesterase LipF, Nitrate reductase, PanC/PanD, Pantothenate synthetase.
9	Cell wall- Antigen 85, LAM, lipoarabinomanan, MmaA4, Methoxy mycolic acid synthase 4, PDIM, phthiocerol dimycocerosates, PcaA, Cyclopropane mycolic acid synthase.
10	Secretion system- ESX-1,ESX-3, ESX-5.
11	Regulation-DevRS, DNA-binding transcriptional activator, HspR, heat shock protein transcriptional repressor, IdeR, Iron-dependent repressor, MprAB, MprA, Response regulator, PhoP, Possible transcriptional regulator, RelA, Bifunctional (p)ppGpp synthase/hydrolase, SigA, RNA polymerase sigma factor, SigE, ,SigF, SigH RNA sigma factor, WhiB3, Redox and pH-responsive transcriptional regulator.
12	Iron uptake -Mycobactin.
13	Toxin- Phospholipase C.

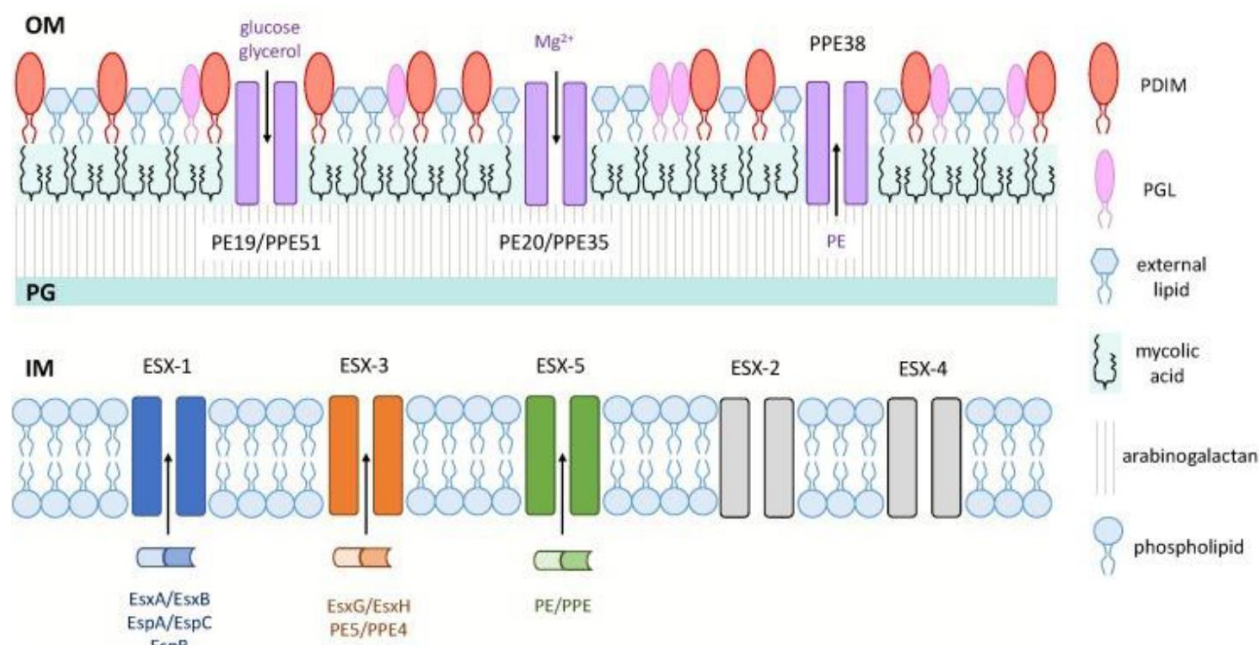


Figura 3: Factores de virulencia de las micobacterias. En la membrana exterior están presentes lipopolisacáridos como PDIM, ácidos micólicos, arabinogalactano, enzimas quelantes de fuentes de carbono como glicerol o glucosa y sideróforos o quelantes de iones. En la membrana interna

están sistemas de secreción de proteínas y proteínas secretadas. Imagen extraída de (Al-Asady y Ali 2023).

3.8 Transcriptómica

La transcriptómica es el estudio del conjunto completo de ARN mensajeros (ARNm) expresados por un organismo bajo determinadas condiciones. A través de la secuenciación de ARN total (RNA-seq), es posible caracterizar en detalle la expresión genética, permitiendo analizar tanto respuestas fisiológicas como mecanismos de patogénesis. Esta tecnología permite identificar los genes que se encuentran activos en un momento específico, lo cual es esencial para comprender los procesos biológicos y las interacciones moleculares entre un patógeno y su hospedador, así como para construir modelos de inferencia sobre dichas interacciones.

En el caso de *Mycobacterium tuberculosis*, existen desafíos técnicos importantes para los estudios transcriptómicos. Al tratarse de un patógeno intracelular y de lento crecimiento, resulta complejo aislar ARN bacteriano en cantidad y calidad suficientes para el análisis. Además, la coexistencia de ARN proveniente tanto de la bacteria como del hospedador complica la separación y análisis diferencial de las muestras.

Sin embargo, avances metodológicos recientes han mejorado significativamente estas limitaciones. Por ejemplo, un estudio empleó lisis celular diferencial y un procesamiento optimizado de muestras, logrando aumentar la detección de genes expresados por *M. tuberculosis* de tan solo 13 a 702 genes, incluyendo múltiples genes clave implicados en su ciclo patogénico (Cornejo-Granados et al. 2021). Este tipo de estrategias no solo incrementan la profundidad del análisis, sino que amplían el potencial de descubrimiento de nuevos blancos terapéuticos y biomarcadores.

Un ejemplo del potencial de la transcriptómica aplicada es el estudio comparativo de la respuesta inmune en ratones infectados con dos cepas cercanamente relacionadas del linaje Beijing de *M. tuberculosis*: la cepa BC y la cepa BL. A pesar de su elevada similitud genética, las respuestas inmunológicas que inducen son notablemente distintas (Cerezo et al. 2021).

Los perfiles transcriptómicos del hospedador fueron analizados en distintos momentos post-infección (días 3, 14, 21 y 60), observándose una evolución diferencial de la inflamación. BC

provocó una respuesta inflamatoria progresiva y sostenida, mientras que BL generó inicialmente una respuesta antiinflamatoria, caracterizada por una alta expresión de IL-10, que permitió una replicación bacteriana más eficiente. Esta fase fue seguida por una inflamación súbita e intensa, lo que precipitó la muerte de los ratones en un promedio de 23 días, en comparación con los 42 días para BC391. En el día 14, los niveles de citocinas en los ratones infectados con BL323 fueron significativamente mayores: 8 veces más TNF- α , 3 veces más IFN- γ , 3 veces más IL-12 y hasta 100 veces más IL-10 que en los ratones infectados con BC391. La coexistencia de mediadores inflamatorios y antiinflamatorios, junto con la expresión de genes inmunosupresores como Nkiras, Dleu2 y Sphk2, sugiere un perfil inmunológico complejo que favorece la progresión agresiva de la infección. (Cerezo et al. 2021).

Un estudio posterior vira la perspectiva de la investigación basada en la expresión del sistema inmune murino hacia el estudio de la variación en la expresión genética entre las variantes BC y BL durante la infección. En el estudio se explica que la progresión de la patología por BC se caracteriza por un aumento en las proteínas de secreción tipo 7, los antígenos secretados por esta vía y una depleción de la capacidad antioxidante del modelo murino por glutatión. BL por otra parte genera la proteína HRP1 un bloqueador del mediador factor nuclear kappa β (NFkB), componente esencial para la producción de citocinas proinflamatorias. El bloqueo de NFkB disminuye la presencia de citocinas proinflamatorias. Y al mismo tiempo no tiene expresión del sistema de secreción tipo 7 al día 14. Sin los pilares de inflamación y antígenos la progresión de la enfermedad se caracteriza por iniciar como antiinflamatoria (Cerezo-Cortés et al. 2022).

El análisis transcriptómico se basa en la comparación de medias de expresión de cada gen donde usualmente varía alguna condición o en este caso la cepa infectiva. Para cada gen de las cepas se responde a la hipótesis de si hay expresión diferencial o no, si difiere la expresión y organizar los genes por aquellos en los que hay mayor variación entre cepas. Observar los genes que varían entre cepas y comparar sus cantidades culmina en entender las diferencias en función o comportamiento entre las cepas bacterianas.

IV. JUSTIFICACIÓN

La tuberculosis es la enfermedad infecciosa líder en mortalidad y en impacto socioeconómico. Sin embargo, algunas manifestaciones de la enfermedad son significativamente más problemáticas que otras. La mayoría de los pacientes con tuberculosis permanecen asintomáticos y sin difundir el agente causal por décadas mientras que otros infectados por cepas de Mtb de virulencia pronunciada pueden sufrir el agravamiento de la enfermedad en cuestión de un año y/o defunción en una fracción del tiempo significativamente menor a aquel en el que otras cepas lo generan. La alta agresividad de las variantes de Mtb alza la necesidad de entender sobre los mecanismos biológicos por los cuales algunas cepas generan el aumento de virulencia.

El secuenciado masivo de los transcritos en conjunto con su análisis logra acceder a la totalidad de expresión genética del objeto de estudio representando un gran potencial analítico pero requiriendo un amplio capital de recursos humanos y computacionales. Sin embargo, seleccionar sólo aquellos genes que están mayormente presentes en una cepa de conocida mayor virulencia es una tarea asequible y con amplio potencial de identificar dianas genéticas cuyos productos impactan sobre el agravamiento de la enfermedad y dilucidar el rol biológico que cumple cada uno de ellos.

Lograr esta tarea consta de acceder a archivos de secuenciado de ARN, el genoma al cual serán alineados, el archivo de anotación de los genes del genoma y utilizar un repertorio ordenado de funciones en algún lenguaje de programación como R y/o python los cuales desembocan en información gráfica comunicable sobre la expresión diferencial.

Desde el punto de vista teórico, el reporte de los niveles de expresión y seleccionado de genes de relevancia para las cepas virulentas permite que expertos del área lideren proyectos de desarrollo de métodos de identificación de cepas problemáticas y apoyo en el desarrollo o asignación de tratamientos para responder al surgimiento de cepas resistentes y de patogenia variante. Los esquemas de tratamiento actuales tienen eficiencia limitada y generan estragos relevantes sobre el estilo de vida del paciente.

Además, realizar análisis transcriptómicos es un área de amplio desafío técnico y poca penetrancia. Difundir la metodología y el entendimiento de las bases favorece el desarrollo de competencia en el capital humano, generación de análisis más profundos, reducción de tiempo en la generación del análisis transcriptómicos y automatización de los algoritmos de análisis.

V. HIPÓTESIS

El análisis transcriptómico de los datos obtenidos en un modelo murino de infección con las cepas de *M. tuberculosis* Beijing clásica y tipo Beijing identificará expresión diferencial en genes asociados a factores de virulencia.

VI. OBJETIVO GENERAL

Elucidar la expresión genética diferencial entre las cepas BC y BL al día 14 y 28 de infección murina.

OBJETIVOS PARTICULARES

1. Generar conteos de transcritos al día 14 y 28 de infección para ambas cepas.
2. Categorizar por rol biológico (metabólico, transcripción o traducción y estrés) los 20 genes de mayor expresión diferencial por BC y BL al día 14 y 28 de infección.
3. Contrastar la variación en presencia y grado de expresión por BC y BL al día 14 y 28 de infección.

VII. MATERIALES Y MÉTODOS

7.1 Obtención de bibliotecas de datos de secuenciación de RNA

Se accedió a la plataforma de [NCBI](https://www.ncbi.nlm.nih.gov/gds/?term=GSE198877) sección Geo Datasets (Gene Expression Omnibus)(<https://www.ncbi.nlm.nih.gov/gds/?term=GSE198877>). Se descargaron los archivos correspondientes a las secuencias de ARN del proyecto número: PRJNA817222.

Los archivos correspondieron a los días 14 y 28 de un modelo de infección con Mtb cepas BC o BL.

Archivos empleados

48 archivos de secuenciación de RNA fueron empleados para realizar análisis transcriptómicos y comparativos. Los 48 archivos se dividen en 24 para cada una de las 2 cepas BC y BL y posteriormente se subdividen en 12 para cada punto temporal; 12 para el día 14 y 12 para el día 28 en cada caso respectivamente.

BC d 14	BL d 14	BC d 28	BL d 28
SRR18348642	SRR18348652	SRR18348670	SRR18348673
SRR18348641	SRR18348623	SRR18348669	SRR18348620
SRR18348640	SRR18348622	SRR18348668	SRR18348621
SRR18348639	SRR18348651	SRR18348667	SRR18348680
SRR18348638	SRR18348650	SRR18348605	SRR18348679
SRR18348637	SRR18348649	SRR18348664	SRR18348678

SRR18348636	SRR18348628	SRR18348663	SRR18348677
SRR18348666	SRR18348627	SRR18348662	SRR18348676
SRR18348665	SRR18348626	SRR18348612	SRR18348675
SRR18348633	SRR18348625	SRR18348611	SRR18348648
SRR18348672	SRR18348624	SRR18348610	SRR18348646
SRR18348671	SRR18348674	SRR18348609	SRR18348645

El formato de los archivos descargados fue Fastq. Cada archivo contiene 2 secuencias: la primera conforme a la información genética y la segunda conforme a la calidad de secuenciado o probabilidad de error del secuenciador. Usé la escala de fred para interpretarla.

Los nombres de los archivos son empleados seguidos de la función fasterqdump del software SRA toolkit disponible en <https://github.com/ncbi/sra-tools/wiki/02.-Installing-SRA-Toolkit> para realizar la descarga de los archivos.

7.2 Transformación de archivos desde fastq a .bam

Los archivos descargados en formato fastq de las secuencias de las cepas de *M. tuberculosis* se importaron a una computadora local (ambiente R) y se transformaron al formato de alineación binaria. bam. Posteriormente se generaron los conteos genéticos mediante 3 funciones de la librería Rsubread, Rsamtools : buildindex, align y featureCounts en el ambiente de programación Rstudio (Liao et al. 2019).

El genoma de referencia contra el cual se realizó la alineación fue el de *Mtb* cepa BC, número de acceso en NCBI (CP017596).

```
# Directorios -----
# Indexar un genoma Rsubread
getwd()
genome_path <- "C:/Users/david/Archivos-Directorios/1. Maestría/0Tesis Maest

BC391path <- "C:/Users/david/Archivos-Directorios/1. Maestría/0Tesis Maestrí
BC391index <- paste("C:/Users/david/Archivos-Directorios/1. Maestría/0Tesis

#AnnotationFiles
list.files("C:/Users/david/Archivos-Directorios/1. Maestría/0Tesis Maestría/
list.files("C:/Users/david/Archivos-Directorios/1. Maestría/0Tesis Maestría/
BCGFF <- "C:/Users/david/Archivos-Directorios/1. Maestría/0Tesis Maestría/1.
BLGFF <- "C:/Users/david/Archivos-Directorios/1. Maestría/0Tesis Maestría/1.

buildindex( #BC391
  basename = BC391index,
  reference = BC391path)

buildindex( #BL
  basename = BLindex,
  reference = BLpath)

|
buildindex(#Musmusculus
  basename = Musmusculusindex,
  reference = Musmusculuspath)
```

Figura 4. Generación de un índice de genoma de referencia y contenedores de archivos de secuenciado. Paso 1: Generar la ubicación del genoma de referencia (BC391path). Paso 2: Generar la ubicación del genoma producto de la indexación (BC391index). Paso 3: Mediante la función buildindex y utilizando ambas ubicaciones generar el indexado.

```

#Alinear los datos Rsamtools
rutaarchivosBC391D14 <- "C:/Users/david/Archivos-Directorios/1. Maestría/0Tesis Maestría/1. AT er
rutaarchivosBL323D14 <- "C:/Users/david/Archivos-Directorios/1. Maestría/0Tesis Maestría/1. AT er

rutaoutputBAMBCD14 <- "C:/Users/david/Archivos-Directorios/1. Maestría/0Tesis Maestría/1. AT er
rutaoutputBAMBLD14 <- "C:/Users/david/Archivos-Directorios/1. Maestría/0Tesis Maestría/1. AT er

list.files(rutaoutputBAMBCD14)
list.files(rutaoutputBAMBLD14)

#Variables genomas indexados
BC391index
BLindex

# Alineamientos BAM -----
#Alineamientos Archivo 1 al 12. BC391.
align( #1 BC
  index = BC391index,
  readfile1 = paste(rutaarchivosBC391D14, "/SRR BR1T1 BC D14.fastq", sep = ""),
  output_file = "C:/Users/david/Archivos-Directorios/1. Maestría/0Tesis Maestría/1. AT er
  nthreads = 4,
  input_format = "FASTQ"
)

```

Figura 5. Alineamiento de archivos de secuenciado a genoma referencia.

Paso 1: Archivos de interés de BC al día 14 (rutaarchivosBC391D14). Paso 2: Generar la dirección en donde se depositarán los archivos alineados en formato .bam. Paso 3: Emplear la función align usando el índice, un archivo a la vez de secuenciado y el lugar en donde será depositado en formato url.

7.3 Tabla de Conteo

Previo a realizar el conteo de transcritos se puede requerir ordenar los archivos e indexarlos mediante las funciones `sortBam()` e `indexBam()` de la librería `samtools`. El uso de ambas funciones requiere como información de entrada la url singular a un archivo `.bam`.

Se empleó la función “`featurecounts`” que toma como archivos de entrada un archivo formato BAM y un archivo de anotaciones de los genes correspondiente al genoma al cual se alineó; los archivos de anotaciones tienen una terminación `.gff3` y también son obtenidos de NCBI y como archivo de salida genera una tabla organizada con el nombre de los genes y el conteo o repeticiones de cada gen. Se utiliza el operador “`$counts`” para acceder a la información de conteos de transcritos de cada archivo.

```
BC391_1_tc <- featureCounts(files = "C:/Users/david/1.Archivos-Directorios/1. Maestría/0Tesis Maestría/1. A1  
                             annot.ext = BCGTF,  
                             isGTFAnnotationFile = TRUE,  
                             GTF.featureType = "gene", # Feature type, e.g., 'exon' for RNA-seq  
                             GTF.attrType = "ID", # Attribute to use for grouping counts (e.g., 'gene_id')  
                             useMetaFeatures = TRUE, # Sum counts over meta-features (e.g., genes)  
                             nthreads = 4)
```

Figura 6. Conteo de transcritos.

Paso 1: Imputar el url al archivo `.bam` en el argumento `files`. Paso2: Imputar el url a un archivo de anotaciones del genoma indexado.

Se generó un marco de datos con la sección de conteos de todos los archivos y una tabla de metadatos orientado a organizar mediante anotaciones los datos. Se requiere para realizar la expresión diferencial.

```
#Generar un marco de datos de metadatos.  
metadatosBCyBL <- data.frame(  
  IDmuestras = muestras,  
  cepa = rep(c("BC391", "BL323"), each = 12),  
  replicadoB = rep(1:3, each = 4, times = 2),  
  replicadoT = rep(1:4, times = 6)  
)  
  
view(metadatosBCyBL)
```

Figura 7. Generación de metadatos.

Marco de datos con los nombres de las muestras, cepa de la cual provienen los datos, número de replicado biológico y replicado técnico. Estos metadatos corresponden a solo 24 archivos ya que las comparaciones fueron realizadas por día, 12 archivos por cepa por día.

7.4 Análisis transcriptómico y normalización

Se empleó la librería DESeq2 para realizar el análisis transcriptómico y la compensación por tamaño de librería. La función `DESeqDataSetFromMatrix` genera una estructura específica para el análisis transcriptómico. Una matriz de los conteos crudos anotados con los metadatos.

Se empleó la función `DESeq` para generar el modelado y pruebas estadísticas de diferencia en expresión entre los grupos de estudio beijing clásico y tipo beijing.

```
ddsBC <- DESeqDataSetFromMatrix(countData = BCdfft1,
                                colData = metadatosBC,
                                design = ~ replicadoB + replicadoT)
ddsBCnormalizado <- DESeq(ddsBC, fitType = "mean")
str(ddsBCnormalizado)
view(ddsBCnormalizado)

conteos_normalizadosBC <- counts(ddsBCnormalizado, normalized = T)
```

Figura 8. Matriz para análisis transcriptómico y normalizado. Se genera el marco de datos compatible para el análisis transcriptómico a partir de los conteos usando `DESeqDataSetFromMatrix`. Se realiza el modelado y conteos normalizados por `DESeq` y `counts` respectivamente.

Se empleó `DESeq2` funciones “collapseReplicates”, “DESeq”, “results” y `ggplot2` para visualizar el gráfico de expresión diferencial. El colapsado de replicados es importante para buscar solamente diferencias entre las cepas y no entre las réplicas biológicas. Posteriormente se emplea `DESeq` para el modelado y `results` para almacenar los resultados. Usando `ggplot2` se grafica en el eje x la tasa de cambio y en el eje y la validez estadística.

```
ddsBCyBLcolapsadoRT <- collapseReplicates(NormalizadoBCyBL, groupby = biological_replicates, run = technical_replicates)
dim(ddsBCyBLcolapsadoRT)

ddsBCyBLcolapsadoRT <- DESeq(ddsBCyBLcolapsadoRT)
dim(ddsBCyBLcolapsadoRT)
head(ddsBCyBLcolapsadoRT)

resBCyBLcol <- results(ddsBCyBLcolapsadoRT, contrast = c("cepa", "BL323", "BC391"))
view(resBCyBLcol)
ggplot(resBCyBL, aes(x=log2FoldChange, y=pvalue, color = significanceval)) + #y=-log10(pvalue)
  geom_point(alpha=0.5, size=2) +
  geom_text(data=top_genes, aes(label=row.names(top_genes)), vjust=1, hjust=1, size=3) +
  geom_text(data = top_genesmore, aes(label=row.names(top_genesmore)), vjust=1, hjust=1, size=3) +
  scale_color_manual(values = c("Yes" = "red", "No" = "black")) + # Red for significant, black for non-significant
  theme_minimal() +
  #geom_hline(yintercept = 0.05, linetype = "dashed", color = "blue")+
  #geom_vline(xintercept = c(-0.807, -1.807, -2.807, -3.807, -4.807), linetype = "dashed", color = "blue")+
  labs(title="Expresión diferencial BL323 vs BC391", x="Log2 Fold Change", y="P-value", subtitle = "Top 10 up/down") +
  theme(legend.position="")
```

Figura 9. Uso de funciones para realizar el análisis transcriptómico y generar el gráfico de volcán.

7.5 Procesamiento de datos

Se exportaron a excel desde R los 20 genes más expresados por cepa y punto temporal empleando la tasa de cambio como variable de filtrado.

```
#de 8 a 30 veces más expresado que en BC. los top 10.
#genes de 2 a 8 veces más expresados en BC que en BL.
top_genesBLmas <- head(resBCyBLcol[order(resBCyBLcol$log2FoldChange), ], 20)
top_genesBCmas <- head(resBCyBLcol[order(resBCyBLcol$log2FoldChange, decreasing = T), ], 20)
view(top_genesBCmas)
view(top_genesBLmas)
dfExpresiónmasenBLqueBC <- data.frame(top_genesBLmas)
dfExpresiónmasenBCqueBL <- data.frame(top_genesBCmas)
view(dfExpresiónmasenBLqueBC)
view(dfExpresiónmasenBCqueBL)
```

Figura 10. Selección de genes de expresión diferencial.

Los 20 genes de tasa relativa menor corresponden a los genes de menor expresión en beijing clásico. Mientras que los genes de mayor tasa relativa corresponden a los genes que de mayor expresión diferencial con sobreexpresión relativa a tipo beijing.

Se empleó el excel para nombrar los genes conforme al archivo de anotaciones genéticas para *M. tuberculosis* (.gff3).

Posteriormente se asignó la clasificación de los genes en grupos de genes metabólicos, de regulación transcripcional y de respuesta a estrés.

```
#Paso 3:loop de reemplazo
#Iterar sobre los grupos_prote[][] y sobre basesota[geneproduct],
#Asignar ese valor a basesota[clasificacion]
pendientes_a_clasificar = []
for group_name, group_values in grupos_prote_dict.items():
    for value in group_values:
        basesota.loc[basesota["geneproduct"]== value, "clasificacion"] = group_name
pendientes_a_clasificar = basesota[basesota["clasificacion"].isna()]["geneproduct"].tolist()

print(basesota["clasificacion"].isna().sum())
print(pendientes_a_clasificar)
```

Figura 11. Código de asignación de clasificación a los genes.

Cuando el nombre de un gen sea encontrado en los valores de un grupo de proteínas se le asigna el nombre del grupo en la columna clasificación.

Gráficos de pay y barras fueron realizados empleando matplotlib para comparar el grado de expresión entre los genes de expresión diferencial mediante código en python.

El proceso general de graficado involucra la organización y filtrado de los datos a ser observados seguido de graficar.

```

#Conteo de frecuencias en grupos de proteínas
BCD14vsBLD14 = basesotaBCD14vsBLD14["clasificacion"].value_counts()
BLD14vsBCD14 = basesotaBLD14vsBCD14["clasificacion"].value_counts()
BCD28vsBLD28 = basesotaBCD28vsBLD28["clasificacion"].value_counts()
BLD28vsBCD28 = basesotaBLD28vsBCD28["clasificacion"].value_counts()

#Generar un marco de datos unido
data_graph_bar_comparison = {"BCD14vsBLD14": BCD14vsBLD14,
"BLD14vsBCD14": BLD14vsBCD14,
"BCD28vsBLD28": BCD28vsBLD28,
"BLD28vsBCD28": BLD28vsBCD28}
df = pd.DataFrame(data_graph_bar_comparison)

#Graficar las frecuencias de grupos de genes.
df.plot(kind='bar', figsize=(10, 6), width=0.8, color = ["black", "red", "green", "skyblue"])

plt.title("Frecuencia de grupos de proteínas para genes sobreexpresados por cepa y día")
plt.xlabel("Grupos de Proteínas")
plt.ylabel("Frecuencia")
plt.xticks(rotation=85)
plt.legend(title="Columns")
plt.tight_layout()
plt.show()

```

Figura 12. Gráfico de frecuencias de grupos de genes.

Se contaron los valores de la columna clasificación mediante `.value_counts`, se unieron los datos en un marco de datos y se graficó empleando la función `.plot`.

7.6 Ajuste de valores de expresión promedio para visualización

En el análisis transcriptómico, las condiciones de BL en los días 14 y 28 se establecen como referencia basal. Por lo tanto, los valores de log2 Fold Change (log2FC) cuantifican la magnitud y dirección de los cambios en la expresión génica de BC en comparación con BL en esos mismos puntos temporales. Para obtener un valor de BL es necesario convertir el promedio de expresión de BC con el antilogaritmo absoluto base 2 de la tasa de cambio de BC. Ya que la tasa de cambio expresa cómo se compara un grado de expresión con aquél del otro grupo es decir BL. Ejemplo: Si un gen de BC tiene -0.55 como tasa de cambio (log2FC) quiere decir que (antilog2 de 0.55) BL tiene 1.515 veces más expresión que BC por lo tanto se multiplica el promedio de expresión por las veces de expresión para conseguir el valor de BL.

Revisión de código en Github:
<https://github.com/DavidSuarez111/MasterinM.Tuberculosis>

VIII. RESULTADOS

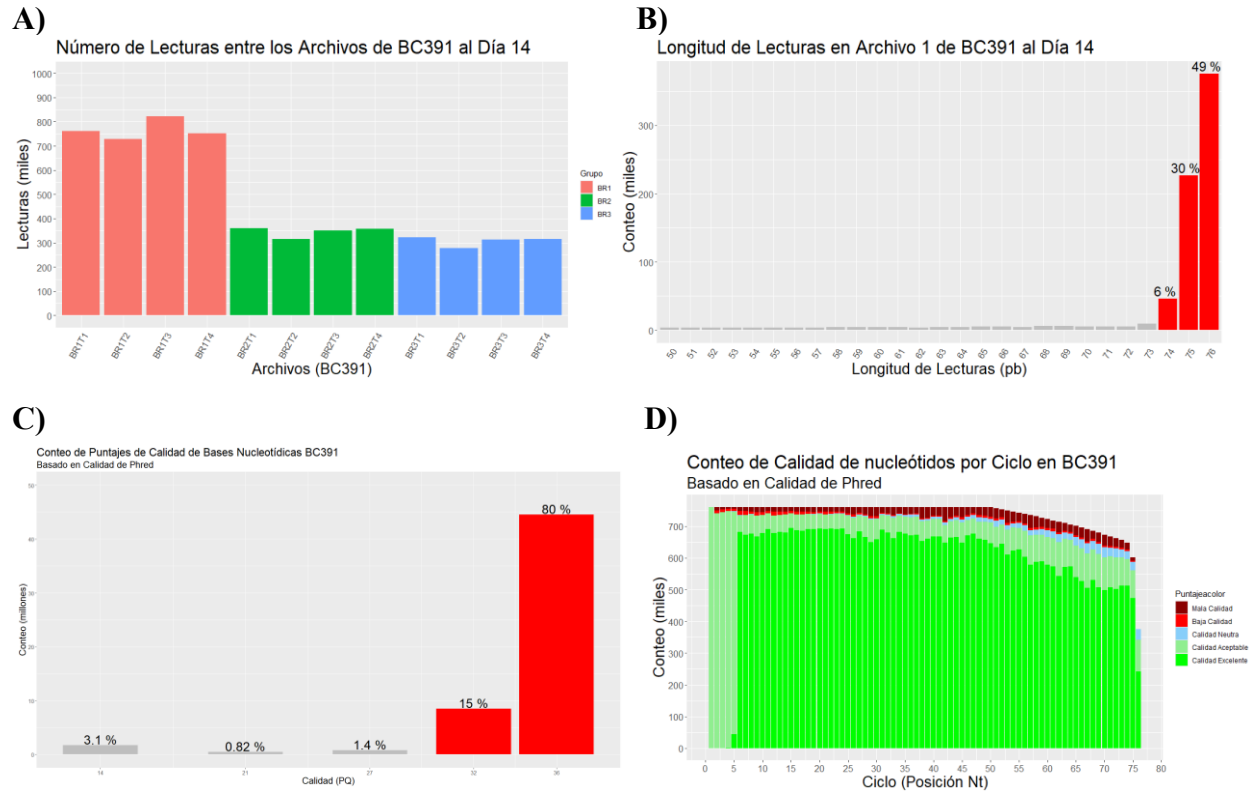


Figura 13: Gráficos de control de calidad. A) Cantidad de lecturas por archivo de los 12 archivos de BC391 al día 14. B) Distribución del rango de tamaños de las lecturas en pares de bases de 1 archivo (de 50 a 76 bases). C) Probabilidad de error en la escala de Phred del secuenciado. El 95% del llamado de bases tiene una calidad mayor a 30 lo equivalente a equivocarse menos de 1 vez por cada 1000 bases asignadas. D) Distribución de calidad del llamado de bases en las posiciones (1-76) en todas las lecturas. Las bases de mayor riesgo de llamado incorrecto se distribuyen homogéneamente a través de las posiciones de las lecturas. Los gráficos fueron replicados para revisar la calidad de los archivos y no está limitado a solo la evidencia en la presente imagen (ver anexos).

Tabla 5: Resumen de alineamientos de archivos por BC391 y BL323 al día 14.

NombreArchivo	Mapeados	LecturasTotales	PorcentajedeMapeo	Resumen
BC1	805	760323	0.106	Promedios de Alineamiento en Porcentaje
BC2	728	727193	0.1	BC: 0.08
BC3	932	822406	0.113	BL: 0.61
BC4	799	751707	0.106	
BC5	243	360094	0.0675	Rango de lecturas Mapeadas
BC6	241	315519	0.0764	BC: 155 a 932
BC7	261	350527	0.0745	BL: 1041 a 4633
BC8	248	356507	0.0696	
BC9	173	323216	0.0535	Promedio Lecturas
BC10	155	278112	0.0557	BC: 414
BC11	192	312284	0.0615	Promedio Lecturas
BC12	188	315266	0.0596	BL: 2806
BL1	3995	486048	0.822	7x librería
BL2	3711	465565	0.797	
BL3	4633	526771	0.88	Promedio Lecturas Totales
BL4	3992	480462	0.831	BC: 472763
BL5	1218	372425	0.327	BL: 437959
BL6	1041	327610	0.318	7% de diferencia en tamaño
BL7	1258	364503	0.345	
BL8	1167	368597	0.317	
BL9	3230	487787	0.662	
BL10	2857	423531	0.675	
BL11	3297	472250	0.698	
BL12	3271	479957	0.682	

Se pueden observar los nombres de archivos, lecturas mapeadas al genoma referencia, lecturas totales que componen los archivos, porcentaje de mapeo y estadística descriptiva en la columna de resumen.

Resumen de alineamientos

24 archivos integrados por RNA de ratón (*Mus musculus*) y RNA de *Mycobacterium tuberculosis* fueron alineados al genoma de *M. tb.* 12 archivos con transcritos provenientes de infección con

M. tb. BC391 y otros 12 con *M. tb.* BL323. Para el caso de muestras con RNA de *M. tb.* BC391 la sumatoria de todas las lecturas en los archivos es de 5.67 millones de los cuales solo 4965 transcritos corresponden a lecturas de *M. tb.* El promedio de lecturas totales por archivo BC391 es de 472763 mientras que el promedio de lecturas alineadas a *M. tb.* es de 414 transcritos. Esto quiere decir que por cada transcrito de *M. tuberculosis* en la muestra se encuentran poco más de 11000 transcritos pertenecientes al genoma de ratón.

Las muestras de secuenciación de RNA de BL323 tienen un total de lecturas de 5.25 millones de las cuales 33670 son transcritos que se alinean con el genoma de *M. tuberculosis*. Esto quiere decir que por cada transcrito de *M. tb.* BL323 hay solo 155 copias del genoma de ratón.

Al comparar las bibliotecas el total de lecturas de los archivos con genes BC391 es 7% superior al total de lecturas de los archivos con genes de BL323 sin embargo el porcentaje de mapeo para BC391 es de 0.0008 % mientras que para BL323 es de 0.061 %. Además hay una proporción de gen de ratón por gen de bacteria 71 veces menor que en las muestras con transcritos de BC391.

Lo más relevante a considerar es que si bien la librería de lecturas totales solo varía 7% a favor de BC391 al considerar los RNA bacterianos extraídos son 6.78 veces mayores para las muestras infectadas por BL323 que aquellas por BC391. De aquí se obtiene el entender que, al menos en estas muestras, al día 14 hay ≈680% más actividad de expresión por BL323 que por BC391.

Para el análisis transcriptómico solamente fueron empleados los transcritos bacterianos, es decir los casi 5000 transcritos por los archivos BC391 y los 33670 transcritos por los archivos BL323 al día 14.

Tabla 6: Resumen de alineamientos de archivos por BC391 y BL323 al día 28.

NombreArchivo	Mapeados	LecturasTotales	PorcentajedeMapeo	Resumen
BCd28_1	1019	316888	0.322	Promedios de Alineamiento en Porcentaje
BCd28_2	926	304443	0.304	BC: 0.39
BC28_3	1122	343236	0.327	BL: 0.72
BC28_4	1041	315347	0.33	
BC28_5	1518	482381	0.315	Rango de lecturas Mapeadas
BC28_6	1415	458231	0.309	BC: 926 a 2408
BC28_7	1695	525108	0.323	BL: 1860 a 4350
BC28_8	1465	479213	0.306	
BC28_9	2212	406085	0.545	Promedio Lecturas
BC28_10	2032	386310	0.526	BC: 1589
BC28_11	2408	442414	0.544	Promedio Lecturas
BC28_12	2216	398867	0.556	BL: 3471
BLd28_1	2217	430480	0.515	2.18x librería alineada
BLd28_2	1860	377710	0.492	
BLd28_3	2220	421417	0.527	Promedio Lecturas Totales
BLd28_4	2210	426017	0.519	BC: 404877
BLd28_5	4337	504917	0.859	BL: 474250
BLd28_6	3773	435718	0.866	14.6% de diferencia en tamaño
BLd28_7	4292	483221	0.888	
BLd28_8	4146	491119	0.844	
BLd28_9	4329	563673	0.768	
BLd28_10	3677	479075	0.768	
BLd28_11	4350	533634	0.815	
BLd28_12	4248	544014	0.781	

Se pueden observar los nombres de archivos, lecturas mapeadas al genoma referencia, lecturas totales que componen los archivos, porcentaje de mapeo y estadística descriptiva en la columna de resumen.

De manera similar a la comparación al día 14, 24 archivos de secuenciación de RNA fueron empleados al día 28; 12 para cada cepa respectivamente.

Para BC391 al día 28 se obtuvo un total de 4.85 millones de lecturas mientras que para BL323 fueron 5.69 millones de lecturas. 14.6% de variación entre la cantidad de lecturas totales. De las 4.85 millones de lecturas de BC391, 19069 son lecturas alineadas a *M. tb* mientras que para los archivos BL323 se tiene un total de 41659 lecturas de transcritos alineados a *M. tb*. Las lecturas de *M. tb*. encontradas en los archivos BC391 y BL323 corresponden al 0.39 y 0.72 por ciento del archivo total.

Los transcritos de *M. tb*. generan una sublibrería que es 2.18 veces más grande para BL323 que para BC391, lo equivalente a decir que hay alrededor del doble de expresión de RNA de BL323 al día 28 en la muestra comparado a BC391 en el mismo punto temporal.

Los promedios por archivo de lecturas totales son de aproximadamente 405 mil para BC391 y 474 mil para BL323 mientras que los promedios mapeados a *M. tb*. son 1589 y 3471 respectivamente.

Para el análisis transcriptómico se utilizaron solamente los 19 mil y 41 mil lecturas mapeadas a *M. tb*.

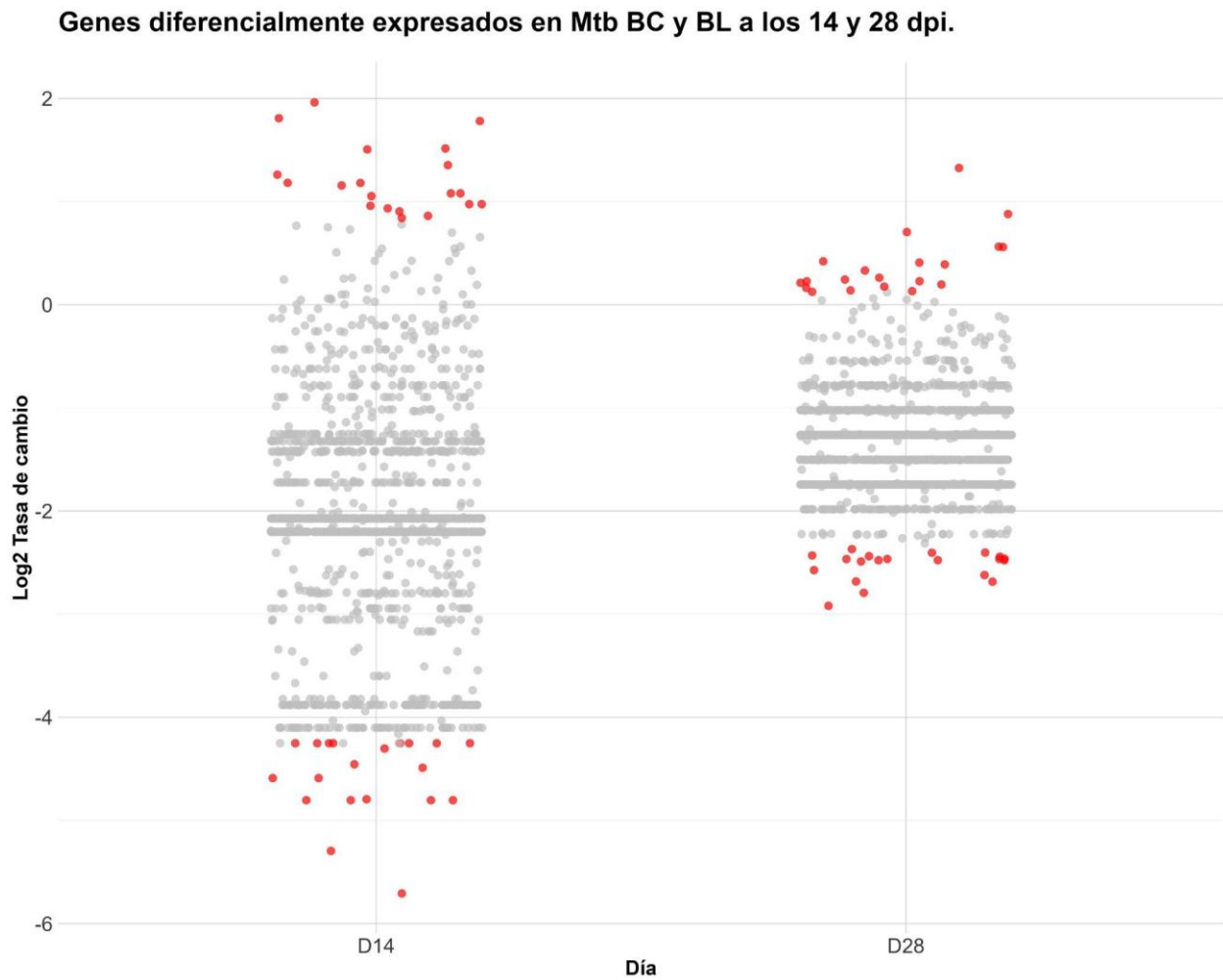


Figura 14: Expresión diferencial de BC y BL al día 14 y 28. Los puntos en rojo señalan los genes diferencialmente expresados (DEG). Los DEG ubicados en la parte superior corresponden a genes expresados a la alza en BC mientras que los DEG ubicados en la parte inferior son sobreexpresados en BL.

Tabla 7: 20 genes de mayor expresión diferencial en BC391 al día 14.

Gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
gene-BJM02_19040	1.0277237	1.9618873	2.246889	0.8731572	0.3825774	0.5112721
gene-BJM02_10680	0.8835841	1.8082302	2.846097	0.6353367	0.5252088	0.6218348
gene-BJM02_02420	0.8629027	1.7808774	3.314742	0.5372596	0.5910883	0.6821627
gene-BJM02_20260	0.7322778	1.5145932	2.749304	0.5509005	0.5817019	0.6756887
gene-BJM02_03580	0.7103282	1.505496	2.985275	0.5043073	0.6140454	0.6945546
gene-BJM02_13210	0.6647479	1.3534096	3.705627	0.365231	0.714939	0.7934928
gene-BJM02_20150	0.6178995	1.2611842	3.909431	0.3226004	0.7469979	0.8132391
gene-BJM02_02415	6.2284482	1.1816234	1.328724	0.8892916	0.3738464	0.5112721
gene-BJM02_01110	1.3292826	1.1804937	2.271009	0.5198102	0.6031958	0.6837362
gene-BJM02_09775	0.571051	1.1565977	3.938239	0.293684	0.7689994	0.8279671
gene-BJM02_16570	0.52842	1.0795534	3.939667	0.2740215	0.7840681	0.8415418
gene-BJM02_18260	0.52842	1.0795534	3.939667	0.2740215	0.7840681	0.8415418
gene-BJM02_14080	0.5383404	1.0529906	3.94095	0.2671921	0.7893213	0.8462101
gene-BJM02_08375	0.5503696	0.9753014	3.944817	0.2472362	0.8047254	0.8590915
gene-BJM02_20560	0.5503696	0.9753014	3.944817	0.2472362	0.8047254	0.8590915
gene-BJM02_01495	0.4936007	0.9594979	3.942839	0.2433521	0.8077327	0.8615378
gene-BJM02_05655	0.491492	0.9340392	3.908467	0.2389784	0.8111223	0.8641864
gene-BJM02_03575	0.4716511	0.9052584	3.835617	0.2360137	0.813422	0.8660584
gene-BJM02_12995	0.5014124	0.8619542	3.946326	0.2184194	0.8271023	0.8800369
gene-BJM02_12650	0.448861	0.8412843	3.945468	0.213228	0.8311491	0.8837535

Tabla 8: 20 genes de mayor expresión en BL323 al día 14.

Gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
gene-BJM02_02830	2.1671825	-5.706976	2.514182	-2.269914	0.02321282	0.4066629
gene-BJM02_11725	1.3432697	-5.295131	2.98641	-1.773076	0.07621614	0.4066629
gene-BJM02_00800	0.8615691	-4.803291	3.316017	-1.448512	0.14747384	0.4066629
gene-BJM02_03820	0.8615691	-4.803291	3.316017	-1.448512	0.14747384	0.4066629
gene-BJM02_11220	0.8615691	-4.803291	3.316017	-1.448512	0.14747384	0.4066629
gene-BJM02_14255	0.8615691	-4.803291	3.316017	-1.448512	0.14747384	0.4066629
gene-BJM02_15125	1.5779437	-4.79379	2.496883	-1.91991	0.05486927	0.4066629
gene-BJM02_01580	1.378089	-4.589072	2.772126	-1.655434	0.09783652	0.4066629
gene-BJM02_17675	1.378089	-4.589072	2.772126	-1.655434	0.09783652	0.4066629
gene-BJM02_11815	0.5698027	-4.488972	4.037482	-1.111825	0.26621357	0.4553878
gene-BJM02_03715	1.3880094	-4.455136	2.738602	-1.626792	0.10378123	0.4066629
gene-BJM02_13020	1.4129083	-4.302803	2.733797	-1.573929	0.11550381	0.4066629
gene-BJM02_02055	0.3798684	-4.251166	3.607415	-1.178452	0.23861633	0.433288
gene-BJM02_04055	0.3798684	-4.251166	3.607415	-1.178452	0.23861633	0.433288
gene-BJM02_05110	0.3798684	-4.251166	3.607415	-1.178452	0.23861633	0.433288
gene-BJM02_05850	0.3798684	-4.251166	3.607415	-1.178452	0.23861633	0.433288
gene-BJM02_06375	0.3798684	-4.251166	3.607415	-1.178452	0.23861633	0.433288
gene-BJM02_07915	0.3798684	-4.251166	3.607415	-1.178452	0.23861633	0.433288
gene-BJM02_08205	0.3798684	-4.251166	3.607415	-1.178452	0.23861633	0.433288
gene-BJM02_11045	0.3798684	-4.251166	3.607415	-1.178452	0.23861633	0.433288

Tabla 9: 20 genes de mayor expresión en BC391 al día 28 al ser comparado con BL323 al día 28.

Gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
gene-BJM02_17200	0.6272582	1.325954	1.077497	1.23058713	0.21847732	0.8079107
gene-BJM02_07780	1.2485988	0.8792174	0.8839982	0.99459188	0.319934796	0.8079107
gene-BJM02_02420	0.371143	0.7048651	1.5837078	0.44507272	0.656267199	0.8079107
gene-BJM02_01490	0.6506408	0.5632137	1.1386528	0.49463164	0.620860166	0.8079107
gene-BJM02_06005	0.26752	0.559705	1.8495279	0.30262049	0.762179098	0.8308978
gene-BJM02_16365	55.0281506	0.4206383	0.1324964	3.17471429	0.001499841	0.8079107
gene-BJM02_11730	45.914277	0.4084811	0.1562905	2.61360216	0.008959329	0.8079107
gene-BJM02_07330	0.298042	0.3908085	1.8156116	0.21524897	0.829573216	0.8602879
gene-BJM02_10530	0.1991657	0.331879	2.0310024	0.16340652	0.870198371	0.8878743
gene-BJM02_06995	0.4280494	0.2622257	1.3780051	0.19029372	0.849078978	0.8767316
gene-BJM02_16550	0.5731932	0.2436789	1.1931225	0.20423632	0.838168834	0.8669339
gene-BJM02_03485	0.9015904	0.2280711	0.9171013	0.24868691	0.803602975	0.838069
gene-BJM02_04115	0.247148	0.2273813	1.6751116	0.13574099	0.89202604	0.90731
gene-BJM02_18445	0.1782806	0.2126116	2.4651879	0.08624561	0.931271174	0.9394234
gene-BJM02_13875	0.5571291	0.1960611	1.2439658	0.15760976	0.874764314	0.8911406
gene-BJM02_11095	0.4580443	0.176259	1.381216	0.12761144	0.898456474	0.9129026
gene-BJM02_17045	0.7991266	0.1638572	1.0241659	0.15999087	0.872888268	0.889692
gene-BJM02_02415	3.2178833	0.13974	0.5220189	0.26769148	0.788936812	0.8308978
gene-BJM02_05700	0.2345893	0.1316668	2.2169655	0.05939054	0.952641046	0.959005
gene-BJM02_16975	0.2166868	0.1249909	2.2936281	0.05449486	0.956540906	0.9624364

Tabla 10: 20 genes de mayor expresión en BL323 al día 28 al ser comparado con BC391 al día 28.

Gene	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
gene-BJM02_09385	0.4552978	-2.919127	1.804104	-1.6180478	0.10565229	0.8079107
gene-BJM02_02270	0.871927	-2.792418	1.105557	-2.5258024	0.01154344	0.8079107
gene-BJM02_00690	0.4393063	-2.685139	2.768602	-0.9698537	0.33211943	0.8079107
gene-BJM02_13365	0.7487134	-2.683179	1.220278	-2.1988259	0.0278903	0.8079107
gene-BJM02_06365	0.6034691	-2.620995	1.393364	-1.881056	0.05996431	0.8079107
gene-BJM02_16580	0.9947414	-2.573229	1.046753	-2.4582968	0.01395977	0.8079107
gene-BJM02_20560	0.2714801	-2.488832	3.002221	-0.8289968	0.40710624	0.8079107
gene-BJM02_13210	0.6856598	-2.477433	1.270924	-1.9493168	0.05125761	0.8079107
gene-BJM02_16110	0.2371164	-2.476723	2.387889	-1.0372021	0.29964166	0.8079107
gene-BJM02_04245	0.4356188	-2.476311	1.822639	-1.3586407	0.17426049	0.8079107
gene-BJM02_06380	0.2209331	-2.465446	2.662865	-0.9258623	0.35451753	0.8079107
gene-BJM02_13010	0.2712841	-2.464485	2.27723	-1.0822295	0.27915057	0.8079107
gene-BJM02_08610	0.2675591	-2.464476	3.003175	-0.8206235	0.41186076	0.8079107
gene-BJM02_09845	0.2135183	-2.46269	2.696259	-0.9133729	0.36104647	0.8079107
gene-BJM02_06255	0.5028192	-2.442162	1.62888	-1.499289	0.13379868	0.8079107
gene-BJM02_20510	0.2344378	-2.437822	2.391131	-1.0195269	0.3079529	0.8079107
gene-BJM02_20345	0.3668097	-2.430551	1.922687	-1.2641429	0.20617872	0.8079107
gene-BJM02_18980	0.2253672	-2.404294	3.000396	-0.8013257	0.42294309	0.8079107
gene-BJM02_19480	0.3376836	-2.402747	1.971423	-1.2187883	0.22292455	0.8079107
gene-BJM02_11900	0.3178317	-2.369548	2.776159	-0.8535346	0.39336292	0.8079107

Las tablas 7 a 10 contienen los genes seleccionados por expresión diferencial. Los nombres de los genes están codificados conforme a la anotación del genoma de BC de ncbi.

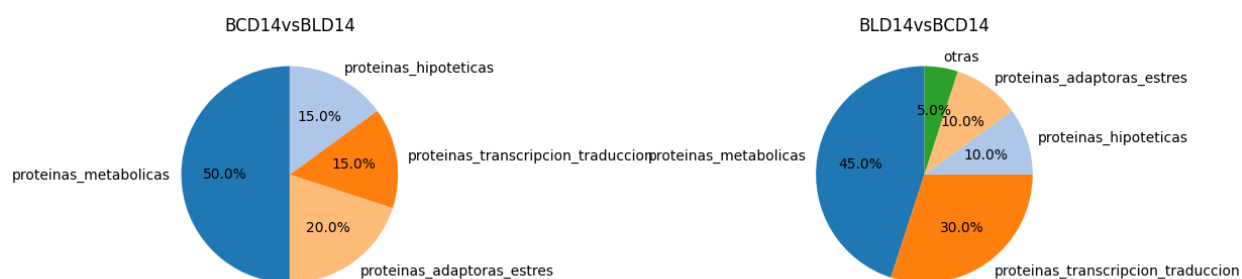


Figura 15: Gráficos de pay de grupos de proteínas en genes sobreexpresados en BC391 y BL323 al día 14.

En BC391 al día 14 se observa mayor presencia de proteínas metabólicas (10/20), seguido por 3 proteínas hipotéticas, 4 proteínas de adaptación al estrés, 3 de transcripción y traducción mientras que en BL323 9 de los 20 genes en sobreexpresión relativa correspondieron a proteínas metabólicas, 6 para proteínas de transcripción y traducción, 2 hipotéticas y 2 de adaptación al estrés.

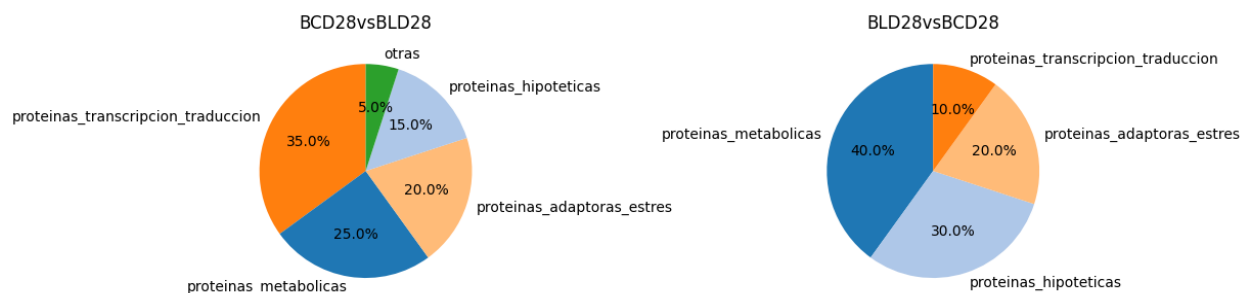


Figura 16: Gráficos de pay de grupos de proteínas en genes sobreexpresados en BC391 y BL323 al día 14.

En BC391 al día 28 se observa menor presencia de proteínas metabólicas 5 de 20, 4 para genes de proteínas de adaptación al estrés, 3 hipotéticas, 5 secuencias para proteínas reguladoras transcripcionales mientras que en BL323 hay más proteínas metabólicas 8 de 20 así como más hipotéticas 6 de 20 también, 4 para proteínas de adaptación al estrés y menor regulación transcripcional.

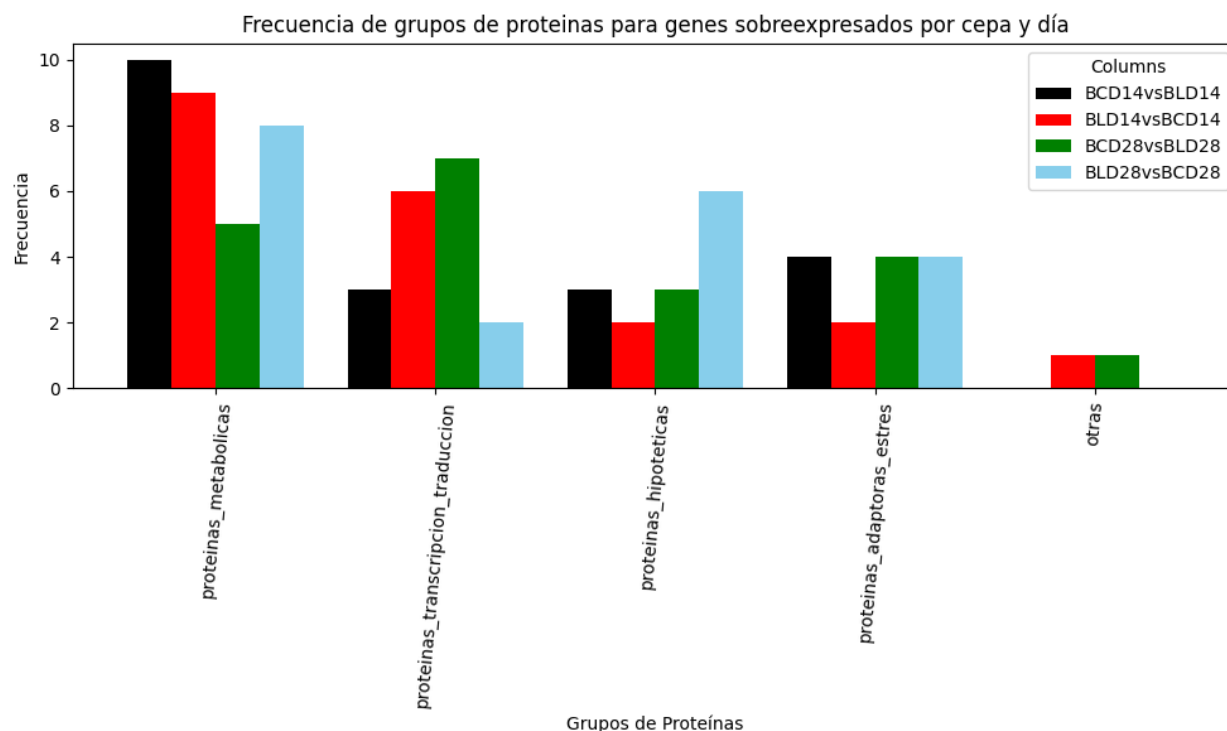


Fig 17: Frecuencia de grupos de proteínas o genes por cepa y día.

En negro se encuentra BC391 al día 14, BL323 al día 14 en rojo, BC391 al día 28 en verde, BL323 al día 28 en celeste. Las proteínas sobreexpresadas para BC391 al día 14 son primordialmente metabólicas (10 de 20 totales) seguidas por 3 hipotéticas, 3 de transcripción traducción y 4 adaptoras a estrés. BL al día 14 hay 9 proteínas metabólicas, 6 proteínas de transcripción o traducción, 2 proteínas hipotéticas, 2 adaptoras a estrés. BC391 al día 28 se caracteriza por 5 genes para proteínas metabólicas, 7 de transcripción o traducción, 3 proteínas hipotéticas, 4 adaptoras a estrés. BL323 al día 28 se caracteriza por 8 proteínas metabólicas, 6 proteínas hipotéticas, 4 proteínas de adaptación al estrés, 2 de regulación transcripcional o traducción.

Hay una mayor representación de genes diferencialmente expresados en el índole metabólico (32 de 80 totales), 18 de regulación transcripcional y traduccional y 14 de adaptación al estrés.

A destacar, hay menor cantidad de genes de regulación transcripcional o traduccional en BL al día 28 seguido de BC al día 14 señalando posible mayor actividad.

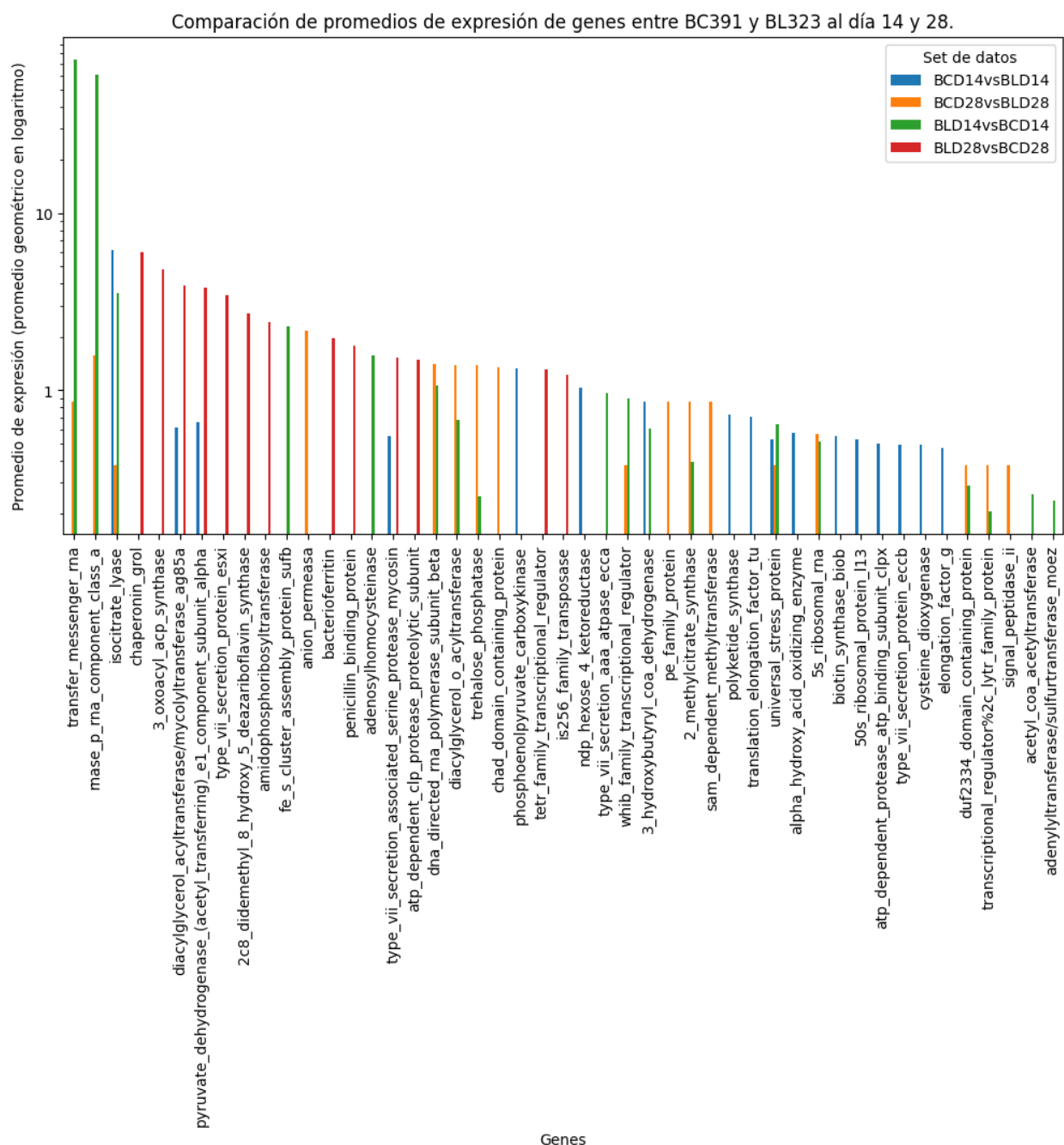


Figura 18: Comparación de los promedios geométricos en logaritmo de los 20 genes de mayor expresión relativa para BC391 y BL323 al día 14 y día 28 posterior a remoción de genes hipotéticos.

47 distintos genes fueron encontrados en los 80 genes provenientes de filtrar por expresión diferencial o Log2.

La figura 18 contiene la totalidad de la información de expresión diferencial. Para fines didácticos, la comparación de expresión genética entre cepas se realizó conforme a la clasificación del grupo de genes en: metabólicos, relacionados a transcripción o traducción y, en tercera instancia, de adaptabilidad y respuesta a estrés.

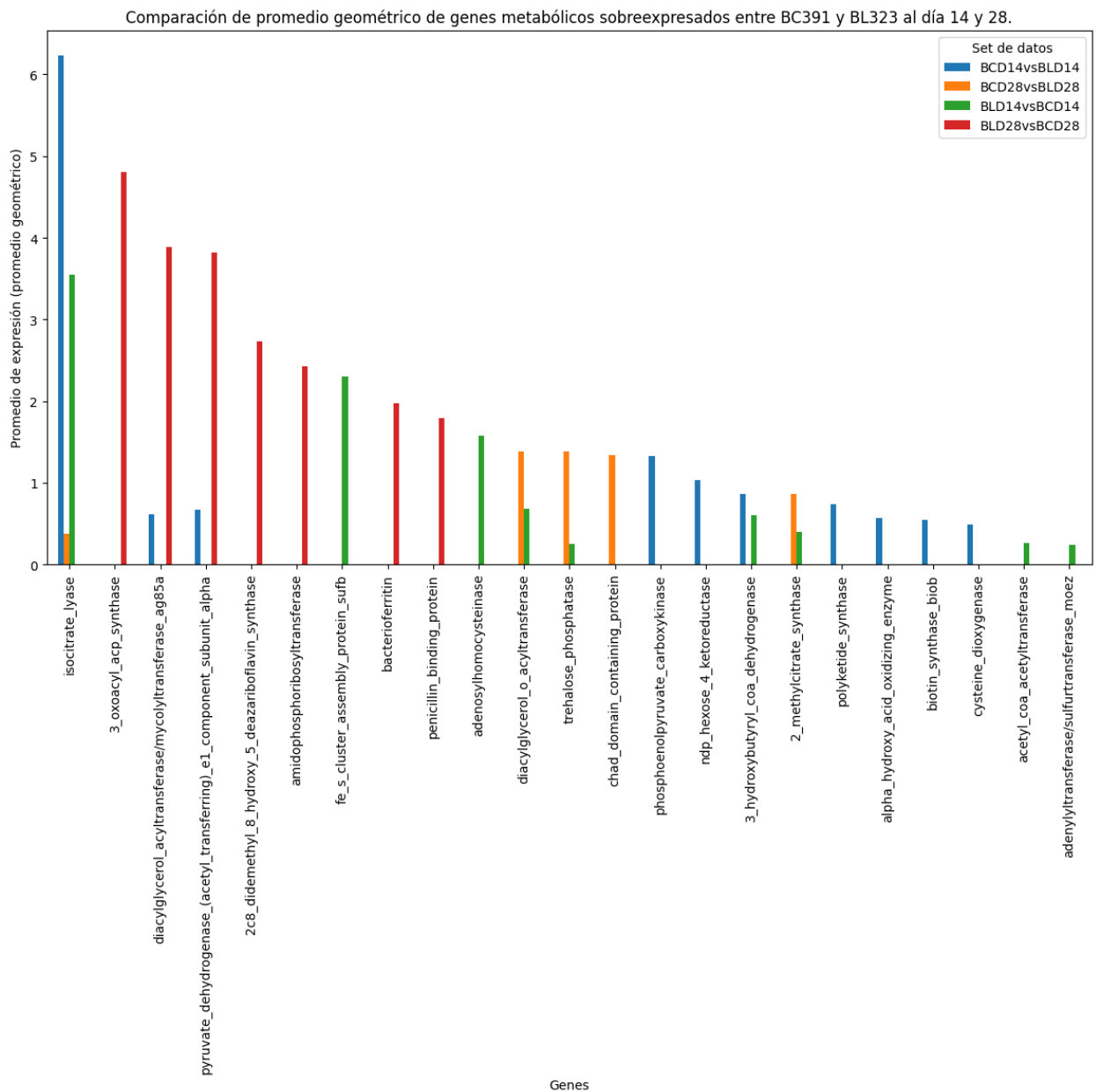


Figura 19: Gráfico de comparación del eje metabólico en genes sobreexpresados por las cepas BC391 y BL323 al día 14 y 28 . Nombres de los genes metabólicos de expresión diferencial en el eje x. En el eje Y una medición del promedio geométrico de expresión genética.

Para BC391 al día 10 de 20 genes diferencialmente expresados son del eje metabólico. BL323 al día 14, 9 de 20. BC391 al día 28, 5 de 20 mientras que BL323 al día 28 son 7 de 20. La predominancia de genes metabólicos al día 14 por BC y al día 28 por BL sugiere que al día 14 la cepa BC es más activa mientras que al día 28 BL es más activa que BC.

El gen mayormente conservado es isocitrato liasa. Un gen relacionado al homeostasis energético en condiciones de hipoxia (Ehrt et al. 2018). Particularmente observamos que BC al día 14 lo expresa en mayor medida y que BL al día 28 no lo expresa significativamente señalando una condición de mayor y menor estrés respectivamente.

El gráfico demuestra la expresión elevada de múltiples genes (barras en rojo) de BL al día 28 señalando mayor actividad transcripcional con su respectiva probable actividad biológica. Los genes de BL al día 28 son variados sintetizadores de nucleótidos (*purF*), ácidos grasos (*kas-A*, *Ag85*), producción energética (*ribD*, piruvato deshidrogenasa).

Se observa también un rezago transcripcional o expresión de genes similares entre BL d14 y BC d28 (barras verdes y amarillas juntas. Figura 19). La estrategia transcripcional entre BL d14 y BC d28 es compartida eficientizando la producción energética por *prpC* quelando minerales (BL por *sufb*; BC por CHAD) y expresando tanto trehalosa como síntesis grasa (*OtsA* y diacilglicerol-o-acil-transferasa).

Para entender el contexto y las diferencias entre la expresión de los genes metabólicos es útil enlistar en una tabla las descripciones de las funciones de los genes diferencialmente expresados en el eje metabólico y añadir si se sobreexpresa por cepa y día.

Tabla 11: Descripción de genes metabólicos expresados diferencialmente por *M. tuberculosis* BC y BL al día 14 y 28.

Gen	Gene ID	Función	BC391 D14	BC391 D28	BL323 D14	BL323 D28
Isocitrato liasa. icl1.	Rv0467 ; kegg .	Conversión de isocitrato en succinato y glioxilato en el ciclo del glioxilato.	Sí	Sí	Sí	No
Fosfoenolpiruvato carboxiquinasa. PckA.	Rv0211	Conversión de oxalacetato a fosfoenolpiruvato. Modula la tasa de gluconeogénesis.	Sí	No	No	No
Hexosa cetoreductasa	Rv2971 .	Metabolismo celular. Genera intermediarios en el metabolismo de carbohidratos.	Sí	No	No	No
Hidroxibutiril CoA deshidrogenasa. BHBD.	Rv0468 .	Conversión de hidroxibutiril-CoA a acetoacetato, metabolismo de ácidos grasos.	Sí	No	Sí	Sí
Sintasa de policétidos. Pks15	Rv2947c	Síntesis de policétidos fenol-tiocerol. Inhibidores de PRR.	Sí	No	No	No
Piruvato deshidrogenasa	Rv2215	Conversión de piruvato en acetil-CoA y CO ₂ para producción de energía y balance redox.	Sí	No	No	No
Micolyl transferasa AG85A	Rv3804c	Integridad de pared celular mediante biosíntesis de ácidos micólicos. Los ácidos micólicos sirven de intermediarios para los compuestos trehalosados.	Sí	No	No	Si
Enzima oxidante de alfa hidroxiácidos. MftD.	Rv0691A	Participa en la oxidación de ácidos grasos ramificados.	Sí	No	No	No
Sintasa de biotina. BioB.	Rv1589	Cofactor enzimático útil para la respiración.	Sí	No	No	No
Cisteína dioxigenasa. CysM.	Rv1336	Metabolismo de la cisteína. Contribuye al equilibrio redox, balance energético y resistencia a antibióticos.	Sí	No	No	No
Diacilglicerol-O-aciltransferasa	Rv2285	Metabolismo de lípidos. Añade un grupo acilo a un diacilglicerol para formar un triacilglicerol. Útil para	No	Si	Sí	No
Trehalosa fosfatasa. OtsA.	Rv1568	Producción de trehalosa a partir de glucosa. Refuerza la membrana contra fármacos y estrés osmótico.	No	Sí	Sí	No
Ensamblaje de clústeres Fe-S.	Rv1461	Los complejos hierro-azufre son componentes de un rango de proteínas energéticas, reparación de	No	No	Sí	No

SufB		ADN y de adaptabilidad a estrés oxidativo. Ejemplo: whiB. (Elchennawi, <i>et. al</i> , 2022)				
Metilcitrato sintasa. PrpC.	Rv1131	Respiración celular. Une propionil-CoA y oxaloacetato para formar metilcittrato.	No	Si	Sí	No
Adenosil homocisteinasa	Rv3248c	Conversión de adenosil homocisteína a homocisteína. Metabolismo de metionina parte de respiración celular.	No	No	Sí	No
Acetil CoA acetiltransferasa .	Rv1135A	Degradación de ácidos grasos, remueve un grupo acetil-CoA para su uso en TCA-respiración.	No	No	Sí	No
Oxoacil ACP sintasa. kasA	Rv2245	Síntesis de ácidos grasos y lípidos. Transfiere un grupo acetilo a un malonilo extendiendo el lípido.	No	No	No	Sí
2C8 didimetil 8 hidroxideazariboflavin sintasa. ribD	Rv2671	Biosíntesis de riboflavinas, cofactores esenciales en reacciones redox en la respiración celular.	No	No	No	Sí
Amidofosforibosiltransferasa. purF.	Rv0808	Biosíntesis de nucleótidos particularmente purinas, esenciales para transcripción genética pero desde los componentes.	No	No	No	Sí
Proteína con dominio CHAD	mtbc0_00236_4	Proteína quelante de metales. (Patel, 2017)	No	Sí	No	No
Bacterioferritina	Rv3841	Almacenamiento y liberación de hierro, esencial para la función celular. Útil para la respiración celular.	No	No	No	Sí
Adenil-sulfurtransferas a moeZ. Similar to moeB.	ARN07497.1	Relacionada con la biosíntesis del cofactor enzimático molibdopterina, que forma parte de enzimas de detoxificación, aprovechamiento de nitrógeno y azufre. Útil para producir genes de supervivencia en condiciones anaeróbicas como narG, NuoG y CDH (Williams et al. 2014).	No	No	Sí	No

Hipervínculos a la información en la segunda columnas y artículos que respaldan la información.
Información consultable en <https://mycobrowser.epfl.ch>.

Comparación de promedio geométrico de genes relacionados a transcripción o traducción sobreexpresados entre BC391 y BL323 al día 14 y 28.

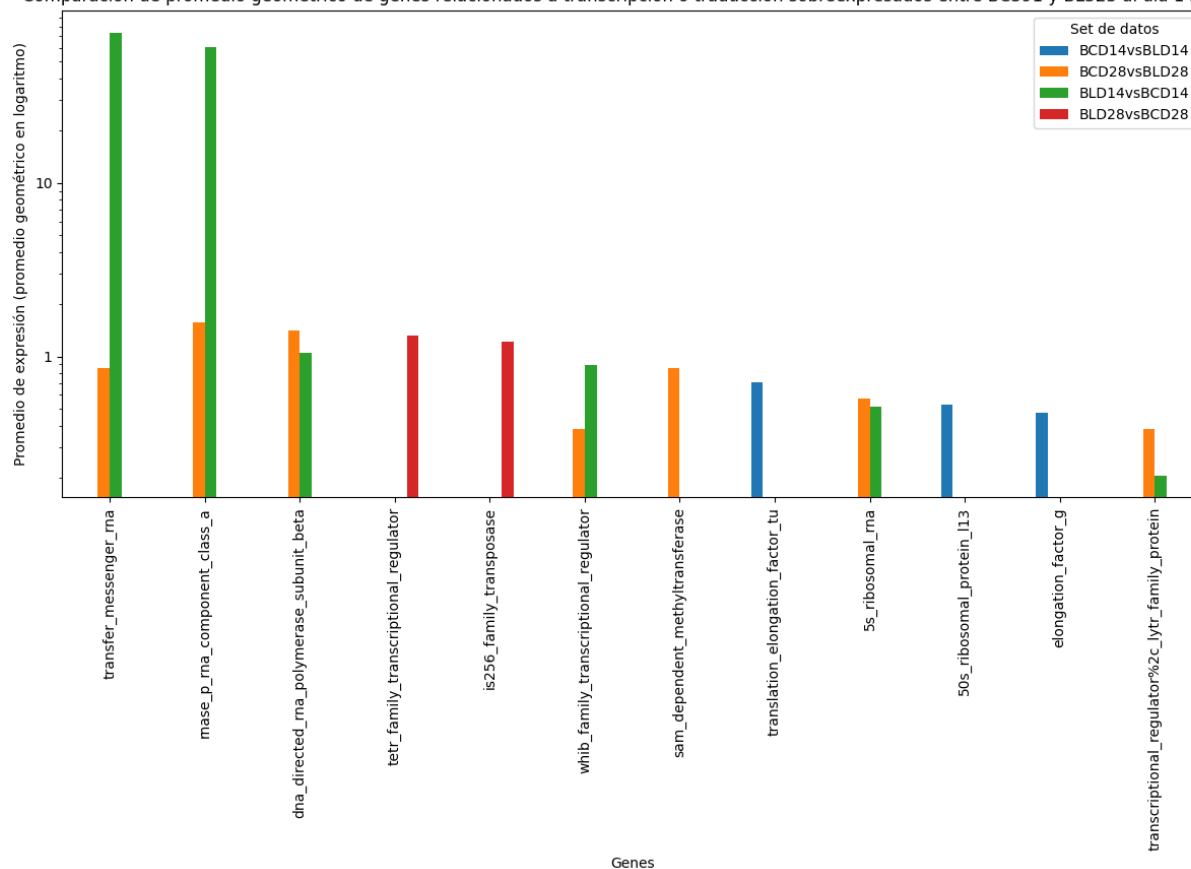


Figura 20: Gráfico de comparación del eje transcripción-traducción en genes sobreexpresados por las cepas BC391 y BL323 al día 14 y 28. El eje X: nombres de los genes de transcripción o traducción. El eje Y es el promedio geométrico de expresión genética en logaritmo.

De los 20 genes sobreexpresados por cepa y punto temporal, BC391 al día 14 sobreexpresa 3 de índole de transcripción o traducción. BL323 al día 14, por otra parte sobreexpresa 6 genes. BC391 y BL323 al día 28 expresan 7 y 2 respectivamente de sus 20 genes sobreexpresados. La predominancia de genes de transcripción y traducción son al día 14 BL323 y al día 28 BC391.

Se observa expresión de genes similares entre BL d14 y BC d28 (barras en verde y amarillo respectivamente en la figura 20). La estrategia transcripcional compartida en el eje de regulación transcripcional-traduccional consta del rescate de mRNA por ARN mensajeros de transferencia (*ssrA*, *SmpB*), excisión de mRNA etiquetado para degradación por ARNasaP (*rnpB*) y síntesis de ARN por polimerasa asociada a ribosomas (*rpoB*, *rrf*). Por otra parte, *whiB* y *lytr* se asocian al ADN para regular su expresión y protegerlo contra estrés oxidativo ambiental.

BC d14 aumenta la expresión transcripcional mediante proteínas de acoplamiento a ribosomas (*tuf*, *fusA* y *rrl*).

BL d28 contiene pocos genes de regulación transcripcional diferencialmente expresados solamente un regulador transcripcional de diversos genes (*tetR*) y una secuencia móvil *IS256*.

Tabla 12: Genes relacionados con transcripción y traducción.

Gen	NCBI ID	Gene Función	BC391 D14	BC391 D28	BL323 D14	BL323 D28
Factor de elongación Tu. <i>tuf</i>	Rv0685	Aumenta la traducción por acoplamiento de tRNAs cargados al ribosoma.	Sí	No	No	No
50S rRNA. <i>rrl</i>	MTB0000020	Subunidad mayor del ribosoma. Indispensable para la traducción.	Sí	No	No	No
Factor de elongación G. <i>fusA</i> .	Rv0684	Facilita el movimiento del ribosoma a lo largo del mRNA durante la traducción.	Sí	No	No	No
RNA mensajero de transferencia. <i>ssrA</i> , <i>SmpB</i> .	Rv3100c	Rescate de ribosomas mediante etiquetado de transcritos en proceso de traducción o sin codón de terminación debido a daño oxidativo.	No	Sí	Sí	No
RNasa P. <i>rnpB</i>	MTB000026	Realiza un corte sobre los tRNA. Maduración de tRNAs.	No	Sí	Sí	No
RNA polimerasa. <i>rpoB</i>	Rv0667	Cataliza la síntesis de RNA a partir de ADN en la transcripción.	No	Sí	Sí	No
<i>whiB</i>	Rv3199	Regulador transcripcional sensible a estrés oxidativo, modula genes antioxidantes, de virulencia y regulación metabólica. Ejemplos: CF10, ESAT6 (evitan apoptosis), DosR, Esx. (Elchennawi, <i>et. al</i> , 2022)	No	Sí	Sí	No
5S rRNA. <i>rrf</i>	MTB000021	Componente menor del ribosoma. Estabiliza al ribosoma para el proceso de traducción.	No	Sí	Sí	No
Regulador transcripcional <i>llytR</i> ; <i>CpsA</i> .	NCBI: ARN08640.1 ; Rv3484	Modula la expresión de genes de remodelamiento de la pared celular, evita la exposición de PAMPs, e inhibe el reclutamiento de la enzima oxidante NADPH al fagosome. (Koster, <i>et. al</i> , 2017)	No	Sí	Sí	No

SAM metiltransferasa. TrmD.	ARN04931.1 Rv2118c	Hay 25 SAM metiltransferasas. La más cercana a la identidad de 254 aminoácidos es TrmD (280 aa). Que metila específicamente RNAs aumentando su estabilidad y traducción (Berney, <i>et. al.</i> , 2015).	No	Sí	No	No
Regulador transcripcional tetR	Rv0144	Los reguladores de transcripción TetR son los más abundantes en micobacterias. Sensan múltiples formas de estrés y se desprenden del ADN facilitando su expresión. Involucrados en regular bombas de flujo, transporte de nutrientes, modulación de pared celular, defensa antioxidante y resistencia a antibióticos. (Balhana, <i>et. al.</i> , 2015)	No	No	No	Sí
Secuencia de inserción IS256	mtbc0_00365 0	Elemento genético móvil que puede alterar la expresión génica contribuyendo a variabilidad genética y adaptación. Asociado al desarrollo de resistencias a antibióticos.	No	No	No	Sí

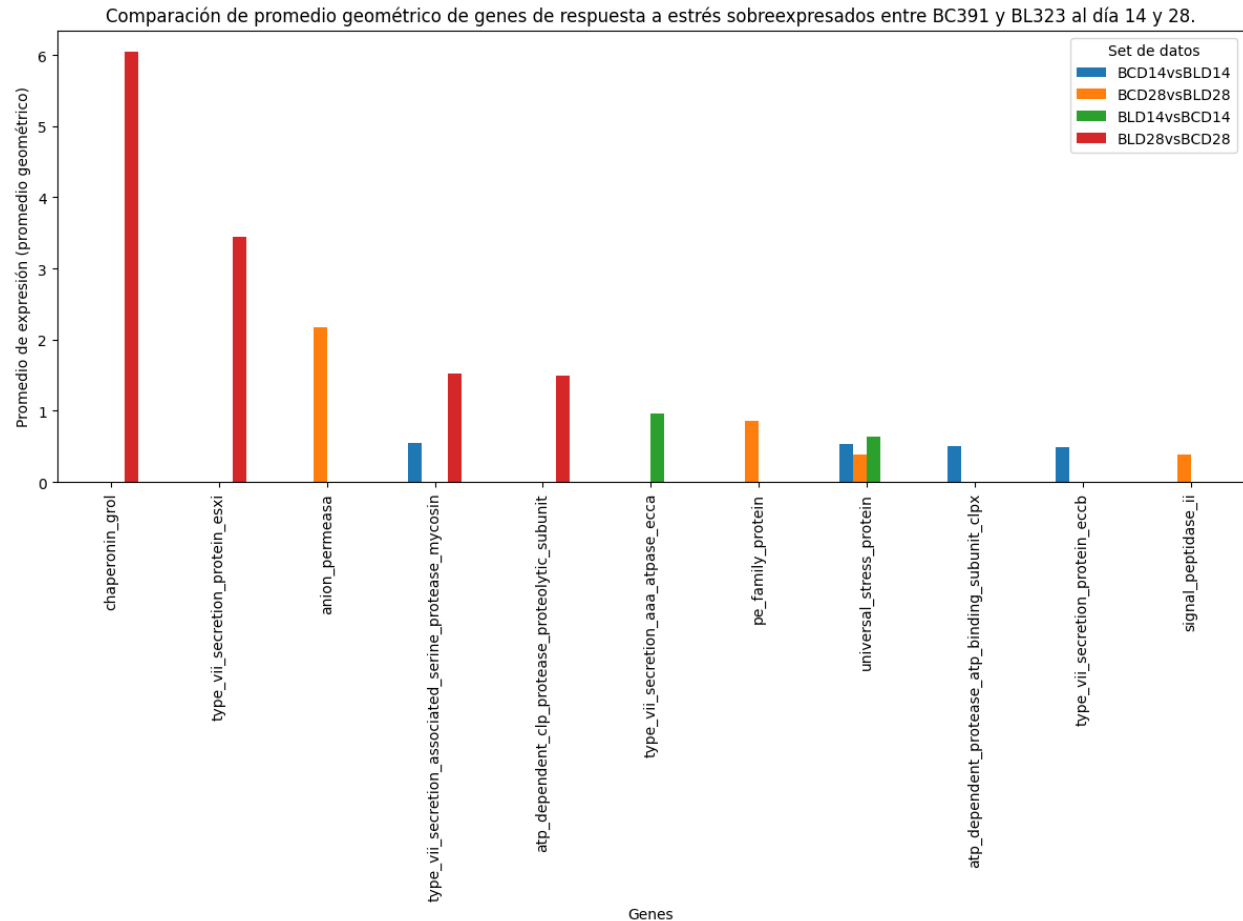


Figura 21: Gráfico de comparación del eje adaptabilidad y estrés en genes sobreexpresados por las cepas BC391 y BL323 al día 14 y 28. En el eje x los nombres de los genes de adaptabilidad diferencialmente expresados. 11 genes totales. El eje Y es el promedio de expresión genética en promedio geométrico.

BC391 al día 14 sobreexpresa 4 de sus 20 genes como genes de respuesta a estrés. BL323 al día 14 2 de sus 20 genes sobreexpresados. Al día 28 para BC391 y BL323 tienen 3 y 5 respectivamente. La resistencia al estrés predomina en BL d28, seguido por BC d14. El gen más conservado en sobreexpresión entre cepas y días es la proteína de respuesta a estrés universal, solamente no se expresa al día 28 por BL323.

BL d28 expresa asistencia de plegamiento de proteínas (chaperonina *groL*), proteasas (*mycP4*, *ClpX*) y *ESAT-6* un factor de virulencia asociado a disrupción de membranas, necrosis tisular y desarrollo de tuberculosis activa (De Martino et al. 2019; Francis et al. 2014).

Tabla 13: Genes sobreexpresión en BC391 y BL323 relacionados a estrés.

Gen	NCBI Gene ID	Función	BC391 D14	BC391 D28	BL323 D14	BL323 D28
USP (Universal Stress Protein)						
<i>TB31.7</i>	Rv2623	Captación de nutrientes.	Sí	Sí	Sí	No
	Rv3449 ; kegg: mycP4	Actividad de proteasa extracelular. Degradación y reciclado de proteínas.	Sí	No	No	Sí
<i>ClpX</i>	Rv2457c	Subunidad de proteasa que define la especificidad a sustrato, puede actuar como chaperona y es dependiente de ATP. Ayuda en la degradación de proteínas dañadas por estrés.	Sí	No	No	No
<i>EccB</i>	MT18B0358	Componente de los sistemas ESX fijado a membrana forma parte del canal de transporte, importante en la secreción de proteínas y lípidos.(Houben, <i>et. al.</i> , 2014)	Sí	No	No	No
<i>EccA</i>	Rv3806	Componente accesorio soluble del sistema ESX de secreción tipo 7. EccA contribuye a la secreción de proteínas y lípidos de membrana. Relacionado con virulencia y respuesta a estrés.(Houben, <i>et. al.</i> , 2014)	No	No	Sí	No
Anión permeasa.	Rv3679	Transportador de aniones al exterior de la célula. Relacionado con la homeostasis celular.	No	Sí	No	No
Chaperonina <i>groL</i>	Rv0440 Rv3417c	Previene mal plegamientos y apoya en el plegamiento de proteínas. Las proteínas se pliegan erroneamente en condiciones de estrés.	No	No	No	Sí
Proteína secretada por EsxI. <i>ESAT-6</i>	Rv1037c	Componente del sistema de secreción tipo 7 ESX-1, relacionado con virulencia y respuesta al estrés. (Francis, <i>et. al.</i> , 2014)	No	No	No	Sí
Penicillin-binding protein (<i>PBP</i>)	Rv3330	Participa en la síntesis de la pared celular particularmente del peptidoglicano, responde al estrés inducido por antibióticos.	No	No	No	Sí
Peptidasa de señal 2. <i>lepB</i>	Rv2903c	Procesa péptidos señal en proteínas secretadas para su maduración.	No	Sí	No	No
PE proteína. 160 variantes.	Rv1818c	Proteínas de pared celular que median adherencia e inmunoevasión pueden tener actividad regulatoria en genes. (Fishbein, <i>et. al.</i> , 2015)	No	Sí	No	No

IX. DISCUSIÓN

Construcción de librerías y sublibrerías

Las librerías de transcritos iniciales al día 14 fueron de 5.67 millones para BC391 y 5.25 millones para BL323, una variación de sólo 7%. Sin embargo, al momento de alinear los transcritos al genoma de *M. tuberculosis*, los transcritos al día 14 se redujeron a 4,965 para BC391 mientras que para BL323 fueron 33,670, es decir, 6.78 veces más muestra en BL323. Es posible que una cepa logre replicar más eficientemente en el modelo, generando más material extraíble al momento del muestreo. Contrario a la previa suposición en el mismo modelo al día 14, se han encontrado alineamientos cercanos entre ambas cepas (65 mil transcritos para BL323 y 77 mil transcritos para BC391), e incluso una mayor cantidad de ARN para la cepa de la cual se obtuvo casi 7 veces menos material transcripcional. Comparando los rendimientos de transcritos alineados a *M. tuberculosis* provenientes de extracción de ARN en el presente proyecto, hay una variación del 700% entre las cepas, mientras que en el punto de referencia por (Cerezo-Cortés et al. 2022) se reporta una variación del 16%. Por el momento, no se puede dilucidar la naturaleza de esta variación, ya sea error experimental o diferencias biológicas. Véase Tabla 5.

Comparar librerías con amplia variación en tamaño genera sesgos en la normalización, ya que los transcritos se multiplican por un factor de conversión para igualar la cantidad de muestra, haciendo que la normalización parezca aumentar la expresión relativa a favor de la muestra más pequeña. Se produce un efecto de dilución en la muestra de mayor tamaño para comparar la misma cantidad de transcritos. Actualmente, no existe un algoritmo de normalización adecuado para muestras que contengan una sublibrería dentro de una librería más grande. Este es un caso particular, producto de la naturaleza de internalización de *M. tuberculosis*. Para resolver esta problemática, se requiere tomar muestras con una cantidad comparable de ARN bacteriano o lograr lo establecido por (Cornejo-Granados et al. 2021) , quienes plantean la obtención de ARN abundante de la micobacteria mediante lisis diferencial capaz de separar el contenido de las células eucariotas o del hospedador.

Diseño experimental

En el presente estudio se analizaron 48 muestras transcriptómicas basadas en los niveles relativos de expresión génica, utilizando el cambio $\log_2$ en la razón de expresión como métrica para identificar genes diferencialmente expresados. Veinticuatro muestras provinieron de ratones infectados con una cepa moderna de Beijing murina (BL) y las otras 24 de ratones infectados con

una cepa clásica de Beijing (BC). Para cada cepa, se recolectaron 12 muestras al día 14 post-infección (dpi) y 12 al día 28 dpi. Se seleccionaron un total de 80 genes, representando 47 genes no redundantes (véase Tabla 1), organizados en tres categorías: genes metabólicos, genes relacionados con la regulación transcripcional o traduccional, y genes relacionados con la respuesta al estrés, con el propósito de visualizar los contrastes de expresión génica entre cepas.

Estrategias metabólicas de BC y BL

Tanto al día 14 como al 28, se analizaron diferencias en la expresión de genes metabólicos, de regulación y asociados al estrés. Al día 14, ambas cepas exhibieron acumulación de intermediarios glucolíticos y una alta expresión del biomarcador de estrés *icl1*, aunque esto fue más pronunciado en BC (Figura 20). BC mostró alta expresión de la enzima fosfoenolpiruvato carboxiquinasa (PepCK) (Tabla 11), facilitando la conversión de intermediarios del ciclo de Krebs en piruvato, el cual puede alimentar posteriormente la gluconeogénesis o la ruta de las pentosas fosfato.

En contraste, BL demostró niveles más bajos de *icl1* y un flujo alternativo de carbono a través del ciclo del metilcitrato, integrando carbono en la rama reductora del ciclo de Krebs y reduciendo la dependencia de *icl1*, lo cual representa una sutil adaptación que mejora su tolerancia al estrés. A pesar de operar bajo menor disponibilidad de carbono, BL prioriza la producción de trehalosa, aunque esto no siempre se refleja en una sobreexpresión evidente de enzimas específicas de la gluconeogénesis.

BL también mostró sobreexpresión de la enzima sintetasa de triglicéridos, particularmente diacilglicerol-O-aciltransferasa, consistente con lo reportado por (Tong et al. 2020) quienes hallaron que las cepas modernas acumulan aproximadamente 2.6 veces más triglicéridos que las clásicas. La acumulación de intermediarios glucolíticos, trehalosa y triglicéridos es típica de micobacterias sometidas a estrés (Ehrt et al. 2018).

Una diferencia metabólica clave radica en que BL acumula trehalosa, mientras que BC muestra mayor expresión del complejo Ag85 y policétidos. Tanto la trehalosa como Ag85 desempeñan funciones cruciales en las interacciones patógeno-hospedador, desencadenando

respuestas inmunes distintas. Ag85 estimula receptores TCR, promoviendo una respuesta inmunitaria Th1. Los linfocitos reclutados a través de esta vía contribuyen a la contención de la infección al generar un entorno rico en IFN- γ , que potencia la capacidad bactericida de los macrófagos (Ehrt et al. 2018; Schick et al. 2020; Wang et al. 2023).

Trehalosa: acumulación y funciones

La trehalosa, un disacárido compuesto por dos moléculas de glucosa, funciona tanto como fuente de carbono como agente protector frente al estrés en el citoplasma y en la pared celular. Puede inhibir la fusión del fagosoma con el lisosoma y proteger proteínas, membranas y ADN frente al estrés oxidativo, vinculándose a la persistencia y resistencia a antibióticos (Kalera et al. 2024; Nobre et al. 2014; Vanaporn y Titball 2020)

Más allá de su rol protector, la trehalosa estimula de manera robusta el reclutamiento de células inmunes innatas mediante la activación del receptor Mincle, una lectina tipo C que desencadena la activación de células mieloides y promueve la respuesta inmunitaria adaptativa Th17 a través de una vía dependiente de Syk (Schick et al. 2020). La trehalosa contribuye a la formación temprana de granulomas, lo cual es en principio protector, pero también puede constituir un nicho permisivo para la replicación bacteriana. Sin embargo, la respuesta Th17, caracterizada por citoquinas como TNF, IL-6 e IL-17, puede inducir inflamación sistémica. Esta respuesta estimula quimiocinas como CXCL1 y CXCL5, generando una infiltración masiva de células mieloides que podría desestabilizar la integridad del granuloma y el equilibrio protector (Lombard et al. 2016). Aún no está claro si la composición del granuloma o la magnitud de la respuesta de células T difiere entre las cepas analizadas, aunque se han reportado eventos tempranos de cavitación en infecciones por la cepa BL (Cerezo et al. 2021).

El reclutamiento de células mieloides inducido por la trehalosa —principal nicho replicativo de *M. tuberculosis*— junto con la acumulación de fuentes de carbono, podría subyacer a la mayor tasa de crecimiento observada en BL comparada con BC (Cerezo et al. 2021). Cabe destacar que, aunque las cepas modernas de Beijing suelen acumular más trehalosa, BC también expresa cantidades significativas (Figura 20), aunque con un retraso considerable en relación a BL. La exploración de puntos temporales más tempranos podría refinar los perfiles metabólicos específicos de cada cepa (Tong et al. 2022).

Interacciones inmunes hospedador-patógeno

Más allá de las estrategias metabólicas, *M. tuberculosis* explota células mieloides para persistir y escapar del confinamiento impuesto por el granuloma. De los DEGs, ESAT-6 funciona como un factor de virulencia que desestabiliza las membranas fagosomales y celulares, permitiendo la fuga bacteriana al citosol y contribuyendo tanto a la activación inmune como a la destrucción tisular. Las células inmunes innatas, especialmente los neutrófilos, pueden ocasionar un daño tisular considerable, contribuyendo a manifestaciones clínicas como la hemoptisis (De Martino et al. 2019).

Convergencia dinámica de perfiles metabólicos en el tiempo

Muchos genes relacionados con los ejes metabólicos se expresaron de forma similar entre BC al día 28 y BL al día 14, sugiriendo una progresión patogénica retrasada o una firma transcriptómica más tardía en BC. Por ejemplo, BC al día 28 expresó enzimas biosintéticas de trehalosa y genes de síntesis de triglicéridos, así como enzimas de la vía del metilcitrato para la incorporación adicional de carbono al ciclo de Krebs. Sin embargo, a diferencia de BL, BC no mostró una alta expresión de mecanismos de secuestro de clusters hierro-azufre necesarios para la producción de enzimas clave.

A pesar de cierta convergencia tras un retraso de 14 días, la expresión de *whiB*, un regulador transcripcional crucial y protector del ADN frente al estrés oxidativo (Tabla 12), permaneció significativamente más alta en BL.

Contraste en la aptitud de supervivencia entre cepas

Además de la acumulación de trehalosa como mecanismo protector, BL expresa significativamente la enzima de quelación de clusters Fe-S *SufB*, promoviendo la producción de *WhiB*. *WhiB* protege el ADN bajo estrés oxidativo y es clave para la supervivencia de BL en un entorno inmunológicamente activado. Todo apunta a que BL está altamente adaptada para sobrevivir en medio de una respuesta inmune innata activa, explotando la inmunidad del hospedador para crear condiciones que favorecen tanto la persistencia bacteriana como el daño tisular mediado por la respuesta inmune.

Durante esta fase crítica, el sistema inmune intenta orientar la respuesta hacia la reparación tisular. Sin embargo, las interacciones entre especies reactivas de oxígeno (ROS) y WhiB pueden provocar la disociación WhiB-ADN, permitiendo la transcripción y traducción de factores de virulencia como ESAT-6. De hecho, ESAT-6 mostró una alta expresión en BL al día 28 (Figura 4). Su actividad favorece la disrupción de membranas, la cavitación del granuloma, la necrosis y la producción compensatoria de IL-10 (De Martino et al. 2019).

Esta cascada permite que las cepas modernas de Beijing adopten un estado metabólico más relajado, reduciendo la acumulación de trehalosa y orientándose hacia la remodelación de la pared celular, reflejada en niveles elevados de Ag85 y en un aumento de la actividad metabólica (síntesis de lípidos, nucleótidos y aminoácidos; Tabla 11, barras rojas en Figura 20), lo cual sustenta finalmente una mayor capacidad replicativa.

X. CONCLUSIÓN

Se identificaron genes diferencialmente expresados relacionados con la virulencia en las cepas BC y BL de *Mycobacterium tuberculosis*, mostrando patrones transcripcionales distintos a los días 14 y 28. Las diferencias observadas en la expresión de moléculas de anclaje a membrana, como la sintasa de ácidos micólicos y policétidos en BC y la acumulación de trehalosa en BL, podrían estar implicadas en la modulación de la respuesta inmune del hospedador. Además, la activación del sistema de secreción tipo VII en BC y la adaptación metabólica de BL —incluyendo la quelación de minerales y la expresión de whiB— apuntan a estrategias diferenciadas para la supervivencia intracelular y evasión inmune.

Al comparar ambos puntos temporales, se observó que BC al día 28 replica parcialmente el perfil transcripcional de BL al día 14, lo que sugiere un rezago en su estrategia patogénica. Por último, la expresión de ESAT6 por parte de BL al día 28 podría estar vinculada a la transición hacia tuberculosis activa.

Este estudio no solo aporta nuevas perspectivas sobre la diversidad funcional entre cepas clínicas de *M. tuberculosis*, sino que además establece un algoritmo de análisis transcriptómico aplicable a futuras investigaciones en caracterización de infecciones, inmunobiología y virulencia bacteriana.

XI. PERSPECTIVAS

El presente proyecto puede ser continuado mediante el análisis comparativo de la expresión diferencial en un modelo murino, con el objetivo de identificar cómo varía la respuesta inmunitaria frente a cada cepa. Explorar las diferencias en la composición de la membrana bacteriana en distintos días post infección podría aportar información clave sobre cómo estas variaciones estructurales modulan la interacción con el sistema inmune. Asimismo, diseñar experimentos que evalúen el reclutamiento de células inmunes y la producción de citocinas permitiría validar la hipótesis de que las cepas inducen respuestas inmunológicas diferenciadas: BC activando predominantemente al sistema inmune adaptativo, y BL al sistema inmune innato.

XII. REFERENCIAS

- Al-Asady, I N, y Ali, J F. 2023. Review Article: Virulence Factors of Mycobacterium Tuberculosis. Journal for Research in Applied Sciences and Biotechnology, 2(3), Article 3. <https://doi.org/10.55544/jrasb.2.3.31>
- Bakuła, Z, Dziurzyński, M, Decewicz, P, Bakonytė, D, Vasiliauskaitė, L, Nakčerienė, B, Krenke, R, Stakėnas, P, y Jagielski, T. 2023. Spoligotyping of *Mycobacterium tuberculosis* – Comparing *in vitro* and *in silico* approaches. Infection, Genetics and Evolution, 115, 105508. <https://doi.org/10.1016/j.meegid.2023.105508>
- Bussi, C, y Gutierrez, M G. 2019. Mycobacterium tuberculosis infection of host cells in space and time. FEMS Microbiology Reviews, 43(4), 341–361. <https://doi.org/10.1093/femsre/fuz006>
- Cerezo, M, Rodríguez, J, López, G, Mata, D, y Hernández, R. 2021. Profiling the immune response to *Mycobacterium tuberculosis* Beijing family infection: a perspective from the transcriptome. Virulence, 12(1), 1689–1704. <https://doi.org/10.1080/21505594.2021.1936432>
- Cerezo-Cortés, M I, Rodríguez-Castillo, J G, Mata-Espinosa, D A, Bini, E I, Barrios-Payan, J, Zatarain-Barrón, Z L, Anzola, J M, Cornejo-Granados, F, Ochoa-Leyva, A, Del Portillo, P, Murcia, M I, y Hernández-Pando, R. 2022. Close Related Drug-Resistance Beijing Isolates of Mycobacterium tuberculosis Reveal a Different Transcriptomic Signature in a Murine Disease Progression Model. International Journal of Molecular Sciences, 23(9),

5157. <https://doi.org/10.3390/ijms23095157>
- Chandra, P, y Grigsby, S. 2022. *Immune evasion and provocation by Mycobacterium tuberculosis* | *Nature Reviews Microbiology*. <https://www.nature.com/articles/s41579-022-00763-4>
- Chang, D P S, y Guan, X L. 2021. Metabolic Versatility of Mycobacterium tuberculosis during Infection and Dormancy. *Metabolites*, 11(2), Article 2. <https://doi.org/10.3390/metabo11020088>
- Cornejo-Granados, F, López-Leal, G, Mata-Espinosa, D A, Barrios-Payán, J, Marquina-Castillo, B, Equihua-Medina, E, Zatarain-Barrón, Z L, Molina-Romero, C, Hernández-Pando, R, y Ochoa-Leyva, A. 2021. Targeted RNA-Seq Reveals the M. tuberculosis Transcriptome from an In Vivo Infection Model. *Biology*, 10(9), 848. <https://doi.org/10.3390/biology10090848>
- De Martino, M, Lodi, L, Galli, L, y Chiappini, E. 2019. Immune Response to Mycobacterium tuberculosis: A Narrative Review. *Frontiers in Pediatrics*, 7, 350. <https://doi.org/10.3389/fped.2019.00350>
- Dorhoi, A, Reece, S T, y Kaufmann, S H E. 2011. For better or for worse: the immune response against *Mycobacterium tuberculosis* balances pathology and protection. *Immunological Reviews*, 240(1), 235–251. <https://doi.org/10.1111/j.1600-065X.2010.00994.x>
- Ehrt, S, Schnappinger, D, y Rhee, K Y. 2018. Metabolic principles of persistence and pathogenicity in Mycobacterium tuberculosis. *Nature Reviews Microbiology*, 16(8), 496–507. <https://doi.org/10.1038/s41579-018-0013-4>
- Ferraris, D M, Miggiano, R, Rossi, F, y Rizzi, M. 2018. Mycobacterium tuberculosis Molecular Determinants of Infection, Survival Strategies, and Vulnerable Targets. *Pathogens*, 7(1), Article 1. <https://doi.org/10.3390/pathogens7010017>
- Francis, R J, Butler, R E, y Stewart, G R. 2014. Mycobacterium tuberculosis ESAT-6 is a leukocidin causing Ca²⁺ influx, necrosis and neutrophil extracellular trap formation. *Cell*

- Death & Disease, 5(10), e1474–e1474. <https://doi.org/10.1038/cddis.2014.394>
- Gagneux, S. 2018. Ecology and evolution of *Mycobacterium tuberculosis*. *Nature Reviews Microbiology*, 16(4), 202–213. <https://doi.org/10.1038/nrmicro.2018.8>
- Gauba, K, Gupta, S, Shekhawat, J, Sharma, P, Yadav, D, y Banerjee, M. 2021. Immunomodulation by epigenome alterations in *Mycobacterium tuberculosis* infection. *Tuberculosis*, 128, 102077. <https://doi.org/10.1016/j.tube.2021.102077>
- Getahun, H, Matteelli, A, Abubakar, I, Aziz, M A, Baddeley, A, Barreira, D, Den Boon, S, Borroto Gutierrez, S M, Bruchfeld, J, Burhan, E, Cavalcante, S, Cedillos, R, Chaisson, R, Chee, C B-E, Chesire, L, Corbett, E, Dara, M, Denholm, J, De Vries, G, ... Raviglione, M. 2015. Management of latent *Mycobacterium tuberculosis* infection: WHO guidelines for low tuberculosis burden countries. *European Respiratory Journal*, 46(6), 1563–1576. <https://doi.org/10.1183/13993003.01245-2015>
- Horsburgh, C R, Barry, C E, y Lange, C. 2015. Treatment of Tuberculosis. *New England Journal of Medicine*, 373(22), 2149–2160. <https://doi.org/10.1056/NEJMr1413919>
- Kak, G, Tiwari, B K, Singh, Y, y Natarajan, K. 2020. Regulation of Interferon- γ receptor (IFN- γ R) expression in macrophages during *Mycobacterium tuberculosis* infection. *Biomolecular Concepts*, 11(1), 76–85. <https://doi.org/10.1515/bmc-2020-0006>
- Kalera, K, Liu, R, Lim, J, Pathirage, R, Swanson, D H, Johnson, U G, Stothard, A I, Lee, J J, Poston, A W, Woodruff, P J, Ronning, D R, Eoh, H, y Swarts, B M. 2024. Targeting *Mycobacterium tuberculosis* Persistence through Inhibition of the Trehalose Catalytic Shift. *ACS Infectious Diseases*, 10(4), 1391–1404. <https://doi.org/10.1021/acsinfecdis.4c00138>
- Kiazyk, S, y Ball, T. 2017. Latent tuberculosis infection: An overview. *Canada Communicable Disease Report*, 43(3–4), 62–66. <https://doi.org/10.14745/ccdr.v43i34a01>
- Kroesen, V M, Madacki, J, Frigui, W, Sayes, F, y Brosch, R. 2019. *Mycobacterial virulence: impact on immunogenicity and vaccine research*. *F1000Research*, 8, 2025.

- <https://doi.org/10.12688/f1000research.20572.1>
- Liao, Y, Smyth, G K, y Shi, W. 2019. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research*, 47(8), e47–e47. <https://doi.org/10.1093/nar/gkz114>
- Lombard, R, Doz, E, Carreras, F, Epardaud, M, Le Vern, Y, Buzoni-Gatel, D, y Winter, N. 2016. IL-17RA in Non-Hematopoietic Cells Controls CXCL-1 and 5 Critical to Recruit Neutrophils to the Lung of Mycobacteria-Infected Mice during the Adaptive Immune Response. *PLOS ONE*, 11(2), e0149455. <https://doi.org/10.1371/journal.pone.0149455>
- Murray, J F, Schraufnagel, D E, y Hopewell, P C. 2015. Treatment of Tuberculosis. A Historical Perspective. *Annals of the American Thoracic Society*, 12(12), 1749–1759. <https://doi.org/10.1513/AnnalsATS.201509-632PS>
- Nobre, A, Alarico, S, Maranha, A, Mendes, V, y Empadinhas, N. 2014. The molecular biology of mycobacterial trehalose in the quest for advanced tuberculosis therapies. *Microbiology*, 160(8), 1547–1570. <https://doi.org/10.1099/mic.0.075895-0>
- Raveslout-Chávez, M M, Van Dis, E, y Stanley, S A. 2021. The Innate Immune Response to *Mycobacterium tuberculosis* Infection. *Annual Review of Immunology*, 39(1), 611–637. <https://doi.org/10.1146/annurev-immunol-093019-010426>
- Sayes, F, Pawlik, A, Frigui, W, Gröschel, M I, Crommelynck, S, Fayolle, C, Cia, F, Bancroft, G J, Bottai, D, Leclerc, C, Brosch, R, y Majlessi, L. 2016. CD4+ T Cells Recognizing PE/PPE Antigens Directly or via Cross Reactivity Are Protective against Pulmonary *Mycobacterium tuberculosis* Infection. *PLOS Pathogens*, 12(7), e1005770. <https://doi.org/10.1371/journal.ppat.1005770>
- Schick, J, Schäfer, J, Alexander, C, Dichtl, S, Murray, P J, Christensen, D, Sorg, U, Pfeffer, K, Schleicher, U, y Lang, R. 2020. Cutting Edge: TNF Is Essential for Mycobacteria-Induced MINCLE Expression, Macrophage Activation, and Th17 Adjuvanticity. *The Journal of Immunology*, 205(2), 323–328. <https://doi.org/10.4049/jimmunol.2000420>

- Talbot, E A, y Raffa, B J. 2015. *Mycobacterium tuberculosis*. En *Molecular Medical Microbiology* pp. 1637–1653. Elsevier. <https://doi.org/10.1016/B978-0-12-397169-2.00092-5>
- Tong, J, Liu, Q, Wu, J, Jiang, Y, Takiff, H E, y Gao, Q. 2020. *Mycobacterium tuberculosis* strains of the modern Beijing sublineage excessively accumulate triacylglycerols *in vitro*. *Tuberculosis*, 120, 101892. <https://doi.org/10.1016/j.tube.2019.101892>
- Tong, J, Meng, L, Bei, C, Liu, Q, Wang, M, Yang, T, Takiff, H E, Zhang, S, Gao, Q, Wang, C, y Yan, B. 2022. Modern Beijing sublineage of *Mycobacterium tuberculosis* shift macrophage into a hyperinflammatory status. *Emerging Microbes & Infections*, 11(1), 715–724. <https://doi.org/10.1080/22221751.2022.2037395>
- Vanaporn, M, y Titball, R W. 2020. Trehalose and bacterial virulence. *Virulence*, 11(1), 1192–1202. <https://doi.org/10.1080/21505594.2020.1809326>
- Wang, N, Liang, Y, Ma, Q, Mi, J, Xue, Y, Yang, Y, Wang, L, y Wu, X. 2023. Mechanisms of *ag85a/b* DNA vaccine conferred immunotherapy and recovery from *Mycobacterium tuberculosis*-induced injury. *Immunity, Inflammation and Disease*, 11(5). <https://doi.org/10.1002/iid3.854>
- Williams, M, Mizrahi, V, y Kana, B D. 2014. Molybdenum cofactor: A key component of *Mycobacterium tuberculosis* pathogenesis? *Critical Reviews in Microbiology*, 40(1), 18–29. <https://doi.org/10.3109/1040841X.2012.749211>