



UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO



FACULTAD DE QUÍMICA

PROGRAMA DE MAESTRÍA Y DOCTORADO
EN CIENCIAS BIOQUÍMICAS

CARACTERIZACIÓN GENÓMICA DE
M. bovis BCG MÉXICO

T E S I S

QUE PARA OBTENER EL GRADO DE :
DOCTOR EN CIENCIAS (BIOQUÍMICAS)

P R E S E N T A :

PATRICIA ORDUÑA ESTRADA



Tutor: DRA. YOLANDA LÓPEZ VIDAL

MÉXICO, D. F.

FEBRERO 2012



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

CARACTERIZACIÓN GENÓMICA DE *M. bovis* BCG MÉXICO

RECONOCIMIENTOS

Esta tesis doctoral se realizó bajo la dirección del Dra. Yolanda López Vidal en el programa de inmunología molecular microbiana, del Departamento de Microbiología y Parasitología de la Facultad de Medicina de la Universidad Nacional Autónoma de México.

El Comité Tutorial que asesoró el desarrollo de esta tesis estuvo formado por:

Dra. Yolanda López Vidal	Facultad de Medicina, UNAM
Dra. Maricarmen Quirasco Baruch	Facultad de Química, UNAM
Dr. Enrique Merino Pérez	Instituto de Biotecnología, UNAM

Se reconoce la colaboración del Dr. Miguel Ángel Cevallos Gaos, del Departamento de Genómica Evolutiva del Centro de Ciencias Genómicas de la UNAM, en cuyo laboratorio se llevó a cabo una parte de la secuenciación del genoma.

El proyecto fue apoyado parcialmente por CONACYT (2007-C01-68729) y PAPIIT-UNAM (IN227611-2). Durante los estudios de doctorado goce de una beca otorgada por CONACYT para la realización de la presente tesis.

Esta tesis fue defendida en examen presentado el día febrero del 2012

El Jurado de Examen Doctoral estuvo constituido por:

Presidente	Dr. Enrique Merino Pérez	Instituto de Biotecnología, UNAM
Vocal	Dra. Bertha González Pedrajo	Instituto de Fisiología Celular, UNAM
Vocal	Dr. Luis Servín González	Instituto de Inv. Biomédicas, UNAM
Vocal	Dr. Dimitris Georgellis	Instituto de Fisiología Celular, UNAM
Secretario	Dr. Raúl Mancilla Jiménez	Instituto de Inv. Biomédicas, UNAM

AGRADECIMIENTOS

Quiero agradecer a todos las personas que de una u otra manera fueron importantes para que éste trabajo llegará a buen término, mencionar a cada una de ellas me resulta imposible.

Agradezco de manera muy especial a la Dra. Yolanda López Vidal, quien me apoyo siempre a lo largo de todo este tiempo y cuya asesoría fue muy valiosa para poder realizar este proyecto. Muchas gracias por los consejos.

También agradezco al Dr. Miguel Ángel Cevallos Gaos que siempre se mostró dispuesto a colaborar con nosotros permitiéndome realizar una estancia en su laboratorio, asesorándome en todo momento.

Además agradezco el apoyo brindado por Adriana Arvizú, Humberto Barrios Camacho e Ismael Hernández en diferentes etapas de este proyecto, su ayuda resultó invaluable.

De manera muy especial agradezco a mis papás y mis hermanas por todo el apoyo que me han brindado durante estos años. Gracias por la paciencia, por escucharme, aunque de repente no supieran de que les hablaba. Gracias por estar siempre ahí.

Finalmente agradezco a todos mis compañeros del laboratorio, a los que están y los que ya salieron. Me gustaría mencionarlos a todos, pero me arriesgo a olvidar un nombre. De todos aprendí algo y siempre recibí apoyo y la amistad de muchos, todos ustedes hicieron agradable mi estancia en el laboratorio .

ÍNDICE

Resumen	V
Abstract	VI
Abreviaturas	VII
I. ANTECEDENTES	1
I.I El complejo <i>Mycobacterium tuberculosis</i>	2
I.II Epidemiología de la tuberculosis	3
Epidemiología de la tuberculosis en México	4
I.III Prevención y control de la tuberculosis	6
I.IV La vacuna BCG	6
Desarrollo de la vacuna BCG	8
Diferencias en eficacia protectora	9
Diferencias morfológicas de las BCGs.....	11
Diferencias bioquímicas de las BCGs	11
Diferencias inmunológicas de las BCGs	12
Diferencias genómicas de las BCGs	13
Diferencias en expresión proteica de las BCGs.....	16
I.V Uso y producción de la vacuna BCG en México	17
II. PLANTEAMIENTO DEL PROBLEMA	19
III. HIPÓTESIS	20
IV. OBJETIVOS	21
V. MATERIALES Y MÉTODOS	22
V.I Cepas empleadas en el estudio	23
V.II Extracción de DNA	23
V.III Determinación del perfil de regiones RD y DU	24
Determinación del perfil de regiones RD por PCR múltiple	24
Determinación del perfil de regiones DU por PCR	25
V.IV Determinación del perfil genético de la BCG México 1931 por PFGE	27
V.V Determinación del genoma de la BCG México 1931	27
Secuenciación	28

Ensamble	28
Cierre	29
Anotación	31
V.VI Análisis del genoma	31
V.VII Identificación de nuevas regiones RD en otras subcepas BCG	32
VI. RESULTADOS	33
VI.I Perfil genético de regiones RD y DU de BCG México	34
Perfil genético de regiones RD	34
Perfil genético de regiones DU	36
VI.II Perfil genético de la BCG México 1931 por PFGE	37
VI.III Secuenciación del genoma completo de la BCG México 1931	39
Secuenciación	39
Ensamble	40
Cierre del genoma	44
Anotación	47
VI.IV Comparación del genoma de BCG México 1931 con BCG Pasteur y Tokio.....	50
Identificación de nuevas regiones RD en otras BCGs	53
VII. DISCUSIÓN DE RESULTADOS	54
VIII. CONCLUSIONES	64
IX. PERSPECTIVAS	65
X. REFERENCIAS	66
XI. ANEXO A (MEDIOS Y TÉCNICAS)	73
XII. ANEXO B (ARTICULO PUBLICADO)	76

RESUMEN

El bacilo de Calmette y Guérin (BCG) es la única vacuna utilizada actualmente para prevenir la tuberculosis. Sin embargo, presenta una alta variabilidad en la protección que confiere contra la tuberculosis pulmonar en adultos (0-80%), una de las explicaciones para esta variabilidad es que la vacuna se produce a partir de diferentes subcepas BCG las cuales presentan diferencias entre sí. La BCG México no ha sido incluida en los estudios de caracterización de las BCGs realizados anteriormente, por lo que nuestro objetivo fue realizar una caracterización genómica de la cepa.

La caracterización de la cepa se realizó determinando el perfil de regiones RD y DU y posteriormente secuenciando el genoma completo. Los resultados obtenidos mostraron que la BCG México 1931 posee un cromosoma circular de 4 350 386 pb con un contenido de G+C del 65%. Se identificaron un total de 3 981 genes, incluyendo 3904 secuencias codificantes, 3 rRNA, 45 tRNA y 29 pseudogenes.

El análisis del genoma de la BCG México 1931 y el perfil de regiones RD y DU obtenido por PCR confirmó la pérdida de las regiones RD1, RD2, N-RD18 y de una copia de la secuencia de inserción IS6110, así como la presencia de la duplicación DU2 en la cepa. Lo cual la ubica dentro del grupo DU2 tipo IV en la genealogía de subcepas BCG.

Por otra parte, la comparación del genoma de la BCG México 1931 con los de BCG Pasteur y Tokio permitió identificar tres nuevas regiones RD de 53 (RDMex01), 655 (RDMex02) y 2 847 pb (RDMex03). Resultando las RDMex01 y RDMex02 únicas para BCG México 1931. Con la identificación 37 SNPs con mutaciones no sinónimas.

En general, los resultados obtenidos mostraron que la BCG México 1931 tiene un genoma de menor tamaño en comparación con BCG Pasteur y Tokio, y que la RDMex02 y RDMex03 son específicas para ésta. La RDMex02 contiene genes implicados en la unión de la BCG a macrófagos, y la RDMex03 contiene genes involucrados en rutas regulatorias.

ABSTRACT

Studies of *Mycobacterium bovis* BCG strains used in different countries and vaccination programs show clear variations in the genomes and immunoprotective properties of BCG strains. The aim of this study was to characterise the genomic profile of the BCG 1931 strain used in Mexico for BCG production.

BCG Mexico 1931 has a circular chromosome of 4,350,386 bp with a G+C content and numbers of genes and pseudogenes similar to those of BCG Tokyo and BCG Pasteur. BCG Mexico 1931 lacks Region of Difference 1 (RD1), RD2 and N-RD18 and one copy of IS6110, indicating that BCG Mexico 1931 belongs to DU2 group IV within the BCG vaccine genealogy. In addition, this strain contains three new RDs, which are 53 (RDMex01), 655 (RDMex02) and 2,847 bp (REDMex03) long, and 37 single-nucleotide polymorphisms representing non-synonymous mutations compared to BCG Pasteur and BCG Tokyo.

BCG Mexico 1931 has a smaller genome than the BCG Pasteur and BCG Tokyo strains. Two specific deletions in BCG Mexico 1931 are described (RDMex02 and RDMex03). The loss of RDMex02 (*fadD23*) is associated with enhanced macrophage binding and RDMex03 contains genes that may be involved in regulatory pathways.

ABREVIATURAS

DNA	Ácido desoxirribonucleico
BAC	Cromosoma artificial bacteriano
BCG	Bacilo de Calmette y Guèrin
BLAST	Programa de búsqueda y alineamiento de tipo local
CDS	Secuencias codificantes
DMSO	Dimetil-sulfóxido
DO	Densidad óptica
DU	Duplicación en tándem
EDTA	Ácido etilen diamino tetra acético
MIRUs	Unidades repetidas intercaladas en micobacterias
OADC	Ovoalbúmina-dextrosa-catalasa
OMS	Organización Mundial de la Salud
pb	pares de bases
PCR	Reacción en cadena de la polimerasa
PFGE	Electroforesis en gel por campos pulsados
PIMM	Programa de Inmunología Molecular Microbiana
RD	Región de diferencia o región ausente
SNP	Polimorfismo de nucleótido simple
TAE	Solución amortiguadora de acetatos
TAES	Tratamiento acortado estrictamente supervisado
TBE	Solución amortiguadora de boratos
TE	Solución de Tris-EDTA
T _m	Temperatura de alineación
UFC	Unidades formadoras de colonia
VIH	Virus de la inmunodeficiencia humana
VNRT	Repeticiones en tándem de número variable

I. ANTECEDENTES

Hoy en día la tuberculosis es una enfermedad con alta incidencia a nivel mundial, siendo uno de los más grandes problemas de salud. En 1993 fue declarada por la Organización Mundial de la Salud (OMS) como una enfermedad reemergente debido a un incremento en el número de casos, asociados principalmente a la coinfección con el virus de inmunodeficiencia humana (VIH) y al número de cepas de *M. tuberculosis* resistentes a los fármacos de primera elección (isoniazida, rifampicina, etambutol y estreptomycin) para el tratamiento de la tuberculosis [1].

La tuberculosis es ocasionada por un grupo de bacterias estrechamente relacionadas denominadas complejo *Mycobacterium tuberculosis*, siendo *M. tuberculosis* el principal agente causal de la enfermedad en humanos [1].

I.1 El complejo *Mycobacterium tuberculosis*

La tuberculosis es ocasionada por un grupo de bacterias estrechamente relacionadas denominadas complejo *Mycobacterium tuberculosis*. A este grupo pertenecen *M. tuberculosis*, el agente causal de la tuberculosis en humanos; *M. bovis*, responsable de la enfermedad en el ganado bovino; *M. africanum*, un grupo de cepas que producen la enfermedad en humanos; *M. microti*, que produce la enfermedad en roedores y; *M. canetti* que se considera la bacteria ancestral del complejo [2].

M. tuberculosis fue aislada por primera vez por Robert Koch en 1882 a partir de biopsias y comúnmente se le denomina bacilo de Koch o bacilo tuberculoso. Este bacilo se considera de crecimiento lento, posee un tiempo de generación de 15 a 20 horas con un crecimiento colonial visible (medio estándar) en un tiempo mínimo de tres semanas y usualmente entre cuatro y seis semanas. Es aerobio, no forma esporas y no presenta movilidad [2].

Una característica importante de las micobacterias en general, es el alto contenido de lípidos de alta masa molecular que presentan en su pared celular, los cuales constituyen alrededor del 60% del peso seco. Esta característica le confiere resistencia a la desecación, la protege contra un gran número de antibióticos y agentes antimicrobianos,

además de ser la limitante para el desarrollo de técnicas de biología molecular que ayuden a su identificación. Además, el alto contenido de lípidos en su pared celular permite identificarlos como bacilos ácido alcohol resistentes debido a que se tiñen con la técnica de Ziehl-Neelsen [2, 3].

La pared celular de las micobacterias se encuentra formada por tres capas: una interna, una intermedia y una externa. La primera capa contiene peptidoglucano y micopéptido, la segunda arabinogalactano y ácidos micólicos y la tercera ceras D y micósidos C [3].

I.II Epidemiología de la tuberculosis

La OMS estima que un tercio de la población (aproximadamente 2 330 millones de personas) está infectada con *M. tuberculosis*, lo que constituye un número importante de personas si se toma en cuenta que una persona con tuberculosis activa infecta en promedio de 10 a 15 personas por año y que de éstas de un 5 al 10% desarrolla la enfermedad en alguna etapa de su vida [1, 2].

En el 2010, la OMS estimó 8.8 millones de nuevos casos de tuberculosis a nivel mundial (equivalente a 128 casos por 100 000 habitantes) y 1.1 millones de muertes. De éstos, el mayor número de casos ocurrió en Asia (59%) y en África (26%). Los países con el mayor número de casos fueron la India (2.0–2.5 millones), China (0.9–1.2 millones), Sudáfrica (0.40–0.59 millones), Indonesia (0.37–0.54 millones) y Pakistán (0.33-0.48 millones), figura 1 [4].

Con relación a la coinfección con el VIH, del 11 al 13% de los casos estimados corresponden a sujetos VIH positivos, con 0.4 millones de muertes en este grupo, que se suman a las 1.1 millones mencionadas previamente [4].

Estimated TB incidence rates, 2010

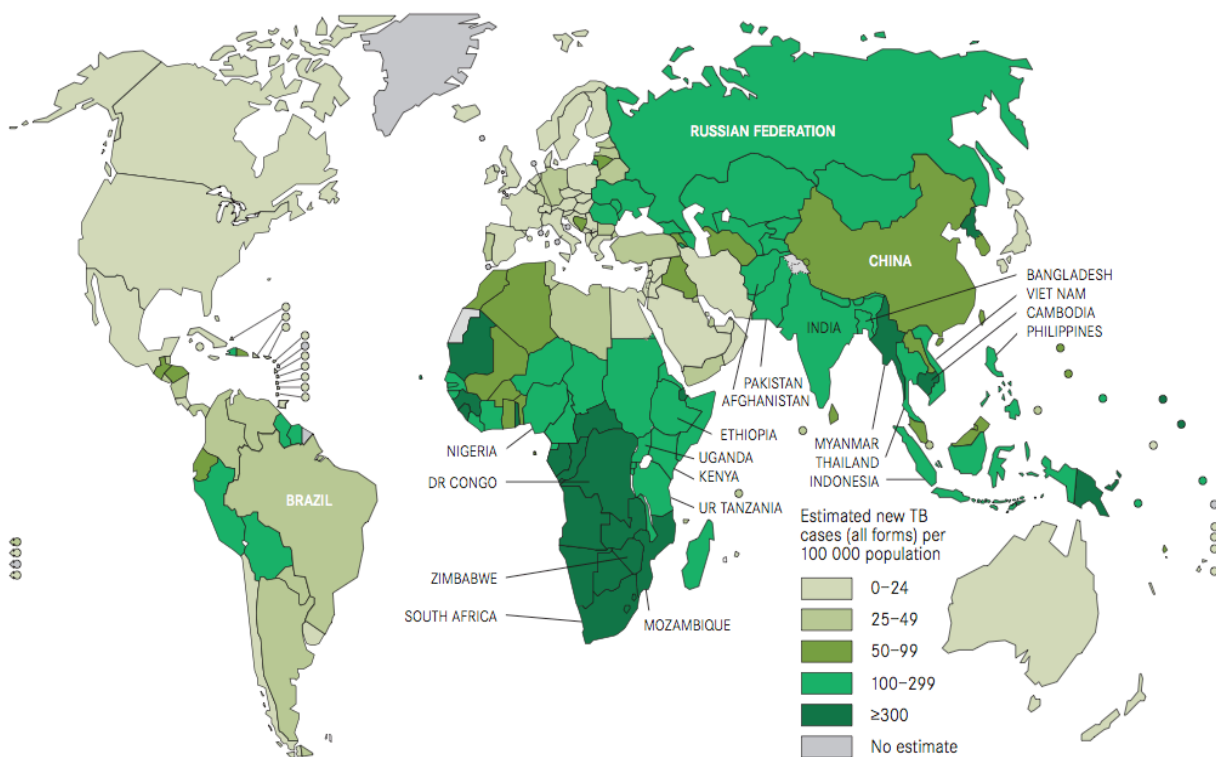


Figura 1. Incidencia de tuberculosis reportada por la OMS en el 2010 [4].

I.II.I Epidemiología de la tuberculosis en México

En México, la tuberculosis pulmonar es un problema de salud pública que afecta a todos los grupos de edad, particularmente a adultos jóvenes y personas de mediana edad. La coinfección con el VIH es aproximadamente del 3% y se estima que afecta al 25% de las personas infectadas con el virus [5].

En el Boletín semanal que emite la Dirección General de Epidemiología de la Secretaría de Salud, entre el año 2000 y el 2009 no se presentaron cambios importantes en el número de casos de tuberculosis reportados en el país, aunque se presentó una tendencia discreta de disminución de casos por año entre el 2004 y 2006, mientras que para el 2008 y 2009 se observa un discreto incremento en el número de casos (figura 2) [6].

Para el 2009 la OMS reportó una incidencia menor de 25 casos por 100 000 habitantes en México [7]. Mientras que para ese mismo año la Dirección General de Epidemiología reportó 14 856 nuevos casos de tuberculosis pulmonar, con la menor incidencia de casos de tuberculosis en el estado de México (2.89) y la mayor incidencia en Baja California (37.72). De esta manera los estados con mayor incidencia fueron Baja California, Guerrero, Tamaulipas, Sonora y Nayarit; mientras que los estados con menor morbilidad fueron el Estado de México, Tlaxcala, Aguascalientes, Zacatecas y Guanajuato [6]. Estas diferencias podrían explicarse por el porcentaje en la cobertura del tratamiento acortado directamente supervisado (TAES) y la virulencia de las cepas prevalentes en cada región. En este año, Nava y cols reportaron que en población adulta en Acapulco, Guerrero el 26% de los pacientes están infectados con *M. tuberculosis* familia Manila [8].

De los casos reportados en el 2009, el grupo etario de 25 a 44 años fue el que presentó un mayor número de casos de tuberculosis con 5 110. Sin embargo, el grupo de personas mayores a los 60 años fue el que se presentó la mayor incidencia (>30) [6].

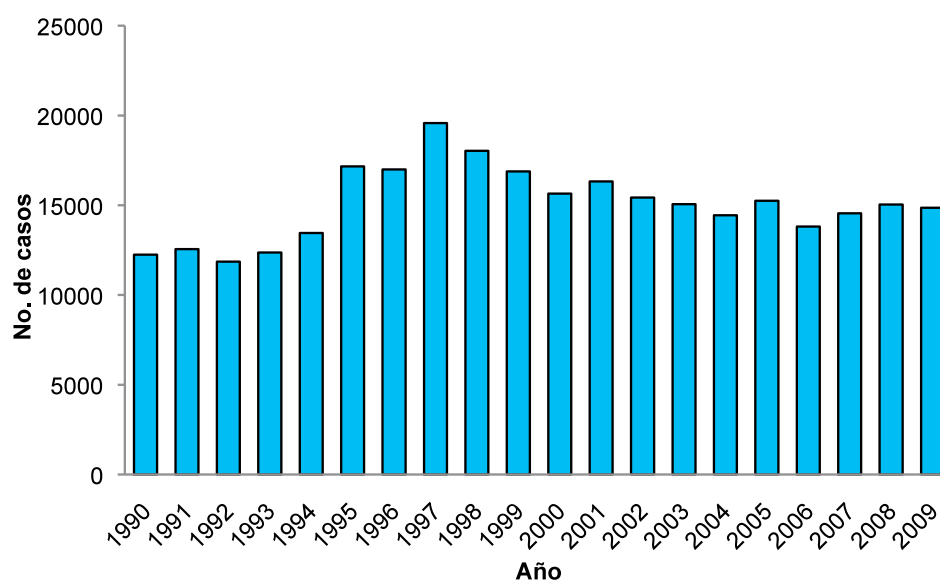


Figura 2. Índice de morbilidad de la tuberculosis en México reportada por la Secretaría de Salud. Datos obtenidos de la Dirección General de Epidemiología y la Organización Panamericana de la Salud [6].

Recientemente, se reportó que la manifestación clínica más frecuente en población pediátrica en el país es la tuberculosis pulmonar, además se encontró que el 27% de los aislamientos presentan resistencia a algún fármaco empleado en el tratamiento de la tuberculosis y 11% corresponden a aislamientos multi-fármaco-resistentes [9].

I.III Prevención y control de la tuberculosis

A partir del incremento en el número de casos de tuberculosis a principios de los noventa, se comenzaron a aplicar nuevas estrategias con la finalidad de controlar la enfermedad.

Las tres medidas utilizadas actualmente a nivel mundial para prevenir y controlar la tuberculosis son: a) la estrategia del Tratamiento Acortado Estrictamente Supervisado (TAES, Alto a la tuberculosis), b) la aplicación de la vacuna BCG, considerada como la única medida preventiva eficaz contra las formas graves de la tuberculosis y c) el desarrollo de nuevas vacunas [1].

El TAES consiste en la detección temprana de la enfermedad mediante baciloscopía y el tratamiento supervisado con fármacos anti-tuberculosos de primera línea (rifampicina, isoniazida, etambutol) el cual debe de ser corto (3-6 meses) y supervisado diariamente para evitar que el paciente lo abandone [1].

La prevención de la tuberculosis se logra con la inmunización de la vacuna Bacilo de Calmette y Guérin (BCG). Esta vacuna es eficaz para proteger contra la tuberculosis miliar y meníngea, sin embargo, presenta una alta variabilidad en la protección que confiere contra la tuberculosis pulmonar en adultos con una eficacia entre el 0 y el 80%; dada esta gran controversia se desarrollan y evalúan nuevas vacunas o refuerzos contra la tuberculosis [10].

I.IV La vacuna BCG

La vacuna BCG se utilizó por primera vez en 1921 para prevenir la tuberculosis y se estima que ha sido administrada a más de tres mil millones de personas alrededor del

mundo [1, 10]. Inicialmente la vacunación con BCG fue por vía oral a recién nacidos y posteriormente por vía intradérmica lo que llevó a una reducción de efectos adversos y a una mayor protección, así como a la administración de una menor dosis de BCG y un mejor control de la dosis efectiva. El último país que utilizó la vía oral fue Brasil en 1973 [10-12].

Actualmente la vacuna se produce a partir de diferentes BCGs, entre las que se encuentran la BCG Connaught, Danesa, Moreau, Pasteur, Rusa y Tokio. A partir de estas BCGs se produce el 90% de las vacunas aplicadas a nivel mundial. La dosis recomendada de la vacuna en niños es de 0.05 mL, aunque generalmente se administra 0.1 mL, volumen que contiene de 200 000 a 3 000 000 UFC dependiendo de la BCG administrada [10, 12]. En la figura 3 se muestra la cobertura de vacunación a nivel mundial reportada por la OMS en el 2009 [1].

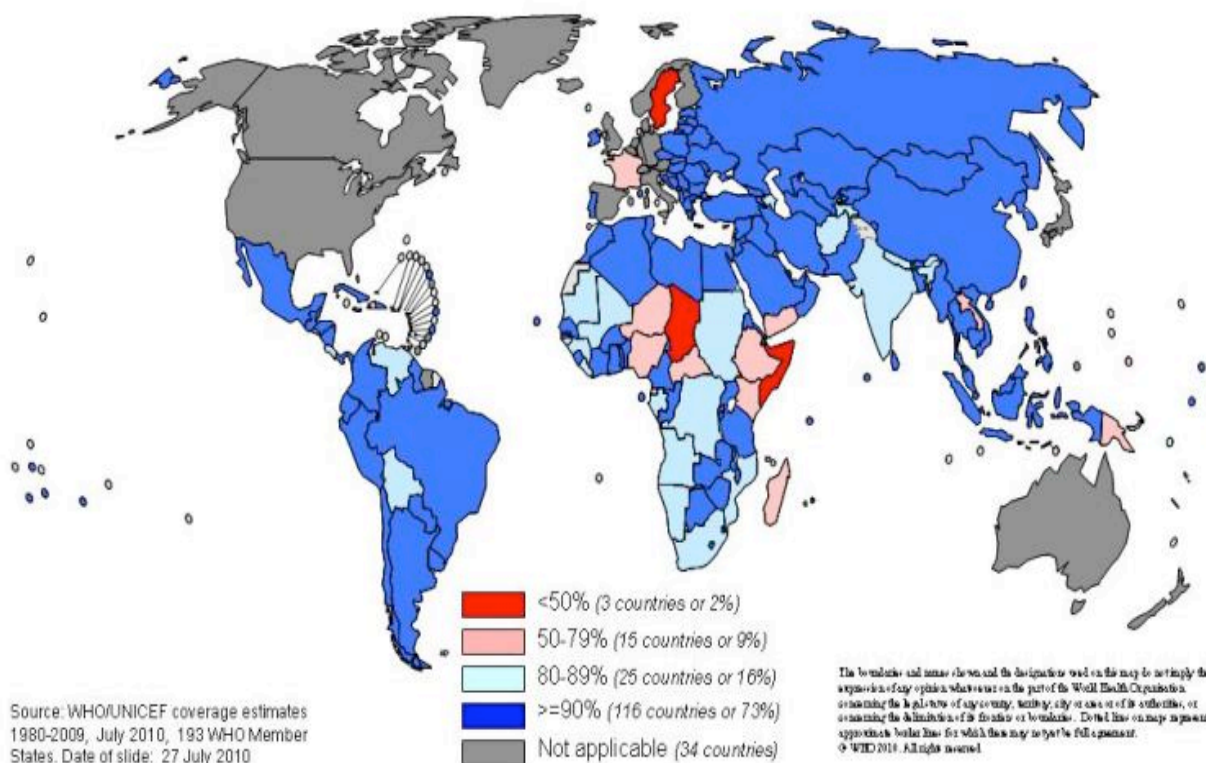


Figura 3. Cobertura de vacunación con BCG reportada por la Organización Mundial de la Salud en el 2009 [1].

En México, la NOM-0.36-SSA2-2002 establece la vacunación con BCG al año de edad, de 0.1mL por vía intradérmica de alguna de las BCGs autorizadas que son la siguientes: BCG Pasteur 1173P2, BCG Montreal, BCG Danesa 1331, BCG Tokio 172 y BCG Glaxo 1077 [13].

I.IV.I Desarrollo de la vacuna BCG

La vacuna BCG es una cepa viva atenuada, obtenida por Albert Calmette y Camille Guérin (Instituto Pasteur, Lille, Francia) a partir de un aislamiento de *M. bovis* proveniente de un caso de mastitis bovina [14, 15].

La BCG fue atenuada después de realizar 230 cultivos, cada dos semanas, en medio de papa glicerizada con bilis de buey (1908-1921). A lo largo de este tiempo se documentó la pérdida de la virulencia del bacilo en diferentes modelos animales como vacas, cobayos, monos, etc., así como un cambio en la morfología de las colonias obtenidas en diferentes cultivos. Calmette realizó el primer reporte de atenuación de la cepa después del decimoquinto cultivo [14-16]. A partir de la obtención de la cepa atenuada, en 1921, la cepa BCG se mantuvo mediante cultivos continuos hasta 1960, año en el cual se implementó la liofilización como medio de preservación de la BCG.

Para 1921, la vacuna BCG líquida comenzó a aplicarse con la finalidad de prevenir la tuberculosis en humanos, la primera administración exitosa se realizó a un niño cuya madre padecía la enfermedad y el cual no contrajo la tuberculosis. Para 1926, se vacunaron 5 183 bebés por vía oral con tres dosis de la vacuna BCG. En los reportes realizados por Calmette se observó una disminución de un 25 a un 2 % en la mortalidad de niños de un año que convivían con madres tuberculosas [14]. Así, entre 1924 y 1929 había en Francia más de 152 000 niños vacunados con BCG, sin reportes de accidentes imputados a la vacuna [11].

Debido a los resultados exitosos de la vacunación con BCG, ésta se distribuyó a diversos países del mundo mediante el envío de una copia del cultivo de BCG que Calmette conservaba en el Instituto Pasteur (Lille, Francia) [15, 17].

Los registros históricos muestran que la primera cepa se envió a Rusia en 1924 y se denominó BCG Moscú. De esta manera, cada cultivo de BCG derivado del Instituto Pasteur se identificó como BCG seguida del nombre del país que la recibió, el laboratorio que la producía o el apellido del investigador que la recibió (por ejemplo: BCG Danesa, BCG Pasteur y BCG Moreau, respectivamente).

Entre 1924-1926 se enviaron treinta y cuatro cultivos de BCG del Instituto Pasteur a diversos países y en 1927 otros 26 recibieron una copia de la cepa BCG original de Francia. En los años siguientes se continuó con el envío a los países que lo solicitaron, donde la cepa se preservó en diferentes medios y bajo distintas condiciones. Esta actividad se conserva hasta el día de hoy. Además, algunas otras BCGs se obtuvieron de las derivadas directamente del Instituto Pasteur; así por ejemplo la BCG Connaught derivó de la BCG Frappier, mientras la BCG Glaxo derivó de la BCG Danesa [14, 15, 17].

Con el paso del tiempo se documentaron diferencias en las características morfológicas (macroscópicas y microscópicas) y en la eficacia protectora que presentan las diferentes BCGs, lo cual se atribuye a los diferentes métodos de conservación utilizados para conservar la cepa [18, 19]. Esto último debido a que estas cepas se conservaron por subcultivos continuos hasta 1960, año a partir del cual comenzó a utilizarse la liofilización como método de conservación de la BCG. De hecho no solo cambia el método de preservación, también cambia el método de producción al método de lote semilla, lo cual lleva a la producción de la vacuna BCG liofilizada.

A la fecha, es clara la existencia de diferencias entre las distintas BCGs, lo cual generó la necesidad de caracterizarlas a diferentes niveles. En las siguientes páginas se describen algunas de las diferencias documentadas entre estas cepas.

I.IV.II Diferencias en eficacia protectora

El estudio de la eficacia de la vacuna BCG se divide en dos grupos: a) la protección que confiere contra la tuberculosis miliar y meníngea en niños y b) la protección que confiere contra la tuberculosis pulmonar en adultos.

Por lo que respecta al primer grupo, existen resultados consistentes de que la BCG protege contra la tuberculosis miliar y meníngea en la población infantil, con una eficacia protectora de alrededor del 80% [20, 21].

Sin embargo, la protección que confiere contra la tuberculosis pulmonar en adultos es muy variable con una eficacia entre el 0 y el 80%. En general, como se aprecia en la figura 4, la eficacia protectora de la vacuna BCG varía de un 0 a un 100% que depende de la BCG utilizada, del año de aplicación y de la región geográfica (basada en su población genética) en la cual se realiza el estudio; por ejemplo la BCG Copenhague mostró un 77% de protección en niños de Inglaterra mientras que, al aplicarse la misma vacuna BCG a población general en Colombia la protección conferida fue del 0% [12, 22]. En un metanálisis realizado en 1994 con 16 diferentes estudios se encontró que la vacuna BCG reduce el riesgo de desarrollar tuberculosis pulmonar en un 50% de los casos [23]. Esta variabilidad derivó en el desarrollo de nuevas vacunas para prevenir la enfermedad [10, 21].

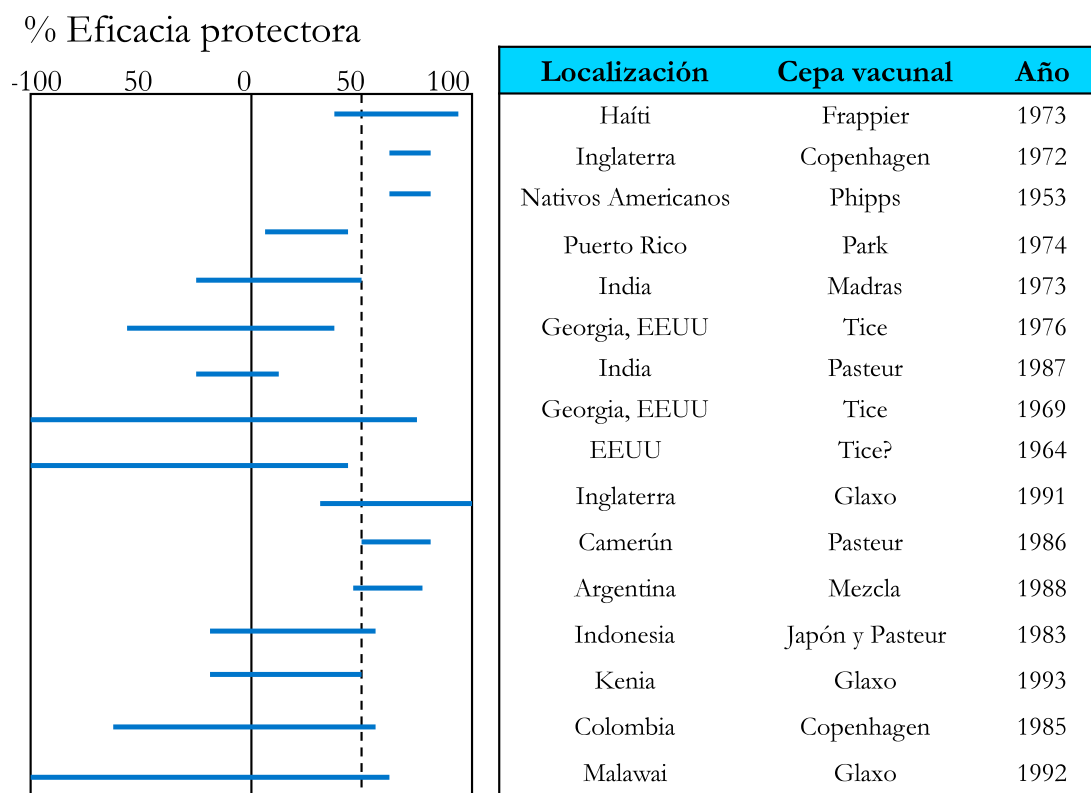


Figura 4. Meta-análisis de estudios sobre la eficacia de la vacuna BCG contra la tuberculosis pulmonar. En el cuadro se muestran el lugar y el año en el que se llevó a cabo el estudio, así como la BCG utilizada. (10)

Las razones por las cuales se da esta variación no son claras, sin embargo, existen diversos factores que afectan la eficacia de la vacuna BCG. Entre éstos se encuentran las diferencias entre las vacunas BCG, el contacto con micobacterias ambientales, factores genéticos propios del humano y la identificación de diferencias en la virulencia de las cepas de *M. tuberculosis* que causan la enfermedad [10, 12].

I.IV.III Diferencias morfológicas de la BCGs

Diferencias morfológicas atribuidas a los diferentes métodos de preservación, como el tamaño, la textura y la forma de las colonias entre las BCGs comenzaron a observarse poco tiempo después de que inició su distribución alrededor del mundo [19]. En 1950 Van Deinse y cols describieron que el número de subcultivos en medio Sauton, así como el tiempo entre éstos provocan cambios en la morfología (diferencias en tamaño y textura de la colonia) y en el tiempo de crecimiento de las BCGs [24]. El reporte de estas diferencias llevó a un estudio más extenso de las características fenotípicas de las BCGs y por consiguiente a la descripción de un mayor número de diferencias [15].

I.IV.IV Diferencias bioquímicas de las BCGs

De igual forma se identificaron diferencias en la producción de antígenos entre las BCGs, como el MPB64 y el MPB70. Los antígenos MPB64 y MPB70 se han descrito como proteínas inmunogénicas en diferentes micobacterias. En las BCGs se describió que las derivadas antes de 1927 del Instituto Pasteur secretan estos antígenos mientras que las BCGs que se distribuyeron posteriormente son incapaces de sintetizarlos [15, 25]. Esta característica permitió clasificarlas en BCGs “tempranas” y “tardías”, considerándose como “tempranas” a las BCGs capaces de secretar estos antígenos.

Otra de las diferencias documentadas a nivel bioquímico entre las BCGs es su capacidad de producción de ácidos micólicos. Estos compuestos se encuentran en la parte intermedia de la pared celular de las micobacterias y son de tres tipos: alfa-micólicos, metoxi-micólicos y ceto-micólicos. Las BCGs derivadas después de 1931 del Instituto

Pasteur son incapaces de producir los ácidos metoximicólicos debido a una mutación puntual en el gen *mma3*. Esta característica también ayuda a clasificar a las BCGs en “tempranas” y “tardías” [26, 27].

Además de los antígenos ya mencionados, entre los que más se han estudiado están ESAT-6 y CFP-10, los cuales se encuentran dentro de la región denominada RD1, que se perdió en todas las BCGs [28]. Estos antígenos no permiten diferenciar a las BCGs, sin embargo, son importantes ya que las diferencian de *M. tuberculosis* y *M. bovis*.

I.IV.V Diferencias inmunológicas de las BCGs

A través del tiempo diversos estudios describieron la respuesta inmunológica inducida por diferentes BCGs tanto *in vivo* como *in vitro*. Los primeros estudios que reportaron diferencias entre las BCGs se basaron en la determinación de la virulencia y persistencia de las BCGs en diferentes modelos animales, así como en la protección conferida con retos de micobacterias virulentas después de la vacunación. Encontrándose diferencias en estos parámetros dependientes de la BCG estudiada, la dosis empleada y el tiempo transcurrido después de la vacunación [18, 29]

Posteriormente, se realizaron estudios en los que se determinó la producción de diferentes citocinas como IL10, IL12, IL4, IL2 e $INF\gamma$ después de la vacunación con o sin reto de micobacterias virulentas [30]; lo cual puso de manifiesto diferencias entre las BCGs. Hoy en día no se cuenta con una correlación entre los resultados obtenidos y éstos varían dependiendo de las BCGs utilizadas, la dosis y el modelo animal, así como del tiempo de determinación que transcurre después que es aplicada la vacuna [29].

Bajo este contexto, en el estudio realizado por Castillo- Rodal se evaluó la protección que confieren 10 BCGs (Pasteur, Danesa, México, Sueca, Tice, Phipps, Frappier, Connaught, Birkhaug y Moreau) contra la tuberculosis en un modelo murino de tuberculosis pulmonar progresiva. Los resultados mostraron que la BCG que confirió la mayor protección fue la BCG Phipps, mientras que la que presentó la menor protección fue la BCG Tice; sin embargo no se encontró una relación entre protección, tamaño del genoma y período de obtención del Instituto Pasteur [31].

I.IV.VI Diferencias genómicas en las BCGs

El campo de la genómica permitió estudiar de manera más exhaustiva las diferencias entre las BCGs. Una de los primeros ensayos aplicados fue la electroforesis en gel por campos pulsados (PFGE), el cual mostró diferentes perfiles entre las BCGs; sin embargo, no fue posible establecer diferencias claras entre ellas [32]. Bajo esta misma estrategia se utilizaron ensayos que identifican polimorfismos en el genoma de las BCGs como la determinación de unidades repetitivas de las micobacterias (MIRUs) y secuencias repetidas de número variable (VNRTs), entre otras [2, 33].

Adicionalmente, un marcador genético ampliamente utilizado es la presencia de la secuencia de inserción IS6110, de la cual presentan dos copias las BCGs derivadas del Instituto Pasteur antes de 1927 y una copia las derivadas después de esta fecha. De esta manera, la presencia de una o dos copias de la IS6110 permite clasificar a las BCGs en dos grupos, tempranas y tardías [2].

Los trabajos con mayor significancia por su contribución en este campo son los realizados por los grupos de Mahairas G, Behr M y Brosch R [17, 34-36]. En éstos se describió la presencia de deleciones genéticas en las BCGs en comparación con *M. tuberculosis* y *M. bovis*, las cuales se denominaron RDs (regiones deletadas o de diferencia).

Mahairas determinó mediante técnicas de hibridación substractiva la existencia de regiones que se encuentran ausentes en las BCGs pero presentes en *M. tuberculosis* H37Rv, como la RD1 [35]. Posteriormente, Brosch y cols identificaron nuevas regiones RD utilizando cromosomas artificiales bacterianos (BACs), las cuales están presentes o ausentes dependiendo de la BCG analizada [36]. Esta última característica permitió diferenciarlas del complejo *M. tuberculosis* e identificarlas mediante la reacción en cadena de la polimerasa (PCR) [37]. Para fines de los años 90's, Behr realizó un estudio de microarreglos con 13 BCGs, en el cual se describieron nuevas regiones RD y además se determinó que estas regiones RD y las descritas anteriormente correlacionan con el año en el cual las BCGs derivaron del Instituto Pasteur. Este hallazgo llevó a la construcción de la primera genealogía de BCGs [17].

Actualmente, se han descrito alrededor de 15 regiones RD entre las BCGs y *M. bovis*, algunas de las cuales son responsables de las diferencias fenotípicas observadas entre ellas [28]. Así por ejemplo, la RD1 es una deleción que remueve 9 genes (*PE35*, *PPE68*, *esat-6*, *cfp10*) y que separa a las BCGs de *M. bovis* [28]. La deleción de esta región en *M. tuberculosis* induce su atenuación; sin embargo, la complementación con los genes codificados en la región RD1 no restaura completamente la virulencia de las BCGs [38, 39]. Mientras que, la RD2 separa a las BCGs tempranas, obtenidas antes de 1927 del Instituto Pasteur, de las BCGs tardías; Kozak y cols describieron recientemente que la presencia de esta región está relacionada con la inmunogenicidad de la BCG [40]. Sin embargo, estas deleciones no explican completamente las diferencias observadas entre las BCGs, lo cual sugiere que existen otras regiones involucradas en la atenuación de la cepa.

Además de las regiones RD, también existen descritas dos duplicaciones (DU) en las BCGs, denominadas DU1 y DU2, siendo la DU1 única de BCG Pasteur. La presencia de estas regiones lleva a la duplicación de genes en las BCGs, sin embargo, la implicación funcional de estas duplicaciones aún no está descrita [41].

Para el 2007 se liberó el genoma de la BCG Pasteur 1173P2, constituido de 4 374 522pb, aproximadamente 30Kb más grande que el de *M. bovis*. La secuenciación de este genoma y su posterior análisis permitió la construcción de una genealogía refinada de BCGs (figura 4), la cual tiene como marcadores las RD mencionadas anteriormente y la DU2. Esta última presenta diferentes tamaños entre las BCGs, lo que permitió agruparlas en cuatro grupos, DU2 tipo I-IV [34].

Recientemente, también se publicó la secuencia del genoma completo de la BCG Tokio 172, la cual tiene un tamaño de 4 371 711pb. La secuenciación y comparación con BCG Pasteur permitió identificar mutaciones en BCG Tokio, específicamente inserciones y deleciones [42].

Adicionalmente, el análisis de las secuencias genómicas de BCG Pasteur y Tokio, así como comparaciones de regiones específicas de las BCGs con las secuencias de *M. tuberculosis* y *M. bovis* llevó a la identificación de polimorfismos de nucleótido simple

(SNPs), tanto únicos como compartidos entre las BCGs [34, 42, 43]. Algunos de los SNPs descritos tienen inferencias funcionales en los genes que afectan, un ejemplo lo constituye el SNP en el gen *mma3*, responsable de la pérdida de la producción de los ácidos metoximicólicos en las BCGs tardías [27, 28]. De hecho, en los últimos años ha tomado fuerza el estudio de SNPs debido a las implicaciones funcionales que pueden tener, sobre todo si se considera que los polimorfismos descritos previamente (deleciones y duplicaciones) continúan sin explicar completamente la atenuación y sobre-atenuación observada en las BCGs.

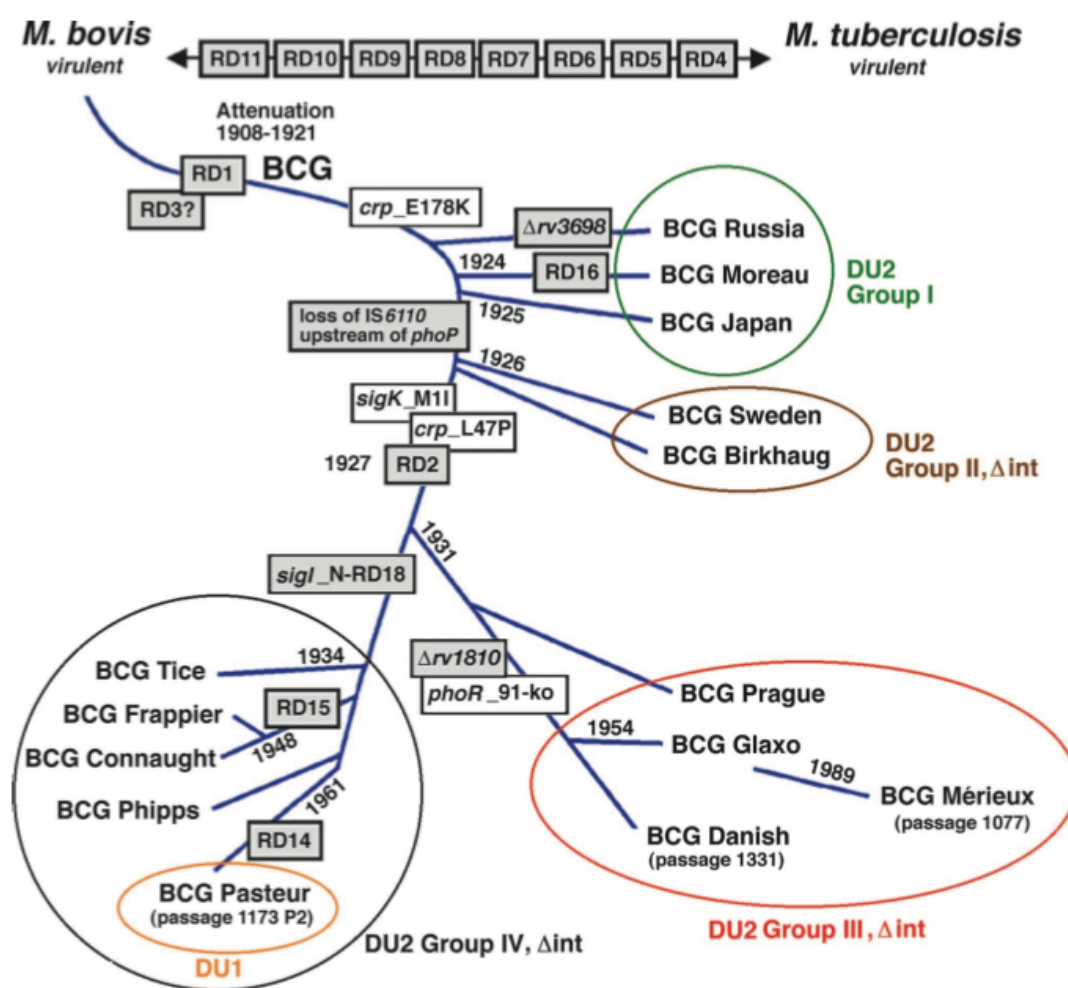


Figura 5. Genealogía de las BCGs, construida con base en regiones RD y DU. La figura representa el orden cronológico en el que las BCGs derivaron del Instituto Pasteur y sus mutaciones [34].

En resumen, el estudio de las BCGs a nivel genómico permitió establecer diferencias específicas entre ellas y, en algunos casos, identificar perfiles genómicos únicos que establece métodos de identificación. Además, de relaciones específicas entre las mutaciones presentes en las BCGs y su atenuación. El conocimiento generado en este campo también contribuye al diseño de vacunas de segunda generación contra la tuberculosis.

I.IV.VII Diferencias en la expresión proteica de las BCGs

Los primeros perfiles proteómicos descritos para las BCGs incluyeron las BCG Chicago, Copenhague y Pasteur. El análisis mostró diferencias en la expresión de proteínas (presencia/ausencia) de la fracción celular, del filtrado de cultivo y en las proteínas de membrana obtenidas bajo condiciones de cultivo estándar [44-46].

Recientemente, Rodríguez y *co/s* realizaron la comparación entre los proteomas de BCG Phipps y BCG Danesa [46]. El análisis mostró que estas BCGs presentan un diferente número de proteínas en la fracción celular bajo condiciones estándar de cultivo. Así, en BCG Phipps se encontró la sobre-expresión de las proteínas alquil hidropéroxido reductasa C (AhpC), de filtrado de cultivo 17 (CFP17) y las chaperonas GroEL1 y GroEL2; mientras que en la BCG Danesa se detectó la sobre expresión de las proteínas GroEL2 y TB39.8, la proteína de 35KDa y una probable alcohol deshidrogenasa. Todas estas proteínas están involucradas en la virulencia, detoxificación y adaptación de diferentes micobacterias.

Las evidencias mencionadas anteriormente llevan a la necesidad de continuar la caracterización minuciosa de las BCGs con la finalidad de comprender sus mecanismos de atenuación y a usar la información generada para un diseño racional de vacunas. Actualmente la OMS establece grupos de trabajo que realizan la caracterización de diferentes BCGs. Así como también la recomendación del uso de métodos estándar para ejecutarlo, entre los que se encuentran la determinación de repeticiones en tándem de número variable (VNRT), el estudio de las regiones RD y DU, pruebas de potencia y de susceptibilidad de la vacuna BCG, entre las más importantes [47].

I.V Uso y producción de la vacuna BCG en México

La vacuna BCG se introdujo en México en 1925 por Fernando Ocaranza, director del Instituto de Higiene, y a partir de 1931 inició la producción de la vacuna BCG líquida en el país a partir de cepas remitidas por Calmette [48-50]. Actualmente se sabe que se realizaron diferentes envíos de la BCG a México, este hecho se respalda con cartas de solicitud dirigidas al Instituto Pasteur por el Dr. Alfonso Reyes, fechadas el 1 de julio de 1926, el 12 de febrero y el 7 de marzo de 1927.

Durante este período la producción de la vacuna BCG se realizó a una escala muy pequeña y bajo solicitud del médico. Así cuando el médico solicitaba la vacuna ésta se producía y entregaba para su aplicación inmediata, el esquema de inmunización consistía en administrar tres dosis de BCG al recién nacido en un período no mayor a 10 días [51].

La aplicación de la vacuna BCG continuó bajo este esquema durante los siguientes 15 años, hasta 1949, año en que se inaugura el Instituto Nacional de Higiene y con él el Laboratorio de Producción de la vacuna BCG cuyas funciones fueron elaborar la vacuna BCG para abastecer todas las necesidades del país. En este período la vacuna BCG se produjo a partir de la BCG enviada directamente del Instituto Pasteur de París con la designación No.870 Serie I (denominada BCG México), la cual se conservó por duplicado mediante cultivos continuos de periodicidad bisemanal en papa glicerizada y biliar [50, 52, 53].

La vacuna BCG México se aplicaba a los niños durante los primeros 10 días de vida y a adultos PPD negativos. La vacunación con BCG inició por vía oral y posteriormente se cambió a la vía intradérmica debido a los efectos de BCGitis causados en niños. Iniciando así en la década de los cincuenta las campañas masivas de vacunación con BCG México en el país [50, 53]. En 1963 una comisión especial de la División de Control de Productos Biológicos de la OMS aprobó el uso de esta vacuna, al demostrar en estudios comparativos con la BCG Danesa (cepa patrón) que la vacuna BCG México producida en el país tenía una potencia y calidad comparables a ésta. Este gran logro propició su distribución a nivel nacional, así como a diversos países de Centro América y América del Sur, vacunando a más de 15 millones de niños [52].

La producción de la vacuna BCG México continuó hasta 1970, año en que se cambia el método de producción al de “lote semilla” (vacuna BCG liofilizada), empleándose la BCG Danesa 1331 para la producción, por medio del cual el liofilizado original no fue objeto de más de 12 cultivos continuos (en realidad nunca más de 8 cultivos); a fin de evitar que la BCG sufriera una posible variación genómica, y en caso de que este fenómeno ocurriera fuera sustituida por otra BCG con diferentes propiedades, además de conservar una calidad homogénea de la vacuna BCG. Con esto comenzó a producirse la vacuna BCG Danesa liofilizada de calidad internacional, la cual poseía ventajas con respecto a la líquida como la de una mejor y mayor conservación en condiciones ambientales adversas [50, 54].

En 1998 la vacuna BCG Danesa dejó de producirse definitivamente en México y a partir de ese año se importa la vacuna BCG disponible comercialmente; por ejemplo BCG Danesa, BCG Tokio y una BCG producida en la India, sin importar su calidad, potencia y eficacia. Lo anterior originó que en México se use una gran diversidad de vacunas BCG y con ello se tenga una eficacia protectora incierta. Existen también reportes, poco frecuentes, de una falta de abastecimiento de la vacuna BCG en los sistemas de salud, lo que origina la ausencia de vacunación por periodos de hasta 6 meses. Además, la importación de la vacuna BCG también lleva a un incremento en el costo de la vacunación, aún cuando el personal entrenado en la preparación de esta vacuna se encuentra activo laboralmente. Esto último nos hace reflexionar en la conveniencia de producir nuevamente la vacuna BCG México en el país.

Tomando como base la historia de estudios y vacunación con la BCG México durante 20 años en nuestro país, decidimos realizar la caracterización genómica de la BCG México, ya que ésta es una de las áreas de investigación mejor estudiada para caracterizar a las BCGs. Además de que nos permitirá determinar con precisión las similitudes y diferencias con otras BCGs y que en el futuro sea considerada para la construcción de vacunas de segunda generación [55]. Para realizar esta caracterización empleamos tres BCGs que representan los diferentes períodos de producción de la vacuna BCG en el país, con la finalidad de realizar la identificación de la BCG México.

II. PLANTEAMIENTO DEL PROBLEMA

Un gran número de estudios demostraron la eficacia protectora de la vacuna BCG contra las formas graves de la tuberculosis. En el país, la BCG México se aprobó por la OMS para su producción y uso, la cual se realizó durante los años de 1949 a 1970 con una cobertura de vacunación de más de 15 millones de niños. La vacunación con la BCG Danesa producida en México se realizó de 1970 a 1998. Actualmente, México compra la vacuna BCG que se encuentre disponible comercialmente.

La caracterización genómica de la BCG México permitirá situarla en la genealogía existente de BCGs y considerar su uso para la producción de la vacuna en México y para el diseño de nuevas vacunas.

III. HIPÓTESIS

La BCG México tendrá un perfil genómico diferente a los perfiles reportados previamente para otras BCGs, en el cual se identificarán regiones genéticas específicas de la cepa que ayuden a explicar sus características fenotípicas.

IV. OBJETIVOS

Objetivo General

Realizar la caracterización genómica de la BCG México.

Objetivos Particulares

- 1) Determinar la presencia de regiones duplicadas (DUs) y de diferencia (RD) en tres BCGs (México 1931, 1988 y 1997) utilizadas para la producción de la vacuna BCG en México.
- 2) Determinar el perfil genómico de la BCG México 1931 por electroforesis en gel por campos pulsados.
- 3) Realizar la secuenciación del genoma completo de la BCG México 1931.

V. MATERIALES Y MÉTODOS

V.I Cepas empleadas en el estudio

Las BCGs utilizadas en el estudio fueron Birkhaug, Connaught, Danesa, Frappier, Moreau, Phipps, Tice, Sueca (colección del PIMM, UNAM), Tokio 172 (donada por la Dra. Mercedes Macías-Parra, Instituto Nacional de Pediatría) y México (proporcionada por el Instituto Nacional de Higiene). En el caso de la BCG México se estudiaron tres BCGs que representan las diferentes etapas de producción de la vacuna BCG en el país y se denominaron BCG México 1931, 1988 y 1997.

El cultivo de todas las BCGs se realizó inicialmente en el medio sólido 7H10 (Anexo X.I) a 37°C por un período aproximado de 2 a 3 semanas. Posteriormente, las BCGs se crecieron en medio Sauton (Anexo X.II) a 37°C y 150 rpm por 15 días. Después de este tiempo los cultivos se cosecharon mediante centrifugación a 3500 rpm y se almacenaron a -70°C hasta su uso.

V.II Extracción de DNA

La extracción de DNA se realizó siguiendo el protocolo descrito en Mycobacterial Protocols con algunas modificaciones [56]. Una pastilla proveniente de aproximadamente 5 mL de cultivo de la cepa se colocó en un tubo eppendorf, se re-suspendió en 600 µL de TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA) y 40 µL de SDS al 10% y se incubó 30 min a 80°C. Posteriormente, se adicionaron 50µL de lisozima (20 mg/mL), 50 µL de RNAsa (10 mg/mL) y 20 µL de EDTA y se incubó la muestra a 37°C por una hora. Al terminar el tiempo de incubación, se adicionaron 30 µL de proteinasa K (50 mg/mL) y se incubó una hora a 65°C. Al final del procedimiento de lisis se realizaron dos extracciones con 1 volumen de fenol/cloroformo/alcohol isoamílico 24:24:1, se separó la fase acuosa y se realizó una extracción más con un volumen de cloroformo. La fase acuosa se precipitó con 1 volumen de isopropanol y 0.1 volúmenes de acetato de sodio 3M. El DNA obtenido se cuantificó por espectroscopía a 260 nm y se realizó una electroforesis en gel de agarosa al 1% para visualizar el DNA obtenido.

V.III Determinación del perfil de regiones RD y DU

La determinación del perfil de regiones RD y DU se realizó mediante PCR. En los dos ensayos se incluyeron las BCGs Danesa, Phipps, Sueca, Tice, Tokio y México 1931, 1988 y 1997 además de que se incluyó DNA de *M. tuberculosis* como control positivo. Todos los ensayos se realizaron por triplicado.

V.III.I Determinación del perfil de regiones RD por PCR múltiple

La determinación del perfil de regiones RD se realizó con el PCR múltiple reportado previamente por la OMS (comunicación personal) y el reportado por Bedwell y cols [37]. En conjunto, estos dos PCR múltiples permitieron identificar las regiones RD1, RD2, RD8, RD14, RD16, N-RD18, RD Dinamarca/Glaxo (D/G) y una copia de la secuencia de inserción IS6110.

El PCR múltiple reportado por la OMS permitió identificar 5 regiones RD. Este comprende el uso de tres oligos por cada región RD; de los cuales uno flanquea del lado izquierdo la región de interés (L, río arriba), otro del lado derecho (R, río abajo) y uno más se encuentra dentro de la región de interés (wtR). Los oligos L y R amplificaron cuando la región estuvo ausente, mientras que el L y wtR amplificaron cuando la región estaba presente. De esta manera, se obtuvieron productos de diferentes tamaños para cada región analizada (tabla 1).

El PCR se realizó en un volumen de 50 μ L con 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM MgCl₂, 0.4 mM dNTP's, 20 pmoles de cada oligo, 2 unidades de *Taq* Gold DNA polimerasa (Applied BioSystem, EEUU) y 100 ng de DNA. Para la amplificación se realizaron 35 ciclos constituidos por: desnaturalización a 94°C, alineación a 55°C y extensión a 72°C. Los productos de PCR se visualizaron por electroforesis en gel de agarosa al 3% (m/v) y solución amortiguadora de TAE 1X por 2 horas a 100 volts. Al término de la cual el gel se tiñó con bromuro de etidio.

Tabla 1. Secuencia de los oligos del PCR múltiple reportado por la OMS.

Región	Oligo	Secuencia	Tamaño del producto amplificado (pb)
RD1	RD1 L	AACAgCgCgTCgCTAgAC	L + R = 331 L + wtR = 598
	RD1 R	TTCAACgggTACTgCgAAT	
	RD1 wtR	ATCCCgTACACAAACCgTTC	
RD2	RD2 L	CgCAAgATCACCCgAgTgg	L + R = 539 L + wtR = 764
	RD2 R	TTgCgAgAAgCgAATTTACA	
	RD2 wtR	ACCATggCAGCAGCAGTAg	
RD14	RD14 L	AACTgAACTgCTCATCggC	L + R = 435 wtR + wtL = 297
	RD14 R	gggATTCggAAgTgTCAC	
	RD14 wtL	gTgTCgATCAACCggCAC	
	RD14 wtR	gtgCAggCgATgTCggTA	
N-RD18	RD18 L	AAATCCATgTCTTAgTgCAg	L + R = 134 L + wtR = 405
	RD18 R	TTgCgACTACCgTTgAgC	
	RD18 wtR	AggTCACgTAACTgCAGCg	
RD D/G	RD-D/G L	gAACCTggTgggAgACCT	L + R = 805 1636 sin delección
	RD-D/G R	CTgCAGCCCTTCgTCTTC	
IS6110	IS6110 L	TTgTCCTgTAgCgAgTCAC	L + R = 689 L + wtR = 318
	IS6110R	gCATATCgTCgTCCAATAg	
	IS6110 wtR	ggTACCTCCTCgATgAACC	

L + R= región ausente, L + wtR= región presente, wtR + wtL= región presente

Por otra parte, el PCR múltiple reportado por Bedwell permitió identificar 5 regiones RD, algunas de las cuales son diferentes a las del PCR anterior. En este PCR se incluyeron dos oligos por región, excepto para la RD1, los cuales amplificaron cuando la región RD estaba presente en la BCG estudiada. Este PCR se realizó con los oligos descritos en la tabla 2 y con las condiciones descritas por Bedwell [37].

V.III.II Determinación del perfil de regiones DU por PCR

La determinación de regiones DU se realizó por PCR con los oligos reportados por Brosch [34], los cuales son específicos para las regiones duplicadas en las BCGs: región duplicada 1 (DU1), región duplicada 2 tipo I (DU2 I), tipo 2 (DU2 II), tipo III (DU2 III) y la tipo IV (DU2 IV). Las secuencias de estos oligos se muestran en la tabla 3. En este ensayo se incluyeron las BCGs Danesa, Phipps, Tice, Tokio y México 1931, 1988 y 1997.

Tabla 2. Secuencia de los oligos del PCR múltiple reportado por Bedwell y cols [37].

Región	Oligo	Secuencia	Tamaño del producto amplificado (pb)
RD1	ET1	AAgCggTTgCCgCCgACCgACC	146 (presencia) 196 (delección)
	ET2	CTggCTATATTCCTgggCCCgg	
	ET3	gAggCgATCTggCggTTTgggg	
RD2	RD2l	CCAgATTCAAATgTCCgACC	315
	RD2r	gTgTCATAggTgATTggCTT	
RD8	RD8l	ACTCCTAgCTTTgCTgTgCgCT	472
	RD8r	gTACTgCgggATTTgCAggTTC	
RD14	RD14l	CAGggTTgAAggAATgCgTgTC	252
	RD14r	CTggTACACCTggggAATCTgg	
RD16	RD16l	ATCgTTCACggACAgCCgTAgT	401
	RD16r	CTCgATCCAaggTCAACCACg	

Para cada par de oligos, la amplificación se realizó en un volumen de 30 μ L con 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM MgCl₂, 0.1 mM dNTP's, 20 pmoles de cada oligo, 1 unidad de *Taq* DNA polimerasa Platinum y 50 ng de DNA. Para la amplificación se realizaron 35 ciclos constituidos por: desnaturalización a 94°C, alineación a 55°C y extensión a 72°C. En todos los casos, los productos de PCR se analizaron por electroforesis horizontal en geles de agarosa al 3% (m/v), en solución amortiguadora TAE 1X por 2 horas a 100 volts y teñidos con bromuro de etidio.

Tabla 3. Secuencia de los oligos utilizados para la detección de DUs.

Región	Iniciador	Secuencia
DU2 I	DU2- IF	CAAAgACACCACCCgATTCT
	DU2- IR	gAAAATgCggACTggAATgT
DU2 II	DU2-IIR	CTggggATTCTggTTggTTA
	DU2-IIF	CCgAgACgTTgAAgAATCg
DU2 III	DU2-IIIR	gATCgACCgCTggCTAACT
	DU2-IIIF	AgCCggTAgTCgAgTTCCAT
DU2 IV	DU2-IVR	CgCAAACCCgTgCAgTC
	DU2-IVR	gTgAgTACgCgCTCAAACg
DU1	DU1-F	ACggTCggTgTCgTCgTCTAAgT
	DU1-R	AgAACTgCAggggTggTACA
Δ Int	Δ int- F	CgTTAAgCgTCgTgTgTTTC
	Δ int- R	TCgACggTCTggACATAgTg

V.IV Determinación del perfil genómico de la BCG México 1931 por PFGE

La electroforesis en gel por campos pulsados (PFGE) permite obtener el perfil del genoma completo de la bacteria de interés con una enzima de restricción de corte poco frecuente, que para BCG podrían ser *DraI*, *AsnI* o *SpeI*. En este trabajo se determinó el perfil genómico de la BCG México 1931 y se comparó con los de las BCGs Danesa, Tice y Phipps.

Para realizar este ensayo se siguió el protocolo descrito en el Mycobacterial Protocols, con algunas modificaciones realizadas durante la estandarización del ensayo [57]. Las BCGs se cultivaron en medio Sauton a una DO_{600nm} 0.8-1.0 e incubadas con glicina (0.2M) 24h antes de finalizar el cultivo. Posteriormente, los cultivos se centrifugaron y resuspendieron en TE, la suspensión bacteriana obtenida se mezcló con un volumen de agarosa al 1.2% y se distribuyó en moldes para obtener bloques de agarosa para cada BCG. Estos bloques se incubaron toda la noche a 37°C en solución de lisis (lisozima 1 mg/mL). Al término, la solución se eliminó y los bloques se colocaron en solución con proteinasa K (2 mg/mL) por 48 horas a 55°C. Transcurrido este tiempo, los bloques se lavaron con TE y se pre-incubaron con solución amortiguadora de restricción enzimática por 1 hora y después toda la noche con *AsnI* (40U) a 37°C. Al día siguiente, se adicionaron 40U más de la enzima y se incubaron 4 horas más a 37°C.

Finalmente, los bloques de agarosa se colocaron en un gel de agarosa al 1% y la electroforesis se realizó por 24 horas (5s, 10s y 15s por 4.5h y 20s, 25s y 35s por 3.5h) a 10°C y 150 volts. El gel se tiñó con bromuro de etidio por 20 minutos.

V.V Determinación del genoma de la BCG México 1931

La obtención del genoma de la BCG México 1931 comprendió diversas etapas entre las que se encuentran la secuenciación, el ensamble y cierre del genoma; además de la anotación del mismo y su comparación con los genomas de otras BCGs.

V.V.I Secuenciación

La secuenciación del genoma de la BCG México 1931 se realizó por el método de pirosecuenciación para lo cual se enviaron 30 µg de DNA genómico a la Unidad de Secuenciación del CINVESTAV Irapuato donde se realizaron dos secuenciaciones completas de su genoma en el equipo GSFLX 454 de Roche, el cual tiene un rendimiento por secuenciación de alrededor de 70 MB.

V.V.II Ensamble

El ensamble del genoma se realizó con el software “GS de novo Assembler” intrínseco del secuenciador GSFLX 454 de Roche y los programas Phred/Phrap/Consed [58].

El primer paso consistió en realizar un ensamble preliminar del genoma usando como genoma de referencia el de BCG Pasteur 1173P2 (GenBank: AM408590). Este ensamble se realizó con las lecturas obtenidas de la pirosecuenciación y permitió identificar diferencias entre las secuencias de BCG México y Pasteur como SNPs, deleciones y duplicaciones. Adicionalmente, se construyó un ensamble utilizando como secuencia de referencia *M. bovis* AF2122/97 (GenBank: BX248333) ya que al emplear como secuencia de referencia la BCG Pasteur no fue posible identificar regiones presentes en México pero ausentes en BCG Pasteur. El análisis de estos dos ensamblados nos permitió obtener una secuencia de referencia, llamada BCG Pasteur modificada, la cual se empleó como secuencia de referencia para el ensamble final del genoma de la BCG México 1931.

A la par se construyó una librería de fósmidos (Anexo X.III) de la cual se seleccionaron 250 clones al azar, se extrajo el DNA con el “High Pure Plasmid Isolation Kit” (Roche, Alemania) y se realizó una secuenciación “fos-end” por el método Sanger con el secuenciador 3730XL DNA (Applied Biosystems, EEUU). La secuenciación “fos-end” comprende la secuenciación de las regiones 5’ y 3’ de los fósmidos, las lecturas obtenidas se incorporaron al ensamble y esto permitió verificar el ensamble del genoma así como definir las zonas de duplicación DU1 y DU2 descritas previamente en las BCGs.

Así, el ensamblaje generado con las lecturas de pirosecuenciación, más las obtenidas de los fósmidos mostró una cobertura del 99% con respecto a la referencia, con 143 espacios distribuidos a lo largo del genoma. Estos espacios se cerraron como se indica en la siguiente sección.

V.V.III Cierre

Ya definidos los espacios en el genoma el siguiente paso fue el planteamiento de la estrategia para su cobertura. Debido a que los espacios fueron relativamente pequeños, de 2 a 2500pb, se decidió diseñar oligos específicos para cada uno ellos. Además se secuenciaron 3 fósmidos. La estrategia seguida se resume en la figura 6.

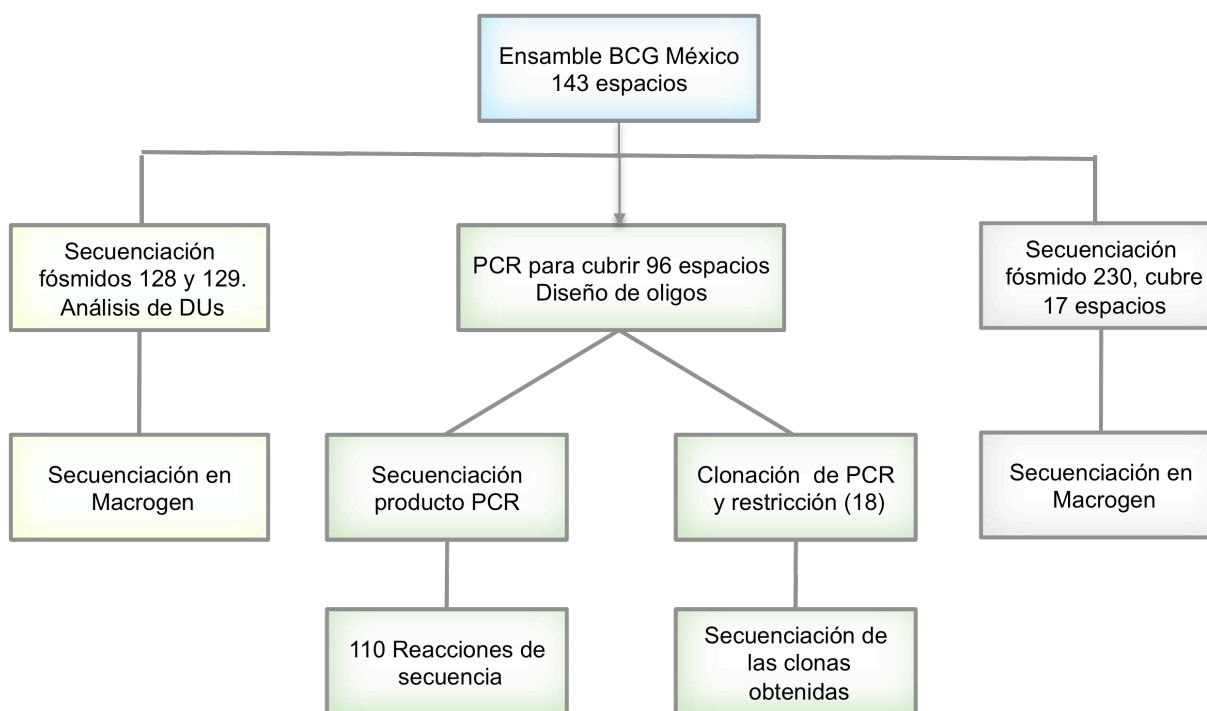


Figura 6. Estrategia seguida para cerrar los espacios en el genoma de la BCG México 1931.

Diseño de oligonucleótidos

Se diseñó un par de oligos para cada espacio presente en el genoma. Sin embargo, en los casos en que los espacios fueron pequeños y cercanos se diseñó un par de oligos para cubrir más de un espacio. Los oligos se diseñaron con ayuda de los programas “Consed” y “Oligo Calc: Oligonucleotide Properties Calculator” [58, 59] bajo las siguientes consideraciones: tamaño entre 17 y 21 bases, T_m de 55 a 60°C y especificidad.

Los oligos se dividieron en tres grupos con base en su T_m (55, 58 y 60°C) para llevar a cabo las reacciones de PCR. A cada mezcla de reacción (50 µL) se le adicionó: solución amortiguadora 1X (600 mM Tris-SO₄ pH 8.9, 180 mM sulfato de amonio), sulfato de magnesio 1.5 mM, dNTP's 0.1 mM, 1 unidad de *Taq* DNA polimerasa High Fidelity Platinum, 20 pmoles de cada oligo y DMSO al 10%. Las diferentes secuencias se amplificaron durante 30 ciclos constituidos por: desnaturalización a 94° C, alineación a la T_m determinada y extensión a 72° C. Los productos de PCR obtenidos se visualizaron mediante electroforesis en gel de agarosa al 2% en solución amortiguadora TBE 0.5X por 80 minutos a 110 volts. En los casos en los que se obtuvieron productos inespecíficos, se realizó una purificación a partir de geles de agarosa (Anexo X.V).

Los productos obtenidos se secuenciaron por el método Sanger con el secuenciador 3730XL DNA (Applied Biosystems, EEUU) en Macrogen (Seúl, Korea) y las lecturas adicionadas a los ensamblajes realizados previamente.

Secuenciación de Fósmidos

Adicionalmente a la secuenciación de los productos de PCR, se realizó la secuenciación de tres fósmidos: 128, 129 y 230. Para la obtención del DNA, se realizó el cultivo de cada uno de los fósmidos en 5 mL de medio Luria (Anexo X.IV) y el DNA se extrajo con el “High Pure Plasmid Isolation Kit” (Roche, Alemania). La secuenciación se realizó por el método Sanger con el secuenciador 3730XL DNA (Applied Biosystems, EEUU) por la compañía Macrogen (Seúl, Korea). Las lecturas generadas se adicionaron al ensamblaje descrito en la sección anterior.

El fósido 230 permitió cubrir una región donde se concentraron 17 espacios en el genoma. Mientras que los fósidos 128 y 129, permitieron definir las regiones DU1 y DU2 en la BCG México.

V.V.IV Anotación

Por lo que respecta a la anotación, ésta se realizó con los programas RAST Server [60], Artemis [61] y BCGList [62]. El primero se empleó para realizar una anotación rápida del genoma, posteriormente los archivos generados se analizaron con el programa Artemis en donde se realizó la anotación del genoma completo. Finalmente, las bases de datos BCGList y Uniprot se utilizaron para el análisis de los genes anotados.

V.VI Análisis del genoma

La secuencia del genoma de la BCG México 1931 se comparó con las secuencias de BCG Pasteur y BCG Tokio (GenBank: AP010918) empleando los programas Consed y BLAST (Basic Local Alignment and Search Tool) [63].

En ambos casos se realizó una comparación directa de las secuencias genómicas. Con el BLAST se logró tener una comparación general entre las tres BCGs analizadas. Sin embargo, la comparación con el programa Consed permitió determinar de manera precisa la presencia de deleciones, duplicaciones y SNPs entre las tres BCGs analizadas.

Finalmente, se realizó la detección de SNPs entre las BCGs México 1931, Pasteur y Tokio con un programa en Perl construido por el Dr. Enrique Merino (IBT, UNAM) en el cual buscó específicamente polimorfismos de un solo nucleótido entre las secuencias de estas BCGs.

V.VII Identificación de nuevas regiones RD en otras BCGs

El análisis del genoma de la BCG México 1931 llevó a la detección de tres deleciones no descritas previamente en las BCGs, las cuales se denominaron RDMex01, RDMex02 y RDMex03. Con la finalidad de conocer si estas deleciones son únicas de la BCG México se diseñaron oligos específicos que permitieron determinar la presencia/ausencia de cada una de éstas regiones en diferentes BCGs.

Los oligos usados fueron 5'-GCCCAAACAGTTCGACGG-3', 5'-CCAGACATATGCGAGGAC-3' para RDMex01; 5'-GCGATAATGGGCGATGTCG-3', 5'-CCGCGGTTGTTGAGTTCG-3' para RDMex02; y 5'-GCAGCAGTGAACGCTTGG-3', 5'-CAGTGGAGCTGAAGGCAG-3' y 5'-GTTGCTTGGACGGCAATCG-3' para RDMex03. Las reacciones de PCR se realizaron en un volumen de 30 μ L con 20mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM MgCl₂, 0.1 mM dNTP's, 20 pmoles de cada iniciador, 0.4 unidades de *Taq* DNA polimerasa Platinum y 20ng de DNA. Para la amplificación se realizaron 35 ciclos constituidos por: desnaturalización a 94° C, alineación a 58° C y extensión a 72° C. Los productos de PCR se analizaron por electroforesis horizontal en geles de agarosa al 2% (m/V) y teñidos con bromuro de etidio.

VI. RESULTADOS

VI.I Perfil genético de regiones RD y DU de BCG México

VI.I.I Perfil genético de regiones RD

Los resultados del PCR múltiple reportado por la OMS se muestran en la tabla 4. En el ensayo se utilizaron 5 BCGs, cuyos perfiles se describieron previamente, y *M. tuberculosis* como control positivo. En todos los casos los perfiles obtenidos con el PCR concordaron con los reportados previamente.

Con respecto a las BCG usadas para la producción de la vacuna BCG en México, los resultados mostraron que la BCG México 1931 presenta un perfil de RDs diferente al de las BCGs México 1988 y 1997. Esta diferencia se presentó en las regiones N-RD18 y RD D/G, la primera presente únicamente en la BCGs México 1988 y 1997 mientras que la segunda se encontró solamente en la BCG México 1931.

Tabla 4. Perfil de regiones RD obtenido con el PCR múltiple reportado por la OMS.

BCG	Perfil de RDs					
	RD1	RD2	RD14	N-RD18	RD D/G	IS6110
<i>M. tuberculosis</i> H37Rv	+	+	+	+	+	+
Danesa	-	-	+	+	-	-
Phipps	-	-	+	-	+	-
Sueca	-	+	+	+	+	-
Tice	-	-	+	-	+	-
Tokio	-	+	+	+	+	+
México 1931	-	-	+	-	+	-
México 1988	-	-	+	+	-	-
México 1997	-	-	+	+	-	-

El símbolo (+) indica que la región esta presente, el (-) indica la pérdida de la región. En el caso de la IS6110 el símbolo (-) indica la pérdida de una copia de esta secuencia.

Adicionalmente, los resultados obtenidos con el PCR múltiple reportado por Bedwell se muestran en la tabla 5. Para este PCR se emplearon las mismas BCGs que en el PCR previamente descrito y los perfiles obtenidos también concuerdan con los reportados previamente para las BCGs analizadas.

En el caso de las BCGs México 1931, 1988 y 1997, los resultados mostraron un perfil idéntico de regiones RD entre éstas, con la pérdida de las regiones RD1 y RD2 y la presencia de las regiones RD8, RD14 y RD16.

En conjunto, los resultados obtenidos en ambos PCR múltiples mostraron que la BCG México 1931 conserva las regiones RD8, RD14, RD16 y RD D/G y perdió las regiones RD1, RD2, N-RD18, así como una copia de la secuencia de inserción IS6110 (tabla 6). Un perfil indistinguible del que presentaron las BCGs Phipps y Tice. Mientras que las cepas BCG México 1988 y 1997 mostraron un perfil idéntico entre sí, con la ausencia de las regiones RD1, RD2, RD D/G y de una copia de la secuencia de inserción IS6110 (tabla 6). Este perfil fue idéntico a la BCG Danesa, la ausencia de la región RD D/G (específica de la BCG Danesa) en la BCG México 1988 y 1997 confirma este hallazgo.

Tabla 5. Perfil de regiones RD obtenido con el PCR múltiple reportado por Bedwell.

BCG	Perfil de RDs				
	RD1	RD2	RD8	RD14	RD 16
<i>M. tuberculosis</i> H37Rv	+	+	+	+	+
Danesa	-	-	+	+	+
Phipps	-	-	+	+	+
Sueca	-	+	+	+	+
Tice	-	-	+	+	+
Tokio	-	+	+	+	+
México 1931	-	-	+	+	+
México 1988	-	-	+	+	+
México 1997	-	-	+	+	+

El símbolo (+) indica que la región esta presente, el (-) indica la pérdida de la región.

Tabla 6. Perfil de regiones RD en las BCGs México.

BCG	Perfil de regiones RD							
	RD1	RD2	RD8	RD14	RD16	N-RD18	RDD/G	IS6110
México 1931	-	-	+	+	+	-	+	+
Phipps	-	-	+	+	+	-	+	+
Tice	-	-	+	+	+	-	+	+
México 1988	-	-	+	+	+	+	-	+
México 1997	-	-	+	+	+	+	-	+
Danesa	-	-	+	+	+	+	-	+

El símbolo (+) indica que la región esta presente, el (-) indica la pérdida de la región. En el caso de la IS6110 el símbolo (-) indica la pérdida de una copia de ésta secuencia.

VI.I.II Perfil genético de regiones DU

La identificación del perfil de regiones DU en las BCGs México se realizó por PCR. En este ensayo se incluyeron 4 BCGs que fueron previamente caracterizadas, con la finalidad de comparar los perfiles obtenidos para las BCGs México con los que presentan estas cepas.

Los resultados mostraron que el perfil de amplificación de la BCG México 1931 fue diferente a las BCGs México 1988 y 1997, basado en la amplificación con los oligos DU2-IV (tabla 7). Además, la BCG México 1931 presentó el mismo perfil de amplificación que BCG Phipps, con la duplicación DU2 tipo IV. En contraste, las BCGs México 1988 y 1997 presentaron la duplicación DU2 tipo III, el mismo perfil que la BCG (tabla 7).

Estos resultados permitieron complementar el perfil genético obtenido para las BCGs México con el análisis de regiones RD.

Tabla 7. Perfil de regiones DU de las BCGs México.

BCG	Amplificación con los oligos					Δ int
	DU1	DU2-I	DU2-II	DU2-III	DU2-IV	
Danesa	+	-	-	+	-	+
Phipps	+	-	-	+	+	+
Tice	+	-	-	+	+	+
Tokio	+	+	-	-	-	-
México 1931	+	-	-	+	+	+
México 1988	+	-	-	+	-	+
México 1997	+	-	-	+	-	+

El símbolo (+) indica una amplificación positiva con los iniciadores probados, mientras que el (-) indica la ausencia de amplificación.

De esta manera, la determinación del perfil de regiones RD y DU demostró que la BCG México 1988 y la BCG México 1997 presentan el mismo perfil que la BCG Danesa, al ubicarse dentro del grupo DU2 tipo III. Mientras que la BCG México 1931 es una cepa diferente a las dos anteriores ya que los perfiles obtenidos la ubican dentro del grupo DU2 tipo IV.

Estos resultados permitieron establecer que la BCG México 1931 es diferente a las BCGs identificadas como BCG México 1988 y 1997 las cuales de acuerdo a los registros históricos y a los perfiles obtenidos son derivadas de la BCG Danesa. Una vez establecida la identidad de la BCG, se procedió a realizar únicamente la caracterización genómica de la BCG México 1931.

VI.II Perfil genético de la BCG México 1931 por PFGE

Para conocer el perfil genómico de la BCG México 1931 se realizó una electroforesis en gel por campos pulsados (PFGE) con la enzima *AsnI*, en el ensayo también se incluyeron las BCGs Danesa, Phipps y Tice. Este ensayo mostró diferentes perfiles entre las BCGs estudiadas, con 18, 18, 15 y 14 bandas resueltas para las BCGs Danesa, México 1931, Phipps y Tice, respectivamente (figura 7).

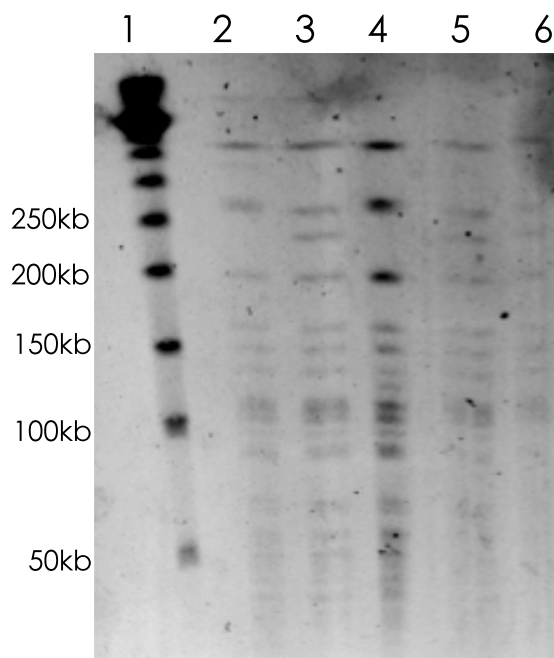


Figura 7. Perfil genético de la BCG México 1931 obtenido por PFGE y restricción enzimática con *AsnI*. Los carriles del 1 al 7 corresponden a: MMØ lambda, BCG Danesa, México 1931, Danesa, Phipps, Tice y MMØ lambda. Gel de agarosa al 1% teñido con bromuro de etidio al 0.02%.

Así, en el perfil genómico de la BCG México 1931 se detectaron 18 bandas en un intervalo de 36 a 450kb y cuya suma fue de 2 600kb, lo que corresponden al 50% del tamaño esperado del genoma que es de aproximadamente de 4 400kb. Este perfil resultó muy similar a los de BCG Phipps y Tice, con la presencia de una banda de 220kb, la cual se encuentra ausente en la BCG Danesa. Sin embargo, la BCG México 1931 presenta diferencias con BCG Phipps y Tice en las bandas que están entre 50 y 100kb.

Aunque en la literatura no existe una descripción detallada del perfil electroforético de BCG Phipps y Tice, Brosch y *cols* mostraron que para la BCG Danesa y Pasteur las bandas ubicadas entre 200 y 300kb corresponden a la región duplicada DU2 [34].

Los resultados obtenidos en este ensayo mostraron que la BCG México 1931 posee un perfil genómico diferente al de las BCGs estudiadas previamente, por lo que se decidió realizar la secuenciación del genoma completo.

VI.III Secuenciación del genoma completo de la BCG México 1931

VI.III.I Secuenciación

Un total de 30µg de DNA genómico de la BCG México 1931 se enviaron a la Unidad de Secuenciación del CINVESTAV Irapuato para su secuenciación. Aquí se realizaron dos corridas de secuenciación en el equipo GSFLX 454 de Roche. Cada secuenciación garantiza 70MB de secuencia, por lo cual después de dos corridas se obtuvo una cobertura de más de 30 veces el genoma.

La Unidad del Cinvestav Irapuato entregó un ensamble preliminar con los resultados generados después de cada una de las corridas de secuenciación. Este ensamble fue realizado con el software “GS De Novo Assembler” intrínseco del secuenciador. En el ensamble las secuencias obtenidas se agrupan en contigs, los cuales están constituidos por un conjunto de secuencias de DNA que se sobrelapan.

De esta manera, el resultado fue un archivo con 1089 contigs de un tamaño entre 100 y 129 413 bases para la primera secuenciación; mientras que las secuencias generadas en la primera y segunda secuenciación se obtuvieron 1521 contigs de un tamaño entre 100 y 175 416 (figura 8A). Además, también se encontró un mayor número de lecturas por contig después de la segunda secuenciación en comparación con la primera secuenciación (figura 8B). Lo que muestra que la segunda secuenciación permitió obtener contigs de mayor longitud y profundidad.

Estos resultados nos permitieron tener una mejor aproximación del comportamiento de las secuencias obtenidas, sin embargo, estos contigs no se emplearon para el ensamble final del genoma debido a que el ensamblador del equipo presenta algunas desventajas, una de ellas es que las secuencias que cubren regiones repetidas se agrupan en un solo contig.

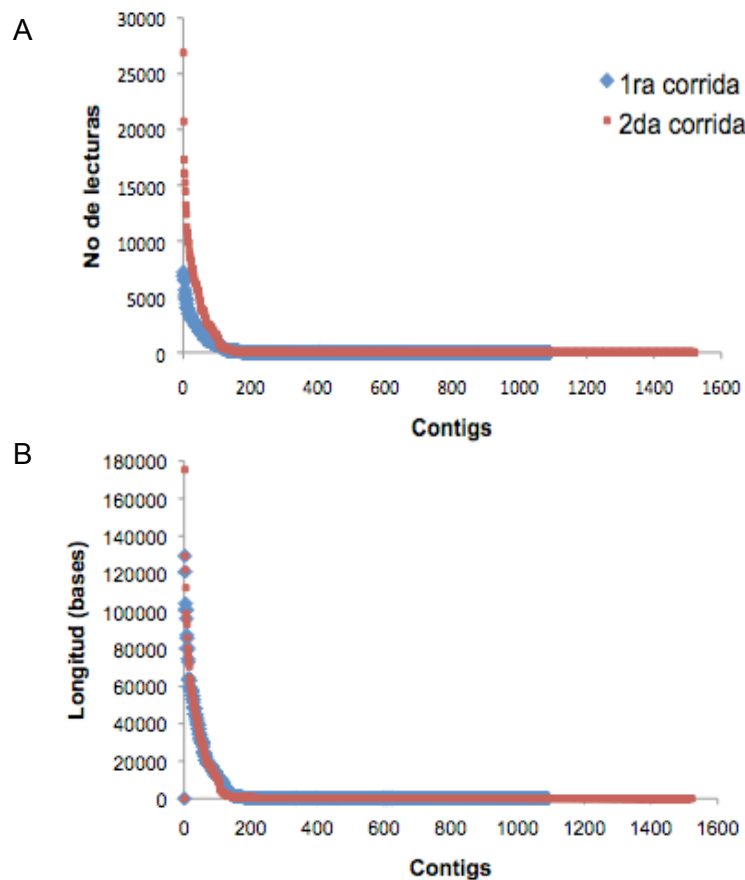
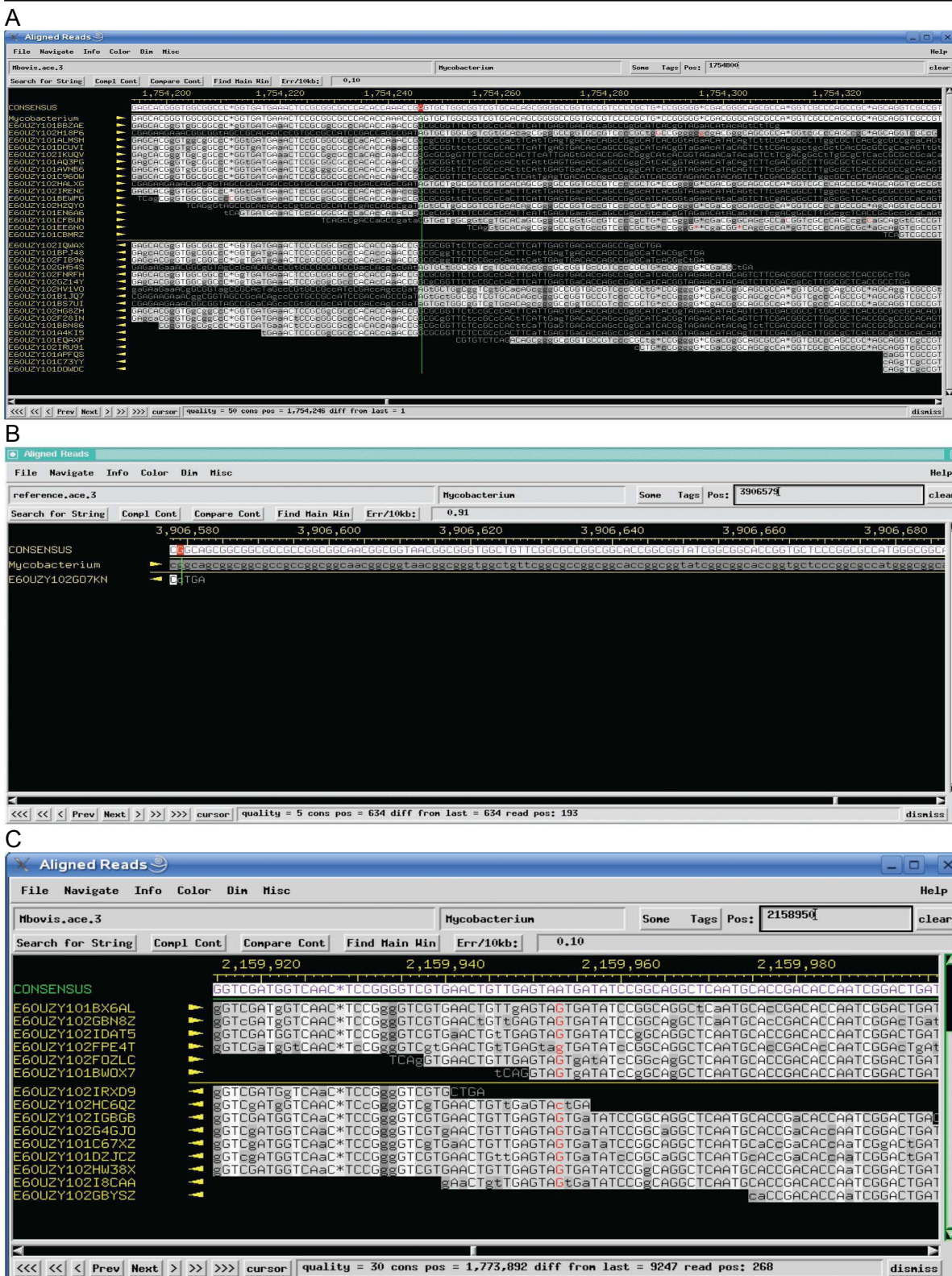


Figura 8. Resultados obtenidos después de la primera y segunda secuenciación con el equipo GSFLX 454 de Roche. La gráfica A muestra la longitud de los contigs (en bases) obtenidos en la primera y segunda secuenciación; mientras que en la gráfica B se observa el número de lecturas que conforman un contig.

VI.III.II Ensamble

El ensamble se realizó con los programas “phred, phrap, consed” del Departamento de Ciencias Genómicas de la Universidad de Washington.

El ensamble preliminar se hizo utilizando como secuencia de referencia el genoma de la BCG Pasteur y las lecturas generadas en las dos corridas de secuenciación. Este ensamble permitió identificar diferencias entre las secuencias como SNPs, deleciones y duplicaciones como se muestra en la figura 9.



Además se identificaron cuatro regiones (55, 655, 2847 y 29667pb) ausentes en BCG México 1931 con respecto a BCG Pasteur. Las cuales inicialmente se identificaron como espacios, pero un análisis posterior mostró que las secuencias en los extremos de estos espacios se sobrelapan indicando la presencia de una delección.

Por otra parte, se construyó un ensamble adicional utilizando como referencia la secuencia de *M. bovis* ya que el ensamble realizado con BCG Pasteur no permitió identificar regiones presentes en BCG México 1931 pero ausentes en BCG Pasteur. Este ensamble confirmó las delecciones en las regiones RD1, RD2, N-RD18 y la presencia de la región RD14 en la cepa.

Para obtener el ensamble final del genoma de la BCG México 1931, se construyó una secuencia de referencia tomando como base el genoma de la BCG Pasteur y los ensamblados mencionados anteriormente, denominada secuencia BCG Pasteur modificada. Así, se adicionó una región de 9 068pb que corresponde a la RD14, región ausente en BCG Pasteur pero presente en BCG México en comparación con *M. bovis*. Además se eliminaron las cuatro regiones que se encuentran ausentes en BCG México en comparación con BCG Pasteur. Lo anterior dio como resultado una secuencia de referencia con un tamaño de 4 350 386pb.

El ensamble final se construyó con la BCG Pasteur modificada, las lecturas obtenidas en las dos corridas de secuenciación y las lecturas generadas durante la secuenciación “fos-end” de la librería de fósidos. En total se consideraron 622 426 lecturas lo que representó una cobertura de 36X el genoma de la BCG México 1931.

En la figura 10 se muestra una representación lineal del genoma de la BCG México 1931 obtenido al término del ensamble. Después del análisis se encontraron 143 espacios en el genoma, lo que representa una cobertura del 99% con respecto a la referencia. Los espacios se encuentran distribuidos a lo largo del genoma, únicamente una región alrededor de 3 900 000 concentra 17 espacios.



Figura 10. Representación lineal del genoma de BCG México 1931. Los rectángulos naranja representan la secuencia cubierta del genoma, los espacios en blanco entre éstos los espacios no cubiertos y las cajas azules los fósmidos secuenciados en la región 5' y 3' (secuenciación "fos-end"). Para la construcción del mapa se empleó el programa *GenVision v0.99 Beta*.

VI.III.III Cierre del genoma

Ya definidos los espacios en el genoma el siguiente paso fue diseñar la estrategia para resolverlos. Debido a que los espacios fueron relativamente pequeños, de 2 a 2 500pb, se decidió diseñar oligos específicos para cubrir cada una de estas regiones. A excepción de la región mencionada anteriormente en donde tenemos 17 espacios, para este caso se consideró la secuenciación del fósmino 230 que cubre esta región.

Diseño de oligonucleótidos

Para cubrir los 143 espacios encontrados en el genoma de BCG México 1931 se diseñaron un total de 110 pares de oligos, con los programas y condiciones descritas previamente en la metodología.

Los oligos se dividieron en tres grupos con base en su T_m (55, 58 y 60°C) para llevar a cabo las reacciones de amplificación respectivas (PCR). Es importante mencionar que para obtener los productos de amplificación esperados fue necesario adicionar dimetil-sulfóxido al 10% (DMSO), esto debido a que la mayoría de las regiones sin cubrir correspondían a regiones con un alto contenido de G+C y el DMSO ayuda a relajar las cadenas de DNA durante la reacción de PCR.

En la figura 11A se muestran los productos de PCR obtenidos para dos pares de oligos con y sin la adición de DMSO en la reacción. Además, se muestra una imagen representativa de los productos de PCR obtenidos a una T_m determinada (figura 11B).

Los productos de PCR se secuenciaron por el método Sanger y posteriormente las secuencias obtenidas se adicionaron al ensamble descrito en la sección anterior. En algunos casos se decidió repetir las secuencias debido a que los cromatogramas fueron de baja calidad, e incluso en algunos otros se diseñaron nuevos oligos.

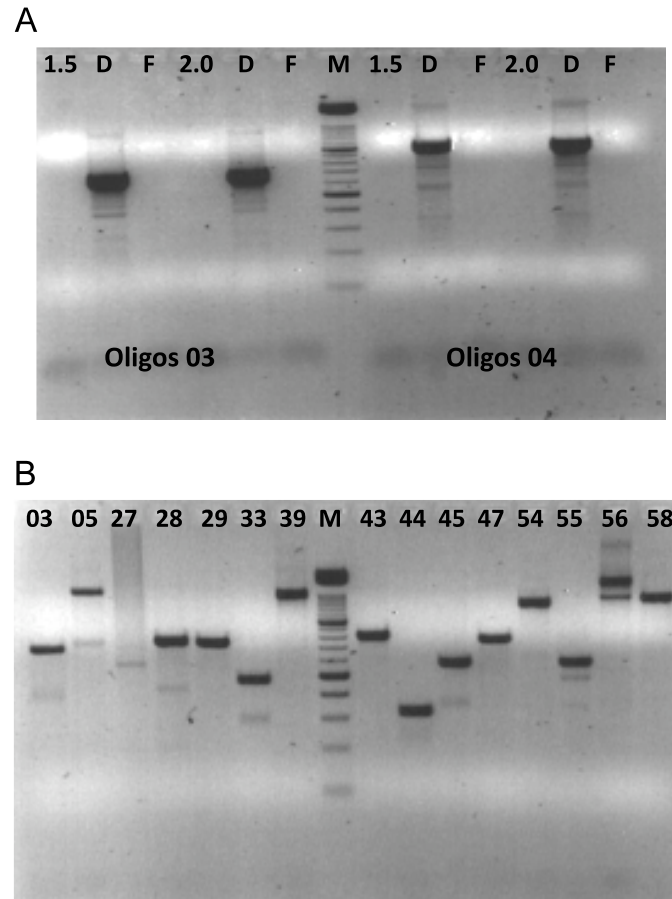


Figura 11. Imagen representativa de los productos de PCR obtenidos. Electroforesis en gel de agarosa al 2%. A) Resultados de la amplificación de dos pares de oligos con dos concentraciones de magnesio (1.5 y 2.0) y DMSO (D) o Formamida (F). B) Resultados de amplificación de 15 pares de oligos a una T_m de 55°C. M=Marcador de masa molecular (100 pb).

Secuenciación de fósmidos

Los fósmidos 128, 129 y 230 se secuenciaron por el método Sanger. Los resultados de la secuenciación se muestran en la tabla 8. Para los fósmidos 128 y 129 se obtuvo una cobertura del 97 y 95%, respectivamente. Mientras que para el fósrido 230 la cobertura obtenida fue del 67%, la cual resultó suficiente para cubrir los 17 espacios encontrados en el genoma en esta región.

Tabla 8. Resultado de la secuenciación de los fósmidos 128, 129 y 230.

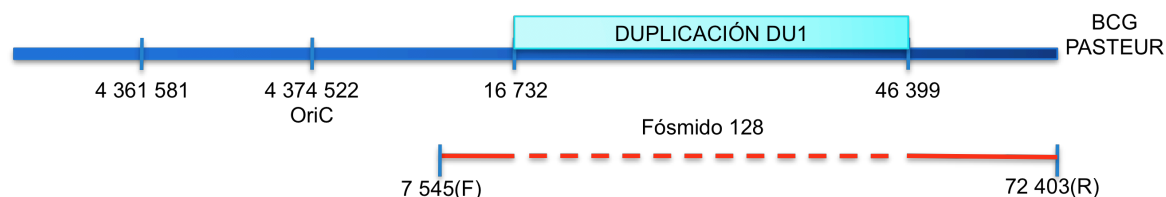
Fósmido	Tamaño (kb)	Número de contigs y tamaño (pb)	Región cubierta
128	42	1) 21604 2) 4817 3) 8202	DU1
129	38	1) 4350 2) 1168 3) 29637	DU2
230	28	1) 813 2) 649 3) 12286 4) 1373 5) 998 6) 2642	17 espacios

En la figura 12 se muestra una representación esquemática de las regiones genéticas que se lograron definir en el genoma de la BCG México 1931 con la secuenciación de los fósmidos 128 y 129. El fósmino 128 comprende la región DU1 y de acuerdo a los resultados obtenidos en la secuenciación, esta duplicación no se encuentra en BCG México 1931, lo cual concuerda con lo que se observó previamente en los ensamblados. Por lo que respecta al fósmino 129, éste cubre la región DU2 y los resultados obtenidos indican que esta duplicación se encuentra presente en BCG México 1931.

La secuenciación de los fósmidos y el diseño de oligos mencionado anteriormente permitió el cierre del genoma de la BCG México 1931. Lo que llevó a determinar que la BCG México 1931 posee un cromosoma circular de 4 350 386pb con un contenido de G+C del 65.7%.

El genoma completo y anotado de la BCG México 1931 se depositó en la base de datos GenBank del NCBI, con el número de acceso CP002095.

FÓSMIDO 128



FÓSMIDO 129

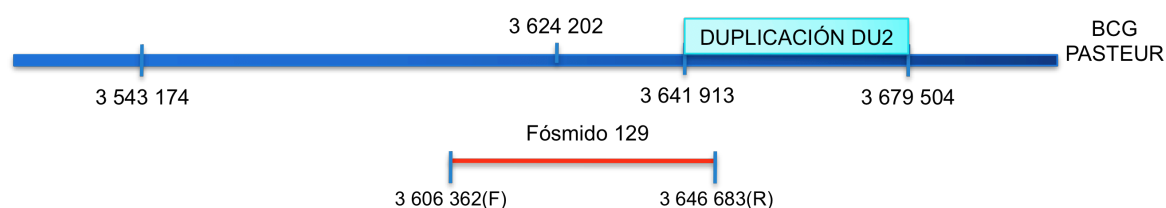


Figura 12. Representación esquemática de las regiones cubiertas por los fósidos 128 y 129, tomando como base la secuencia del genoma de BCG Pasteur.

VI.III.IV Anotación del genoma

La anotación preliminar del genoma se realizó con el programa RAST Server, el cual a partir de la secuencia del genoma genera una anotación utilizando una base de datos conocida como Pfam. El archivo generado se analizó con el programa Artemis en donde se verificó gen por gen que éstos estuvieran bien anotados, para lo cual se empleó la base de datos BCGList y Uniprot. Estas últimas contienen los genes anotados y realizando un alineamiento se verificó que los marcos de lectura correspondieran con los descritos previamente o si no determinar si existía algún polimorfismo y realizar la anotación correspondiente. En la figura 13 se observa una imagen de la anotación realizada en este programa. La anotación y predicción de las secuencias codificantes (CDS) mostró que el genoma completo de la BCG México 1931 contiene 3 904 CDS, 3 genes que codifican para rRNA y 45 genes para tRNA, adicionalmente se identificaron 29 pseudogenes (Figura 14) [64].

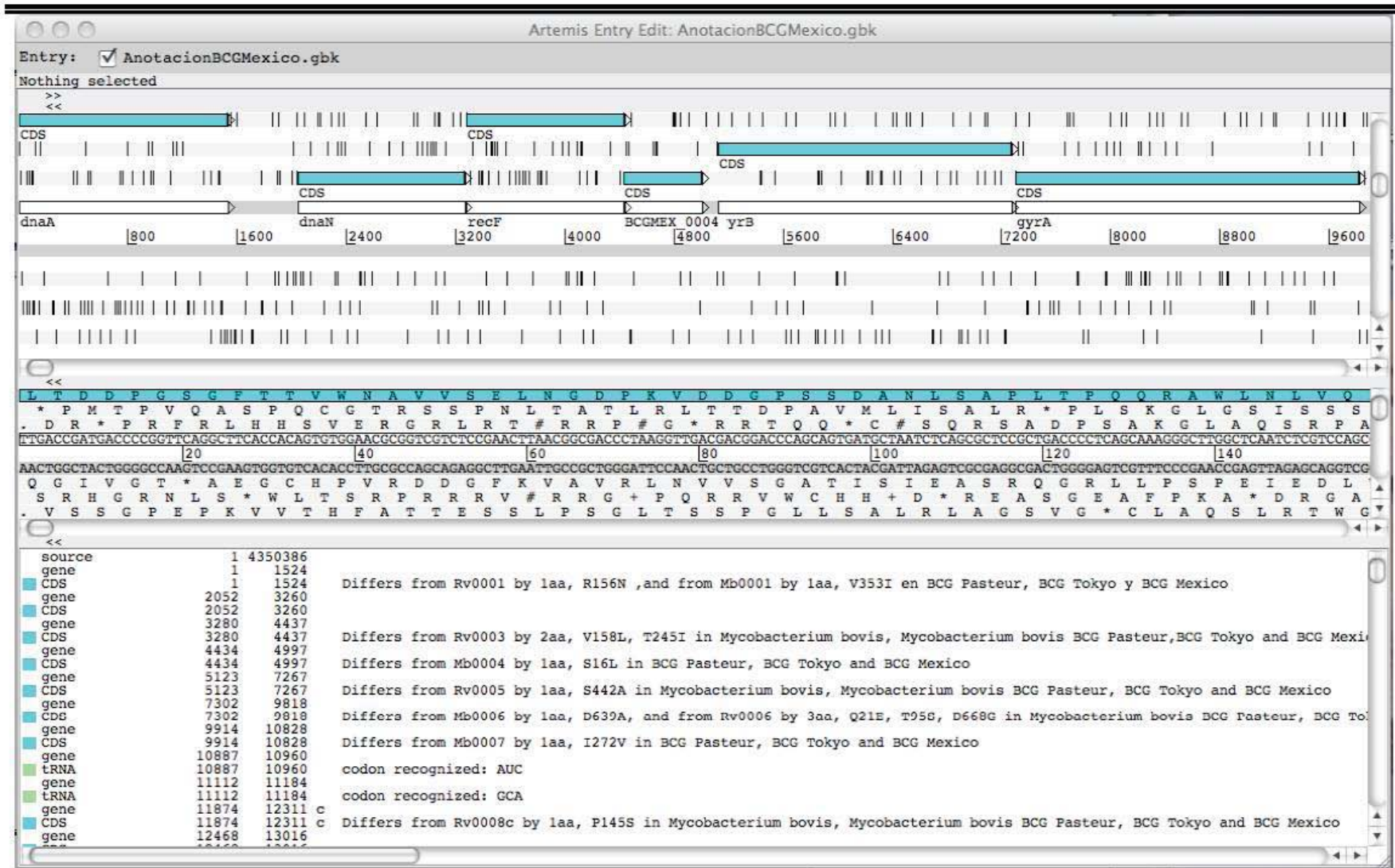


Figura 13. Anotación del genoma de BCG México 1931 en el programa Artemis. En la parte superior se observan los genes y los CDS anotados, en la parte media se observa la secuencia de nucleótidos y aminoácidos, y finalmente en la parte inferior se observan las anotaciones realizadas.

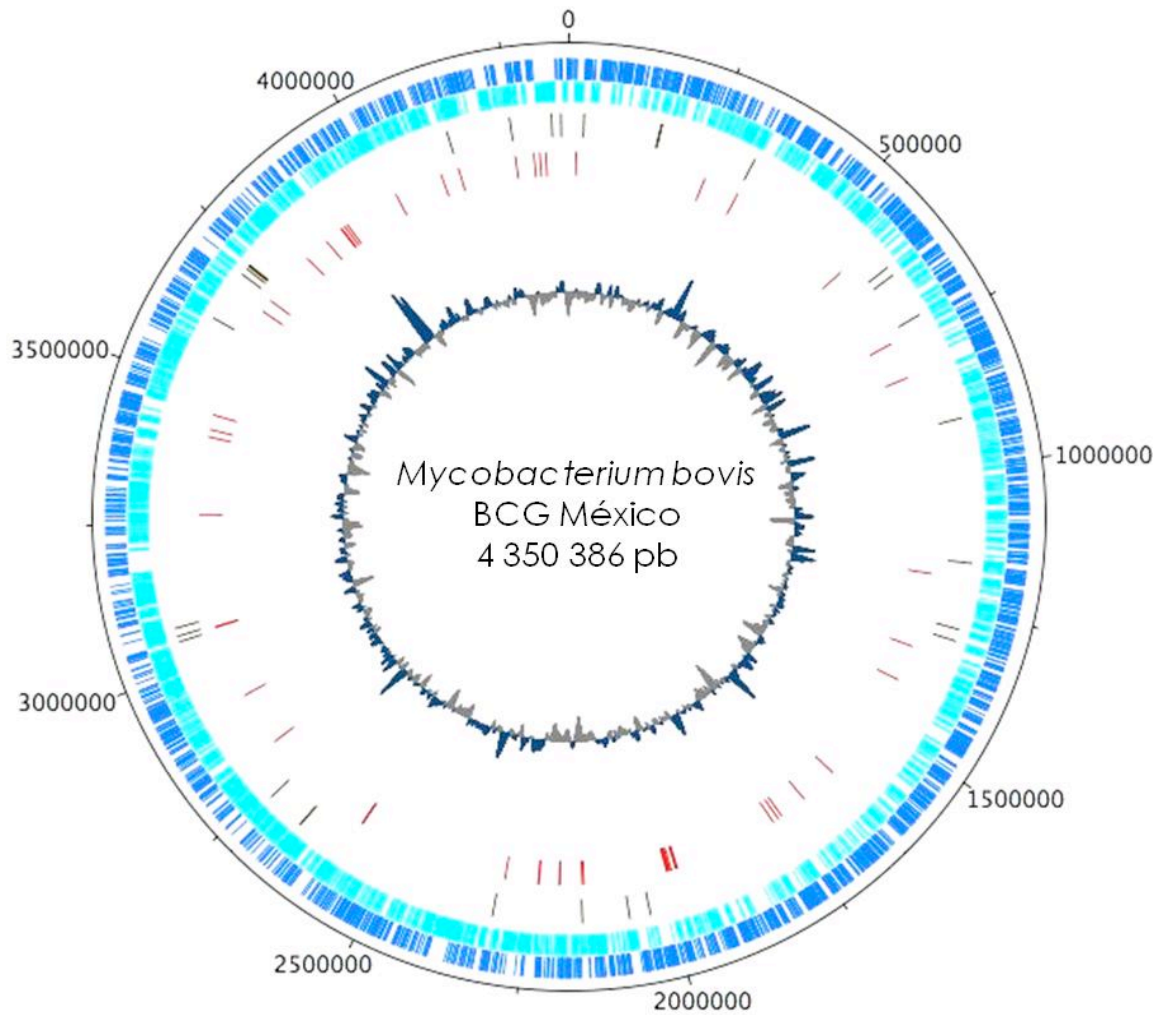


Figura 14. Mapa circular del cromosoma de la BCG México 1931. La escala está en pares de base y es mostrada en el círculo exterior. El círculo azul fuerte muestra los CDS codificados en la cadena "directa" y el azul claro los codificados en la cadena "complementaria". Los círculos verde y rojo representan pseudogenes y regiones de diferencia, respectivamente. Finalmente, los círculos del centro representan el contenido de G+C.

VI.IV Comparación del genoma de BCG México 1931 con BCG Pasteur y Tokio

La comparación del genoma completo de BCG México 1931 con los genomas de BCG Pasteur y BCG Tokio permitió la identificación de nuevas regiones de diferencia, así como la identificación de las regiones RD y DU descritas previamente.

En total, se identificaron tres deleciones en BCG México 1931 en comparación con BCG Pasteur. Estas regiones se denominaron RDMex01 (53 pb), RDMex02 (655 pb) y RDMex03 (2 847 pb), respectivamente (tabla 9). La RDMex01 comprende una región intergénica entre los genes *rpsN1* y *rpsH*, la RDMex02 trunca el gen *fadD23*, mientras que la RDMex03 provoca la pérdida de tres proteínas hipotéticas (*BCG3923*, *BCG3924* y *BCG3926*) y un regulador transcripcional (*whiB6*) [64].

Además de estas deleciones también se determinó la presencia de SNPs entre las BCGs estudiadas, que para fines de esta comparación nosotros definimos como polimorfismos de un solo nucleótido entre las secuencias analizadas. Así, se identificaron un total de 33 SNPs con respecto a BCG Pasteur y 77 con respecto a BCG Tokio [64]. Del total de SNPs, alrededor de 23 son consistentes con los reportados en dos estudios previos [42, 43].

Tabla 9. Regiones RD identificadas en BCG México 1931.

Secuencia pérdida	Localización ¹	Longitud (pb)	Genes afectados ¹
RDMex01 ²	844357...844410	53	Región intergénica entre <i>rpsN1</i> y <i>rpsH</i>
RDMex02 (Δ <i>fadD23</i>)	4272815...4273470	655	<i>fadD23</i>
RDMex03 (Δ Rv3860-64)	4308821...4311668	2847	<i>BCG3923</i> <i>BCG3924</i> <i>whiB6</i> <i>BCG3926</i>

¹ La localización y el nombre de los genes afectados corresponden al genoma de BCG Pasteur; ² Región ausente en BCG Danesa.

Del total de SNPs identificados, 37 corresponden a mutaciones no sinónimas (nsSNP) afectando regiones codificantes (tabla 10). De los cuales 4 son específicos de la BCG México 1931 (*BCGMEX_0506c*, *BCGMEX_1957*, *BCGMEX_2390*, *ponA2*). Este análisis permitió la identificación de nsSNP en los genes *mma3*, *sigK* y *BCG3734*; los cuales han sido descritos en estudios previos [28, 43].

En general, los genes afectados en mayor proporción por nsSNP codifican proteínas del metabolismo de lípidos (16%), del metabolismo y la respiración (16%), de la familia PE/PPE (16%) e hipotéticas (19%), ver figura 15A. Si comparamos estos datos con las proporciones de estas proteínas en el genoma (figura 15B) podemos ver que una gran proporción de nsSNPs esta concentrada en las proteínas del metabolismo de lípidos y en las proteínas de la familia PE/PPE.

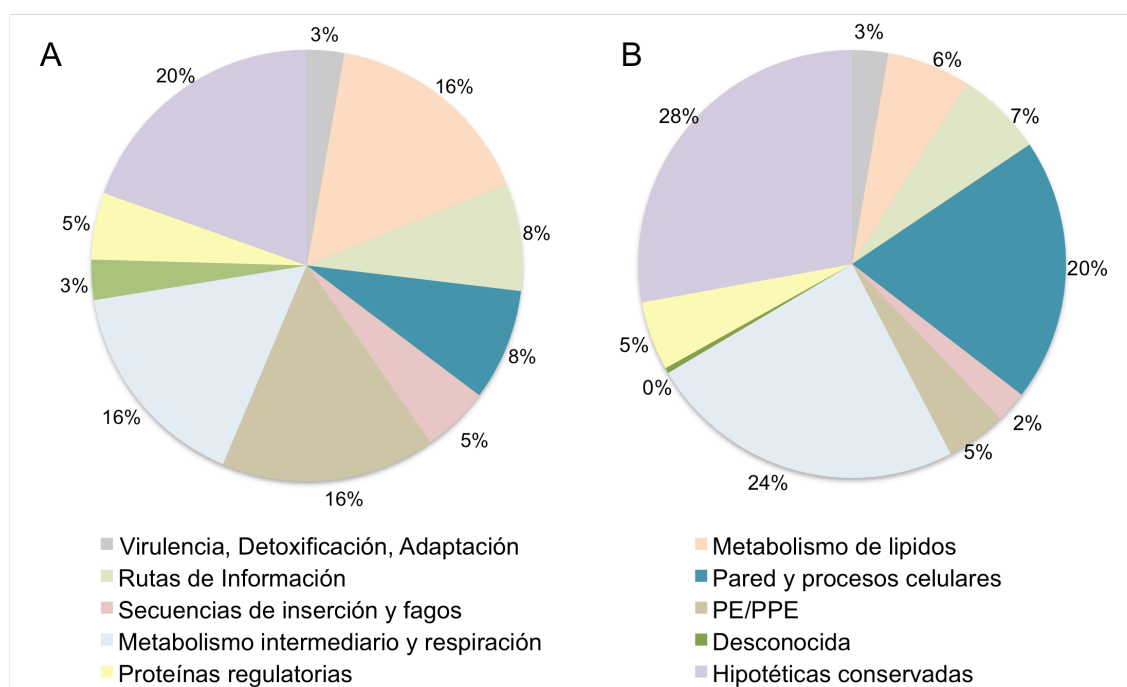


Figura 15. Categoría funcional de las proteínas afectadas por SNPs no sinónimos en BCG México 1931 (A), los porcentajes se calcularon tomando como 100% las proteínas con nsSNPs. En el panel B se muestra la categoría funcional de las proteínas del genoma completo, los porcentajes se calcularon tomando como 100% el número total de proteínas.

Tabla 10. SNPs no sinónimos identificados en BCG México 1931.

SNP	Posición ¹	Gen	Mutación	Cambio	En comparación con BCG
1	39256	BCGMEX_0037c	NS	prolina-leucina	Pasteur
2	89198	BCGMEX_0084	NS	cisteína-arginina	Tokio
3	338587	<i>PE_PGRS4</i>	NS	leucina-lisina	Tokio
4	535543	sigK	NS	isoleucina-metionina	Tokio
5	562401	<i>pcaA</i>	NS	treonina-prolina	Tokio
6	582595	<i>regX3</i>	NS	arginina-cisteína	Tokio
7	585522	<i>BCGMEX_0506c</i>	NS	serina-glicina	Tokio y Pasteur
8	644544	BCGMEX_0568	NS	treonina-isoleucina	Tokio
9	739067	mmaA3	NS	ác. aspártico-glicina	Tokio
10	1165878	<i>BCGMEX_1072c</i>	NS	ác. glutámico-glicina	Tokio
11	1294737	<i>narJ</i>	NS	prolina-leucina	Pasteur
12	1465719	<i>BCGMEX_1346c</i>	NS	serina-glicina	Tokio
13	1671425	BCGMEX_1520	inserción	----	Tokio
14	1908146	PPE22	NS	alanina-ác. aspártico	Pasteur
15	1987564	<i>BCGMEX_1787c</i>	NA	----	Pasteur
16	2171779	<i>BCGMEX_1957</i>	NS	prolina-alanina	Tokio y Pasteur
17	2261942	<i>pks12</i>	NS	lisina-ác. glutámico	Pasteur
18	2268073	<i>pks12</i>	NS	glicina-ác. aspártico	Pasteur
19	2599245	hrcA	NS	glutamina-arginina	Tokio
20	2625029	<i>BCGMEX_2390</i>	NS*	prolina-arginina	Tokio y Pasteur
21	2637892	<i>PE_PGRS41</i>	NS	glicina-cisteína	Tokio
22	2744951	<i>PE_PGRS43b</i>	inserción	----	Tokio
23	2839572	BCGMEX_2587c	NS	arginina-cisteína	Pasteur
24	3294834	ilvN	NS	treonina-isoleucina	Pasteur
25	3433045	PPE50	NS	lisina-ác. glutámico	Tokio
26	3433175	<i>PPE50</i>	NS	ác. glutámico-glicina	Tokio
27	3537078	<i>BCGMEX_3256</i>	NS	ác. aspártico-glicina	Tokio
28	3592845	<i>BCGMEX_3305c</i>	NS	isoleucina-treonina	Tokio
29	3674969	<i>sdhB</i>	NS	alanina-treonina	Tokio
30	3785919	<i>BCGMEX_3469</i>	NS	leucina-arginina	Tokio
31	4015927	hpt	NS	asparagina-aspártico	Tokio
32	4059882	BCGMEX_3724	NS*	serina-treonina	Pasteur
33	4060893	<i>acs</i>	inserción	----	Tokio
34	4069095	BCGMEX_3734	NS	prolina-leucina	Tokio
35	4075693	<i>ponA2</i>	NS	arginina-glicina	Tokio y Pasteur
36	4087843	<i>BCGMEX_3752</i>	NS	alanina-prolina	Tokio
37	4252183	<i>fadD23</i>	pseudogen	----	BCG Pasteur

¹ La localización y el nombre de los genes afectados corresponden al genoma de BCG México 1931; las letras en negro muestran los SNPs reportados previamente; NS= no sinónimo; *= región no anotada en BCG Pasteur y Tokio, pero sí en otras micobacterias.

VI.IV.I Identificación de nuevas regiones RD en otras BCGs

Como se describió previamente la BCG México 1931 presenta tres deleciones (RDMex01, RDMex02 y RDMex03), no descritas previamente, con respecto a BCG Pasteur. Con la finalidad de determinar si estas deleciones se encuentran únicamente en BCG México se diseñaron oligos específicos para el análisis de estas regiones RD, con los criterios y condiciones de amplificación descritos en la metodología.

Los resultados obtenidos se muestran en la tabla 11, en ésta se observa que la RDMex02 se encuentra ausente en BCG México y BCG Danesa, pero presente en todas las demás BCGs estudiadas. Mientras que las regiones RDMex02 y RDMex03 están ausentes únicamente en BCG México 1931, lo cual demuestra que estas deleciones son específicas de la BCG México y por lo tanto son un marcador molecular útil para su identificación.

Tabla 11. Análisis de las regiones RDMex01, RDMex02 y RDMex03 en diferentes BCGs.

BCG	México	Phipps	Tokio	Moreau	Connaught	Australia	Tice	Frappier	Danesa
RDMex01	-	+	+	+	+	+	+	+	-
RDMex02	-	+	+	+	+	+	+	+	+
RDMex03	-	+	+	+	+	+	+	+	+

El símbolo (+) indica la presencia de la región en el genoma de las BCGs, mientras el símbolo (-) indica la pérdida de la región.

VII. DISCUSIÓN

La única vacuna disponible actualmente para prevenir la tuberculosis es la BCG, la cual derivó en diferentes cepas. A través del tiempo se documentaron diferencias a nivel genómico, proteico e inmunológico entre las BCGs, lo que ha llevado a su caracterización. En este momento las BCGs mejor caracterizadas son las usadas para la producción de la vacuna en diferentes países alrededor del mundo, como la BCG Danesa, Tokio, Moreau y la BCG Pasteur que se utiliza como cepa de referencia.

A la fecha, ninguna de las BCGs empleadas para producir la vacuna en México ha sido caracterizada, lo cual daría información sobre la cepa. Además de que la caracterización global y en este caso la genómica resulta esencial para producir nuevamente la vacuna en el país o para el desarrollo de vacunas de segunda generación contra la tuberculosis [55].

Por lo que en el laboratorio de Inmunología Molecular Microbiana de la Facultad de Medicina se decidió caracterizar a nivel genómico las BCGs que se emplearon para la producción de la vacuna BCG en México. Inicialmente, se comenzó con la caracterización genética de tres BCGs donadas por el Instituto Nacional de Higiene y las cuales denominamos BCG México 1931, 1988 y 1997.

Los resultados obtenidos en este estudio mostraron que la cepa BCG México 1931 se ubica dentro del grupo DU2 tipo IV, mientras que las cepas BCG México 1988 y 1997 pertenecen al grupo DU2 tipo III. Estos resultados concuerdan con los registros históricos, donde se establece que la producción de la vacuna comenzó a partir de una cepa enviada directamente del Instituto Pasteur a México (BCG México 1931) y posteriormente se realizó la producción a partir de lotes semilla de la BCG Danesa que pertenece al grupo DU2 tipo III (BCG México 1988 y 1997).

Estos resultados confirman la identidad de la BCG México 1931, demostrando que es una cepa diferente a la BCG Danesa. Lo anterior es importante debido a que estas dos etapas de producción de la vacuna BCG en México originaron confusiones con respecto a la identidad de la cepa, de hecho en dos trabajos previos, se ubicaba a la BCG México dentro del mismo grupo que la BCG Danesa (DU2 tipo III) [30, 31].

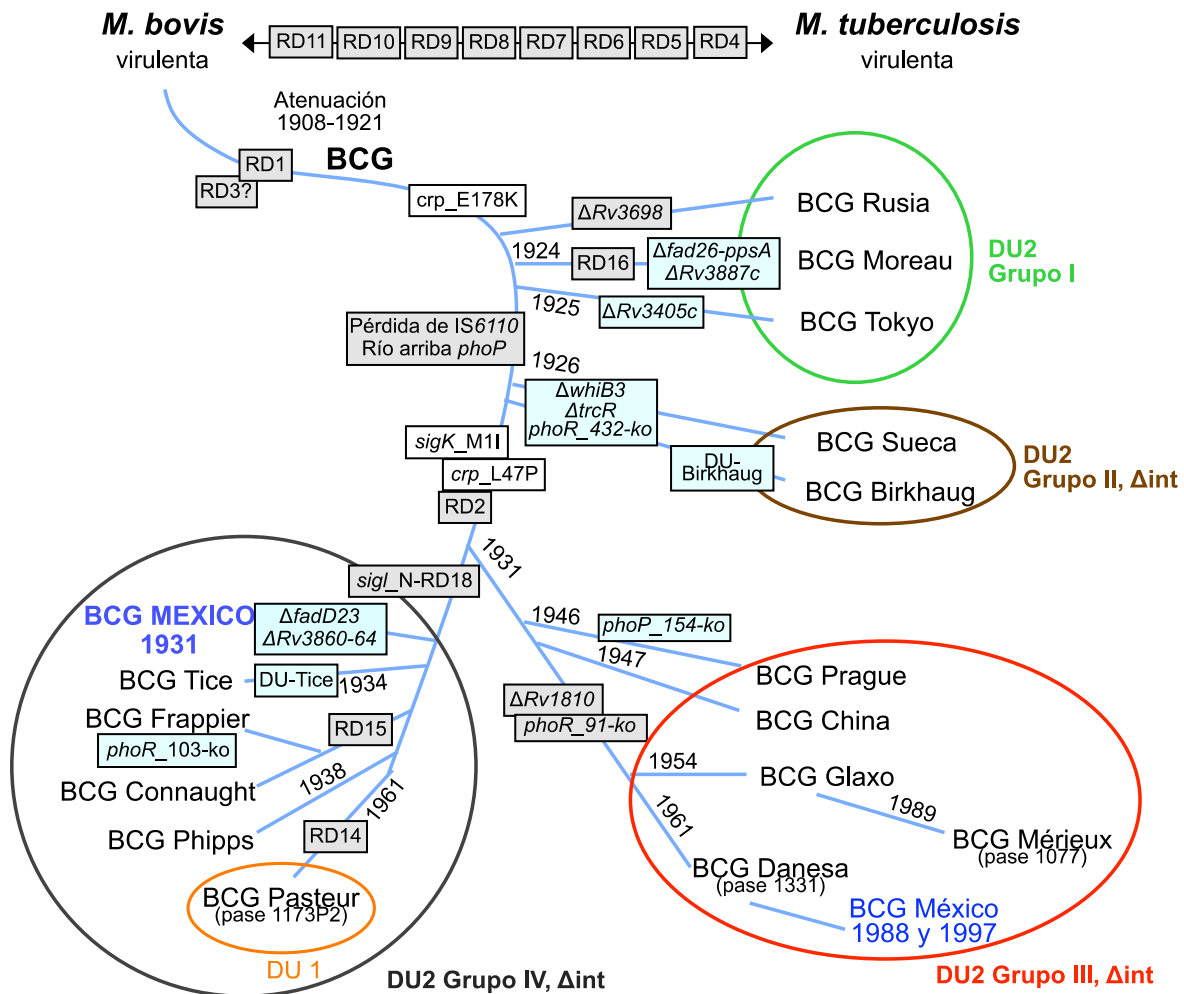


Figura 16. Genealogía de vacunas BCG, modificada de Brosch y cols [34]. La BCG México 1931 se incluyó en este estudio, además también se incorporó la información generada posteriormente a la publicación de la figura (recuadros verdes).

Con base en estos resultados se decidió continuar únicamente con la caracterización genética de la BCG México 1931, la cual corresponde a la cepa enviada del Instituto Pasteur a México. El perfil de regiones RD y DU permitió ubicarla dentro del grupo de DU2 tipo IV (figura 16); sin embargo, no permitió diferenciarla de la BCG Phipps y Tice que también pertenecen a este grupo. Motivo por el cual se decidió determinar su perfil genómico por electroforesis en gel por campos pulsados (PFGE).

Los resultados obtenidos por PFGE correlacionaron con los obtenidos previamente, además de que permitieron determinar que la BCG México presenta diferencias con respecto a BCG Phipps y Tice, y por lo tanto no corresponde a ninguna de las BCGs caracterizadas previamente.

Posteriormente, con la finalidad de tener la caracterización completa de la cepa a nivel genético se realizó la secuenciación del genoma completo de la BCG México 1931. El primer paso comprendió la secuenciación de la BCG por pirosecuenciación, el cual es un método de secuenciación masiva tipo “shotgun” que permite obtener alrededor de 70MB de secuencia por corrida. Nosotros realizamos dos corridas de secuenciación lo cual permitió obtener una cobertura de 36X el genoma. Lo cual es una cobertura alta si se considera que con este tipo de técnicas conforme se incrementa la redundancia la probabilidad de cerrar completamente el genoma disminuye y el costo se incrementa considerablemente [65].

En este sentido es importante mencionar que, dentro de las técnicas de secuenciación masiva, se decidió trabajar con GSFLX 454 de Roche debido a permitió obtener lecturas relativamente largas (250 pb) y a que estaba disponible en el Cinvestav de Irapuato. Sin embargo, actualmente están disponibles plataformas que garantizan lecturas de mayor tamaño como 454 Titanium FLX de Roche y otras que se pueden realizar a un menor costo como Genome Analyzer de Illumina.

Una vez realizada la secuenciación se procedió al ensamble de las secuencias. El primer ensamble lo realizó la Unidad de Secuenciación del Cinvestav Irapuato con el programa “GS de novo Assembler” intrínseco del secuenciador GSFLX 454 de Roche. Sin embargo, no se tomó éste como el ensamble final debido a que presentaba algunas desventajas, la principal es que no discriminaba entre secuencias repetidas. Esto provocaba que las secuencias repetidas se descartaran del ensamble y por lo tanto hubiera un mayor número de espacios en el genoma.

Debido a esto se decidió construir un ensamble nuevo con los programas Phred/Phrap/Consed, lo que permitió que las lecturas se distribuyeran en partes iguales en regiones con secuencias repetidas. El ensamble se construyó utilizando como

secuencia de referencia la BCG Pasteur, la cual era la única secuencia de BCG disponible cuando se realizó el ensamble. Para realizar el ensamble también se consideraron las lecturas generadas de la secuenciación “fos-end” de los fósmidos, las 250 clonas secuenciadas representaban una cobertura de 1X el genoma, para validar el ensamble.

Una vez ensamblado el genoma se procedió a identificar el número total de espacios en el genoma de la BCG México 1931 y a su posterior cierre. Debido a que se encontró un número relativamente pequeño de espacios (143) se decidió seguir como estrategia para el cierre del genoma el diseño de oligos específicos para cada uno de los espacios. Lo que resultó factible debido a que los espacios fueron pequeños (2-2500pb) y por lo tanto fue posible obtener productos de PCR que cubrieran estas regiones. Uno de los principales retos para el diseño de oligos fue que justamente la mayoría de ellos comprendían regiones repetidas y/o con alto contenido de G+C, en algunos casos estos obstáculos se resolvieron con el diseño de oligos más grande o un poco alejados de la región de interés. Es decir, por ejemplo si el espacio era de 500pb, el oligo se diseñó 200pb río arriba y 300pb río abajo. Además en todos los casos se tuvo que adicionar DMSO a las reacciones de PCR, lo que permitió que las cadenas DNA se relajaran a las T_M establecidas y por lo tanto que se obtuvieran los productos esperados.

También es necesario establecer, que en algunos casos la amplificaciones con los oligos no dieron los productos esperados, o las reacciones de secuencia siempre fueron de baja calidad. Para estos casos se diseñaron nuevos oligos.

Adicionalmente, al diseño de oligos también fue necesario secuenciar tres fósmidos. La secuenciación del fósmino 230 permitió cubrir 17 espacios que se encontraban concentrados en una sola región del genoma. Mientras que la secuenciación de los fósmidos 128 y 129 permitió definir el ensamble de la BCG México 1931, el primero llevó a la conclusión de que duplicación DU1 no se encuentra en el genoma de esta cepa y el segundo que sí presenta la duplicación DU2 y que ésta es muy similar a la descrita para BCG Pasteur.

Después de cerrar el genoma, se definió el tamaño de la BCG México 1931 (4 350 386pb) resultando 20Kb menor que el de BCG Pasteur (4 374 522pb) y BCG Tokio (4 371 711pb) [34, 42]. Esta diferencia se debe en el caso de BCG Pasteur a la ausencia de la

duplicación DU1, a deleciones específicas y a la presencia de la región RD14 en la BCG México 1931. Mientras que con BCG Tokio la diferencia es debida a la pérdida de las regiones RD2, N-RD18 y de una copia de la secuencia de inserción IS6110; así como a diferencias en duplicación DU2.

Una vez cerrado el genoma, se realizó la anotación completa con el programa Artemis, usando como apoyo el programa RAST Server y las bases de datos BCG List y BLAST. Durante la anotación se anotaron proteínas hipotéticas no anotadas previamente en BCG Pasteur y BCG Tokio (secuencia disponible cuando se realizó la anotación) pero sí en otras cepas de micobacterias. Al respecto, nosotros no encontramos un criterio de anotación para las proteínas hipotéticas (proteínas con función desconocida), de hecho observamos que en general estas proteínas hipotéticas están anotadas en algunos genomas pero en otros no a pesar de que el marco de lectura es evidente. Tal vez lo único que tienen en común es que tienen una secuencia corta (100- 400pb), aunque tampoco hay un límite establecido en este sentido.

La anotación completa del genoma de la BCG México 1931 mostró que ésta BCG contiene menos genes anotados (3 904) que BCG Pasteur (3 954) y Tokio (4 033) [34, 42]. Esta variación se debe al diferente número de regiones RD entre las cepas y a las diferencias en el criterio de anotación de las proteínas hipotéticas.

Finalmente, ya que se tuvo la secuencia y la anotación de la BCG México 1931 realizamos una comparación con las secuencias de las BCG Tokio y Pasteur. Esta comparación permitió identificar tres regiones ausentes en BCG México 1931 con respecto a estas BCGs, las cuales se denominaron RDMex01 (53pb), RDMex02 (655pb) y RDMex03 (2 847pb). Interesantemente, el análisis de estas regiones en nueve BCGs mostró que la RDMex02 y la RDMex03 están ausentes únicamente en la BCG México 1931, con lo cual pueden ser consideradas como marcadores moleculares de la cepa.

La RDMex01 comprende una deleción intergénica entre los genes *rpsN1* y *rpsH* que codifican dos subunidades de la proteína ribosomal 30S. El efecto biológico de esta mutación es desconocido.

La RDMex02 trunca el gen *fadD23*, el cual codifica una probable acyl-CoA ligasa implicada en la degradación de lípidos. Lynett y cols describieron que esta proteína está involucrada en la producción de sulfolípidos y que la disrupción del gen *fadD23* resulta en un incremento de la asociación entre la bacteria y el macrófago [66]. Lo que nos lleva a pensar que ésta delección puede explicar el fenotipo observado en un estudio realizado en el laboratorio por Molina donde se encontró que la BCG México presenta una mayor asociación a macrófagos (THP-1) en comparación con las BCGs Danesa, Moreau, Phipps y Tokio [67]. Esto resulta de interés debido a que esta mayor asociación a macrófagos puede tener como consecuencia una mayor internalización de la cepa.

Finalmente, la RDMex03 es la delección más grande encontrada en BCG México 1931, remueve 4 genes: 3 proteínas hipotéticas (*BCG3923*, *BCG3924* y *BCG3926*) y un probable regulador transcripcional (*whiB6*). Este regulador pertenece a una familia de proteínas denominada WhiB (1-7), las cuales se ha especulado que pueden jugar un papel clave en la persistencia o latencia de las micobacterias. De hecho existe evidencia de que sugiere que estas proteínas son factores transcripcionales involucrados en la regulación de diferentes procesos celulares como la división celular, patogénesis y resistencia a antibióticos. Incluso han sido propuestas como parte de un nuevo sistema redox en las micobacterias [28, 68]. La ausencia de este gen en la cepa puede ocasionar una respuesta diferencial de la BCG México 1931 al estrés oxidativo en comparación con otras BCGs, lo cual puede estar implicado en la sobrevida de la cepa dentro del macrófago.

Por lo que además de ser consideradas como marcadores moleculares de la cepa (figura 16), estas delecciones pueden explicar algunas de las diferencias fenotípicas observadas entre la BCG México 1931 y otras BCGs. Para poder establecer esto es necesario realizar ensayos que demuestren claramente cuál es el impacto funcional de estas delecciones.

La comparación entre las BCGs también evidenció la presencia de SNPs, 33 con respecto a BCG Pasteur y 77 con respecto a BCG Tokio. Del total de SNPs identificados 23 son consistentes con los reportados en dos estudios previos [42, 43]. En uno de estos trabajos, García Pelayo y cols construyeron una filogenia de BCGs basada en la presencia de SNPs [43]. Los SNPs encontrados en BCG México 1931 correlacionan con

esta filogenia en donde se ubica a la BCG México 1931 en el mismo grupo que BCG Tice, lo cual concuerda con el perfil de regiones RD y DU descritos previamente.

Dentro de los SNPs identificados, 37 correspondieron a mutaciones no sinónimas. De estas mutaciones 17 caen dentro de proteínas hipotéticas y resto en proteínas pertenecientes a diferentes procesos celulares. Cuando estas mutaciones caen dentro del sitio catalítico, en el caso de enzimas, o en sitios que puedan afectar el plegamiento o reconocimiento de las proteínas su función se puede ver comprometida.

Para algunos de los nsSNPs identificados ya se describieron implicaciones funcionales específicas. Así por ejemplo, el SNP en el gen *mmaA3* es responsable de la ausencia de ácidos metoximicólicos en las BCGs obtenidas después de 1927 del Instituto Pasteur. Lo anterior es consistente con los hallazgos de Hayashi y *cols*, quienes describieron que la BCG México 1931 no presenta ácidos metoximicólicos en su pared celular [30].

Con respecto al SNP en el gen *sigK*, éste es responsable de una baja expresión de los antígenos MPB70 y MPB83 en las BCGs tardías, incluyendo la BCG México 1931 [25]. Mientras que la mutación en el gen *BCG3734*, homólogo del regulador global CRP, es específica de las BCGs y es responsable del incremento de la unión de CRP a su DNA blanco [69].

La implicación funcional del resto de los nsSNPs identificados no ha sido descrita. Sin embargo, en algunos de éstos se originan cambios que involucran los aminoácidos prolina y cisteína (tabla 10), los cuales se sabe que se encuentran en giros beta y gamma de las proteínas o en puentes disulfuro, respectivamente; lo que puede originar que se presenten cambios en la estructura de estas proteínas. Entre los genes que presentaron estos cambios se encuentran *narJ*, *pcaA*, *regX3* y *PE_PGRS41*. Interesantemente, mutaciones en los genes *regX3* y *pcaA* han mostrado que éstos son importantes para la virulencia de *M. tuberculosis* [70-72].

Adicionalmente, se buscó si los nsSNPs identificados afectan la región catalítica o sitios de unión a proteínas en aquellas proteínas que ya han sido caracterizadas [73], en ninguno de los casos encontramos esta relación.

Además de los nsSNPs mencionados, se encontró un SNP específico de BCG México 1931 que causa un cambio de aminoácido en el gen *ponA2*. Mutaciones de este gen en *M. tuberculosis* han sido asociadas a una mayor sensibilidad al estrés por calor (24h a 45°C) y a la exposición al H₂O₂. Además, una cepa de *M. tuberculosis* con mutación en el gen *ponA2* presentó una menor sobrevivencia que la cepa silvestre [74].

La identificación de SNPs en la BCG México 1931 resulta de interés debido a que como se describió, algunos de ellos tienen implicaciones funcionales sobre la cepa, aunque en otros casos a pesar de que se tienen evidencias indirectas del posible papel de estos SNPs aún es necesario demostrar si tienen o no alguna implicación funcional.

Un grupo de proteínas en los que también se encontraron SNPs en la BCG México 1931 es la familia PE/PPE. Estas proteínas están constituidas de secuencias repetidas ricas en prolina y ácido glutámico, las cuales juegan un papel importante en la evasión de la respuesta inmune del hospedero y posiblemente en la variación antigénica que presentan las micobacterias. Nosotros identificamos SNPs en los genes PPE22 y PPE50, interesantemente en estudios previos se ha demostrado que la proteína PPE22 induce una respuesta de células B, mientras que la PPE50 es requerida para el crecimiento de las micobacterias *in vitro* [75, 76]. También se encontraron diferencias en el tamaño de las proteínas PE_PGRS54 y PPE_PGRS55 en BCG México 1931 con respecto a las encontradas en BCG Tokio, lo cual concuerda con un artículo publicado previamente en donde se describió que en la BCG Tokio estas proteínas tienen una secuencia más corta [42]. Las implicaciones funcionales de esta variación en el tamaño permanecen desconocidas.

Interesantemente, se identificaron seis SNPs en los genes *PE_PGRS4*, *PPE22*, *PE_PGRS41*, *PPE50* y *PE_PGRS43b* (tabla 10). Estos genes codifican proteínas de la familia PE/PPE, la cual está constituida por secuencias repetidas ricas en prolina y ácido glutámico. Además, se observó que los genes *PE_PGRS54* (6 285 pb) y *PPE_PGRS55* (5 433 pb) tienen una mayor longitud en BCG México 1931 en comparación con su secuencia homóloga en BCG Tokio (6,153 y 5,088 pb, respectivamente).

En resumen, la caracterización completa de la BCG México 1931 permitió la identificación de polimorfismos específicos de la cepa los cuales tienen un impacto en las características fenotípicas que presenta la BCG México. Entre ellas el perfil proteico e inmunológico, así como sus propiedades bioquímicas. A este respecto, recientemente se publicó un artículo en donde se describen las características bioquímicas de 14 BCGs, interesantemente la BCG México es la que presenta un perfil bioquímico más parecido al de *M. bovis* que cualquier otra BCG [77].

Lo descrito en este trabajo muestra que dentro de los genes que presentan más polimorfismos en las BCGs se encuentran aquellos que afectan genes de sistemas de dos componentes y de transporte (*esX*, *senX3-regX3*, *phoP-phoR*), proteínas reguladoras (familia *whiB*, *crp*, *trcR*) y del metabolismo de lípidos (*fadD*, *ppsA*). Algunos de los cuales representan ejemplos claros de la relación que existe entre las diferencias genéticas y fenotípicas observadas en las BCGs. Sin embargo, las implicaciones funcionales de la mayoría de estas diferencias genéticas son aún desconocidas, su estudio permitirá explicar los diferentes fenotipos descritos en las BCGs.

VIII. CONCLUSIONES

La determinación del perfil de regiones RD y DU de las BCGs utilizadas para la producción de la vacuna en México permitió identificar claramente dos períodos de producción. El primero en el cual se produjo la vacuna a partir de una BCG enviada del Instituto Pasteur (1949-1970), denominada BCG México. Y el segundo a partir de una BCG donada por el Instituto Danés (BCG Danesa, 1970-1998), donde las BCGs analizadas poseen un perfil idéntico al de ésta. Estos resultados y la posterior secuenciación del genoma de la BCG México permitieron ubicar a la BCG México dentro del grupo DU2-tipo IV de la genealogía de BCGs.

Además, el análisis del genoma de la BCG México 1931 permitió identificar regiones posiblemente implicadas en las características fenotípicas de la cepa, como la RDMex02, la RDMex03 y la presencia de SNPs. Algunos de los cuales podrían estar involucrados en la atenuación de la cepa.

IX. PERSPECTIVAS

La caracterización de la BCG México 1931 puede permitir considerar el uso de ésta para volver a producir la vacuna en el país. Además de que puede ser considerada para el desarrollo de nuevas vacunas contra la tuberculosis, ya que con su genoma descrito es factible realizar modificaciones genéticas dirigidas que permitan tener una BCG con mejores características que las BCGs usadas actualmente.

La perspectiva directa de este trabajo es realizar el estudio de la implicación funcional que tienen las diferencias genéticas encontradas en la BCG México 1931, específicamente en las regiones RD y SNPs identificados.

La secuenciación del genoma completo de la BCG México también puede llevar al estudio de las secuencias regulatorias, expresión de genes, y a realizar diversos análisis genómicos *in silico*.

X. REFERENCIAS

1. **Tuberculosis.** [<http://www.who.int/topics/tuberculosis/en/>]
 2. Cole ST, Eisenach KD, McMurray DN, WRJ J. **Tuberculosis and the Tubercle Bacillus.** Washington, USA: ASM Press, 2005.
 3. Riley LW. **Of mice, men, and elephants: Mycobacterium tuberculosis cell envelope lipids and pathogenesis.** J Clin Invest. 2006 Jun;116:1475-8.
 4. WHO. **Global tuberculosis control 2011.** WHO REPORT 2011.
 5. Báez AR. **Epidemiología de la tuberculosis en México.** Salud Pública de México. 2003;43.
 6. **Dirección General de Epidemiología.** [www.dgepi.salud.com.mx]
 7. WHO. **WHO REPORT 2010: Global Tuberculosis Control.** Switzerland, 2010.
 8. Nava-Aguilera E, Lopez-Vidal Y, Harris E *et al.* **Clustering of Mycobacterium tuberculosis cases in Acapulco: Spoligotyping and risk factors.** Clin Dev Immunol. 2011;2011:408375.
 9. Macías Parra M, Kumate Rodríguez J, Arredondo García JL *et al.* **Mycobacterium tuberculosis complex genotype diversity and drug resistance profiles in a pediatric population in Mexico.** Tuberculosis Research and Treatment. 2011;2011.
 10. WHO. **BCG Vaccine.** Weekly Epidemiological Record. 2004 23 January 2004.
 11. P. León A. **Tuberculosis y BCG. 50 años de Investigación en México.** México: Marquez Lopez Editores 1992.
 12. Fine P, Carneiro I, Milstein J, Clements C. **Issues relating to the use of BCG in immunization programmes.** Department of vaccines and biologicals WHO. 1999.
 13. Norma Oficial Mexicana: **NOM-036-SSA2-2002. Prevención y control de enfermedades. Aplicación de vacunas, toxoides, sueros, antitoxinas e inmunoglobulinas en el humano.** México: Diario Oficial de la Federación 2003.
 14. Hawgood BJ. **Doctor Albert Calmette 1863-1933: founder of antivenomous serotherapy and of antituberculous BCG vaccination.** Toxicon. 1999 Sep;37:1241-58.
 15. Oettinger T, Jorgensen M, Ladefoged A, Haslov K, Andersen P. **Development of the Mycobacterium bovis BCG vaccine: review of the historical and biochemical evidence for a genealogical tree.** Tuber Lung Dis. 1999;79:243-50.
-

16. Bonah C. **The 'experimental stable' of the BCG vaccine: safety, efficacy, proof, and standards, 1921-1933.** *Stud Hist Philos Biol Biomed Sci.* 2005 Dec;36:696-721.
 17. Behr MA, Wilson MA, Gill WP *et al.* **Comparative genomics of BCG vaccines by whole-genome DNA microarray.** *Science.* 1999 May 28;284:1520-3.
 18. Dubos RJ, Pierce CH. **Differential characteristics in vitro and in vivo of several substrains of BCG. I. Multiplication and survival in vitro.** *Am Rev Tuberc.* 1956 Nov;74:655-66.
 19. Aronson JD, Schneider P. **The problem of standardization of BCG vaccine.** *Am J Public Health Nations Health.* 1950 May;40:533-44.
 20. Trunz BB, Fine P, Dye C. **Effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness.** *Lancet.* 2006 Apr 8;367:1173-80.
 21. Andersen P, Doherty TM. **The success and failure of BCG - implications for a novel tuberculosis vaccine.** *Nat Rev Microbiol.* 2005 Aug;3:656-62.
 22. Fine PE. **BCG vaccination against tuberculosis and leprosy.** *Br Med Bull.* 1988 Jul;44:691-703.
 23. Colditz GA, Brewer TF, Berkey CS *et al.* **Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature.** *JAMA.* 1994 Mar 2;271:698-702.
 24. Van Deirse F, Senechal F. **BCG on Sauton's medium; effect of a long series of subcultures on the morphological and biological properties of BCG cultures.** *Bull World Health Organ.* 1950;2:347-54.
 25. Charlet D, Mostowy S, Alexander D, Sit L, Wiker HG, Behr MA. **Reduced expression of antigenic proteins MPB70 and MPB83 in Mycobacterium bovis BCG strains due to a start codon mutation in sigK.** *Mol Microbiol.* 2005 Jun;56:1302-13.
 26. Belley A, Alexander D, Di Pietrantonio T *et al.* **Impact of methoxymycolic acid production by Mycobacterium bovis BCG vaccines.** *Infect Immun.* 2004 May;72:2803-9.
 27. Behr MA, Schroeder BG, Brinkman JN, Slayden RA, Barry CE, 3rd. **A point mutation in the *mma3* gene is responsible for impaired methoxymycolic acid production in Mycobacterium bovis BCG strains obtained after 1927.** *J Bacteriol.* 2000 Jun;182:3394-9.
-

-
28. Liu J, Tran V, Leung AS, Alexander DC, Zhu B. **BCG vaccines: their mechanisms of attenuation and impact on safety and protective efficacy.** Hum Vaccin. 2009 Feb;5:70-8.
 29. Ritz N, Hanekom WA, Robins-Browne R, Britton WJ, Curtis N. **Influence of BCG vaccine strain on the immune response and protection against tuberculosis.** FEMS Microbiol Rev. 2008 Aug;32:821-41.
 30. Hayashi D, Takii T, Fujiwara N *et al.* **Comparable studies of immunostimulating activities in vitro among Mycobacterium bovis bacillus Calmette-Guerin (BCG) substrains.** FEMS Immunol Med Microbiol. 2009 Jul;56:116-28.
 31. Castillo-Rodal AI, Castanon-Arreola M, Hernandez-Pando R, Calva JJ, Sada-Diaz E, Lopez-Vidal Y. **Mycobacterium bovis BCG substrains confer different levels of protection against Mycobacterium tuberculosis infection in a BALB/c model of progressive pulmonary tuberculosis.** Infect Immun. 2006 Mar;74:1718-24.
 32. Varnerot A, Clement F, Gheorghiu M, Vincent-Levy-Frebault V. **Pulsed field gel electrophoresis of representatives of Mycobacterium tuberculosis and Mycobacterium bovis BCG strains.** FEMS Microbiol Lett. 1992 Nov 1;77:155-60.
 33. Frothingham R. **Differentiation of strains in Mycobacterium tuberculosis complex by DNA sequence polymorphisms, including rapid identification of M. bovis BCG.** J Clin Microbiol. 1995 Apr;33:840-4.
 34. Brosch R, Gordon SV, Garnier T *et al.* **Genome plasticity of BCG and impact on vaccine efficacy.** Proc Natl Acad Sci U S A. 2007 Mar 27;104:5596-601.
 35. Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK. **Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis.** J Bacteriol. 1996 Mar;178:1274-82.
 36. Brosch R, Gordon SV, Marmiesse M *et al.* **A new evolutionary scenario for the Mycobacterium tuberculosis complex.** Proc Natl Acad Sci U S A. 2002 Mar 19;99:3684-9.
 37. Bedwell J, Kairo SK, Behr MA, Bygraves JA. **Identification of substrains of BCG vaccine using multiplex PCR.** Vaccine. 2001 Feb 28;19:2146-51.
 38. Lewis KN, Liao R, Guinn KM *et al.* **Deletion of RD1 from Mycobacterium tuberculosis mimics bacille Calmette-Guerin attenuation.** J Infect Dis. 2003 Jan 1;187:117-23.
-

39. Pym AS, Brodin P, Brosch R, Huerre M, Cole ST. **Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti***. *Mol Microbiol*. 2002 Nov;46:709-17.
 40. Kozak R, Behr MA. **Divergence of immunologic and protective responses of different BCG strains in a murine model**. *Vaccine*. 2010 Feb 4;29:1519-26.
 41. Brosch R, Gordon SV, Buchrieser C, Pym AS, Garnier T, Cole ST. **Comparative genomics uncovers large tandem chromosomal duplications in *Mycobacterium bovis* BCG Pasteur**. *Yeast*. 2000 Jun 30;17:111-23.
 42. Seki M, Honda I, Fujita I, Yano I, Yamamoto S, Koyama A. **Whole genome sequence analysis of *Mycobacterium bovis* bacillus Calmette-Guerin (BCG) Tokyo 172: a comparative study of BCG vaccine substrains**. *Vaccine*. 2009 Mar 10;27:1710-6.
 43. Garcia Pelayo MC, Uplekar S, Keniry A *et al*. **A comprehensive survey of single nucleotide polymorphisms (SNPs) across *Mycobacterium bovis* strains and *M. bovis* BCG vaccine strains refines the genealogy and defines a minimal set of SNPs that separate virulent *M. bovis* strains and *M. bovis* BCG strains**. *Infect Immun*. 2009 May;77:2230-8.
 44. Rodriguez-Alvarez M, Palomec-Nava ID, Mendoza-Hernandez G, Lopez-Vidal Y. **The secretome of a recombinant BCG substrain reveals differences in hypothetical proteins**. *Vaccine*. 2010 May 21;28:3997-4001.
 45. Jungblut PR, Schaible UE, Mollenkopf HJ *et al*. **Comparative proteome analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG strains: towards functional genomics of microbial pathogens**. *Mol Microbiol*. 1999 Sep;33:1103-17.
 46. Rodriguez-Alvarez M, Mendoza-Hernandez G, Encarnacion S, Calva JJ, Lopez-Vidal Y. **Phenotypic differences between BCG vaccines at the proteome level**. *Tuberculosis (Edinb)*. 2009 Mar;89:126-35.
 47. Walker KB, Brennan MJ, Ho MM *et al*. **The second Geneva Consensus: Recommendations for novel live TB vaccines**. *Vaccine*. 2010 Mar 8;28:2259-70.
 48. Instituto-Nacional-de-Higiene. **Cien años de lucha por la salud**. Primera Edición ed. México: Grupo DESEA, S.A. de C.V., 1995.
 49. Garza-Ramos J. Las vacunas en México. En: Salud Sd, ed. **Simposio Avances en el uso de Vacunas 1885-1985**. México 1986:129-30.
-

-
50. Escobar A. Historia de las vacunas y las vacunaciones. En: salud Sd, ed. **Vacunas ciencia y salud**. México 1992:15-6, 195.
 51. **Instituto de Higiene**. Secretaría de Salud. México, 1931.
 52. P. León A. **S.S.A. Laboratorio de Producción de Vacuna B.C.G.** Salud Pública de México. 1963;V:1006-11.
 53. P. León A, Jiménez M. **La vacunación contra la tuberculosis en México**. Boletín de la oficina Sanitaria Panamericana. 1950;29:7-13.
 54. Fernández d Castro J. **Estado actual de la vacunación con BCG**. Gaceta Médica de México. 1976;III:271-9.
 55. Castanon-Arreola M, Lopez-Vidal Y. **A second-generation anti TB vaccine is long overdue**. Annals of clinical microbiology and antimicrobials. 2004 Jun 3;3:10.
 56. Sonnenberg JTBaMG. Isolation of Genomic DNA from Mycobacteria. En: Stoker TPaNG, ed. **Mycobacteria Protocols**. New Jersey: Human Press 1998.
 57. Philipp WJ, Gordon S, Talenti A, Cole ST. Pulsed Field Gel Electrophoresis for Mycobacteria. En: TpaNG S, ed. **Mycobacterial Protocols**. New Yersey: Human Press 1998.
 58. Phred, Phrap, Consed. [<http://www.phrap.org/phredphrapconsed.html>]
 59. **Oligo Calc: Oligonucleotide Properties Calculator**. [<http://www.basic.northwestern.edu/biotools/oligocalc.html>]
 60. **Rapid Annotation using Subsystem Technology**. [<http://rast.nmpdr.org/>]
 61. Rutherford K, Parkhill J, Crook J *et al*. **Artemis: sequence visualization and annotation**. Bioinformatics. 2000 Oct;16:944-5.
 62. **BCGList** [<http://genolist.pasteur.fr/BCGList/>]
 63. **Basic Local Alignment and Search Tool**. [<http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi>]
 64. Orduna P, Cevallos MA, de Leon SP *et al*. **Genomic and proteomic analyses of Mycobacterium bovis BCG Mexico 1931 reveal a diverse immunogenic repertoire against tuberculosis infection**. BMC Genomics. 2011;12:493.
 65. Roach J. **Random subcloning**. Genome Res. 1995;5:464-73.
 66. Lynett J, Stokes RW. **Selection of transposon mutants of Mycobacterium tuberculosis with increased macrophage infectivity identifies fadD23 to be involved in sulfolipid production and association with macrophages**. Microbiology. 2007 Sep;153:3133-40.
-

67. Molina-Olvera G. **Sobrevivencia y multiplicidad de distintas subcepas de Mycobacterium bovis BCG en macrófagos de la línea celular THP-I.** México: Universidad Nacional Autónoma de México, 2010.
 68. Alam MS, Garg SK, Agrawal P. **Studies on structural and functional divergence among seven WhiB proteins of Mycobacterium tuberculosis H37Rv.** FEBS J. 2009 Jan;276:76-93.
 69. Bai G, Gazdik MA, Schaak DD, McDonough KA. **The Mycobacterium bovis BCG cyclic AMP receptor-like protein is a functional DNA binding protein in vitro and in vivo, but its activity differs from that of its M. tuberculosis ortholog, Rv3676.** Infect Immun. 2007 Nov;75:5509-17.
 70. Parish T, Smith DA, Roberts G, Betts J, Stoker NG. **The senX3-regX3 two-component regulatory system of Mycobacterium tuberculosis is required for virulence.** Microbiology. 2003 Jun;149:1423-35.
 71. Glickman MS, Cox JS, Jacobs WR, Jr. **A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of Mycobacterium tuberculosis.** Mol Cell. 2000 Apr;5:717-27.
 72. Glover RT, Kriakov J, Garforth SJ, Baughn AD, Jacobs WR, Jr. **The two-component regulatory system senX3-regX3 regulates phosphate-dependent gene expression in Mycobacterium smegmatis.** J Bacteriol. 2007 Aug;189:5495-503.
 73. **Universal Protein Resource.** [<http://www.uniprot.org/>]
 74. Vandal OH, Roberts JA, Odaira T, Schnappinger D, Nathan CF, Ehrt S. **Acid-susceptible mutants of Mycobacterium tuberculosis share hypersusceptibility to cell wall and oxidative stress and to the host environment.** J Bacteriol. 2009 Jan;191:625-31.
 75. Sampson SL. **Mycobacterial PE/PPE proteins at the host-pathogen interface.** Clin Dev Immunol. 2011;2011:497203.
 76. Sassetti CM, Boyd DH, Rubin EJ. **Genes required for mycobacterial growth defined by high density mutagenesis.** Mol Microbiol. 2003 Apr;48:77-84.
 77. Hayashi D, Takii T, Mukai T *et al.* **Biochemical characteristics among Mycobacterium bovis BCG substrains.** FEMS Microbiol Lett. 2010 May;306:103-9.
-

XI. ANEXO A
MEDIOS Y TÉCNICAS

X.I Preparación de Medio Middlebrook 7H10

El medio 7H10 (19g) se resuspendió en 900mL de agua y se le adicionaron 5mL de glicerol. Posteriormente, el medio se disolvió y esterilizó por 10 minutos en autoclave bajo condiciones estándar de presión y temperatura. Al finalizar, se le adicionaron 100mL de medio de enriquecimiento Middlebrook OADC en condiciones de esterilidad.

X.II Preparación de Medio Sauton

El medio Sauton se preparó con 4.0 g de L-Asparagina, 2.0g de ácido cítrico, 0.5g de KH_2PO_4 , 0.5g de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.05g de citrato de amonio férrico, 60mL de glicerol y un 0.05% de Tween 80 por litro de agua. Antes llegar al volumen final, el medio se a un pH de 7.4 con NaOH 1M. El medio se esterilizó 15 minutos en autoclave bajo condiciones estándar.

X.III Librería genómica con fragmentos $\geq 40\text{kb}$

La librería de fósmidos se construyó con el kit “CopyControl™ Fosmid Library Production” (Epicentre Technologies, EEUU) siguiendo la metodología descrita a continuación.

Primero se realizó una electroforesis en gel de agarosa al 0.8% por 14 horas a 35 volts con la finalidad de verificar que el DNA genómico contaba con un tamaño de alrededor de 40kb. Posteriormente, se llevó a cabo una reacción de reparación con 14 μg de DNA y una solución con la enzima de reparación, la muestra se incubó 45 minutos a temperatura ambiente. Al terminó, la reacción se cargó en un gel de agarosa de bajo punto de fusión al 0.8%, el cual se resolvió por electroforesis en gel de agarosa a 35 volts y 14 horas. De este gel se cortó un fragmento de agarosa conteniendo el DNA de interés y se recuperó eluyendo a través de una columna de Dakron y precipitándolo con n-Butanol. Después, se realizó la reacción de ligación con 260ng del DNA reparado, 1 μg del vector CopyControl pCC1FOS y 5U de ligasa, la ligación se dejó toda la noche a temperatura

ambiente. Al día siguiente, el producto de la reacción de ligación se incubó por 3 horas a 30°C con los fagos incluidos en el kit “Max Plax Lambda Packaging Extracts”. Terminado el tiempo de incubación se adicionó solución amortiguadora PDB (Phage Dilution Buffer) a un volumen final de 1mL. Posteriormente, se realizaron diluciones seriadas 10^{-2} , 10^{-4} , 10^{-5} y 10^{-6} en PDB, 10µL de cada dilución se incubaron 20 min a 37°C con bacterias *E. coli* EPI300-T1®. Finalmente, las bacterias infectadas se inocularon en medio LB (Anexo X.IV) con cloramfenicol (15µg/mL) y las placas se incubaron toda la noche a 37°C.

La librería obtenida presentó 7, 000 UFC/mL, lo cual permitió tener representado al menos una vez (500 clonas) el genoma de la BCG México 1931.

X.IV Preparación de medio LB

El medio Luria se preparó con 4.0 g de L-Asparagina, 2.0g de ácido cítrico, 0.5g de KH_2PO_4 , 0.5g de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.05g de citrato de amonio férrico, 60mL de glicerol y un 0.05% de Tween 80 por litro de agua. Después de disolver todos los componentes, el medio se ajustó a un pH de 7.4 con NaOH 1M. El medio se esterilizó 15 minutos en autoclave bajo condiciones estándar.

X.V Purificación de DNA a partir de Geles de Agarosa

Para purificar el producto de PCR de interés se corrió un gel de agarosa (en este caso al 3% en TBE) y se cortó la bandas de interés. La banda se transfirió a un tubo eppendorf de 1.5mL, al tubo se le adicionó 1mL de resina y se colocó a baño María a 70°C hasta que la agarosa se fundió completamente. Después, la resina se pasó por una minicolumna Wizard, y posteriormente se realizó un lavado con 2mL de isopropanol al 80%. La muestra se centrifugó a 10 000 rpm por 2 minutos y la columna se transfirió a un tubo nuevo. Finalmente, el DNA se eluyó de la columna con agua caliente (60°C) y centrifugación por 20 segundos. El producto obtenido se almacenó a -20°C hasta su uso.

XII. ANEXO B
ARTÍCULO PUBLICADO

RESEARCH ARTICLE

Open Access

Genomic and proteomic analyses of *Mycobacterium bovis* BCG Mexico 1931 reveal a diverse immunogenic repertoire against tuberculosis infection

Patricia Orduña¹, Miguel A Cevallos², Samuel Ponce de León³, Adriana Arvizu¹, Ismael L Hernández-González², Guillermo Mendoza-Hernández⁴ and Yolanda López-Vidal^{1*}

Abstract

Background: Studies of *Mycobacterium bovis* BCG strains used in different countries and vaccination programs show clear variations in the genomes and immune protective properties of BCG strains. The aim of this study was to characterise the genomic and immune proteomic profile of the BCG 1931 strain used in Mexico.

Results: BCG Mexico 1931 has a circular chromosome of 4,350,386 bp with a G+C content and numbers of genes and pseudogenes similar to those of BCG Tokyo and BCG Pasteur. BCG Mexico 1931 lacks Region of Difference 1 (RD1), RD2 and N-RD18 and one copy of IS6110, indicating that BCG Mexico 1931 belongs to DU2 group IV within the BCG vaccine genealogy. In addition, this strain contains three new RDs, which are 53 (RDMex01), 655 (RDMex02) and 2,847 bp (REDMex03) long, and 55 single-nucleotide polymorphisms representing non-synonymous mutations compared to BCG Pasteur and BCG Tokyo. In a comparative proteomic analysis, the BCG Mexico 1931, Danish, Phipps and Tokyo strains showed 812, 794, 791 and 701 protein spots, respectively. The same analysis showed that BCG Mexico 1931 shares 62% of its protein spots with the BCG Danish strain, 61% with the BCG Phipps strain and only 48% with the BCG Tokyo strain. Thirty-nine reactive spots were detected in BCG Mexico 1931 using sera from subjects with active tuberculosis infections and positive tuberculin skin tests.

Conclusions: BCG Mexico 1931 has a smaller genome than the BCG Pasteur and BCG Tokyo strains. Two specific deletions in BCG Mexico 1931 are described (RDMex02 and RDMex03). The loss of RDMex02 (*fadD23*) is associated with enhanced macrophage binding and RDMex03 contains genes that may be involved in regulatory pathways. We also describe new antigenic proteins for the first time.

Background

Tuberculosis (TB) remains a major health problem worldwide; the World Health Organisation (WHO) estimates that there were 9.4 million new cases and 1.7 million deaths from TB in 2009 [1]. *Bacillus Calmette-Guérin* (BCG) is currently the only available vaccine against tuberculosis. This vaccine protects against the most severe forms of the disease, millitary and meningeal tuberculosis; however, it is highly variable in its ability

to protect against pulmonary tuberculosis (0-80%). There are several reasons for this variability, including differences between BCG substrains, exposure to non-tuberculous mycobacteria (NTMs), the nutritional or genetic background of the population, differences in trial methods and variations between different clinical *Mycobacterium tuberculosis* strains [2-6].

Use of BCG in the early 1920s proved effective in protecting against TB, leading to distribution of the vaccine in many countries. This distribution process and subsequent preservation resulted in the generation of numerous BCG substrains with different morphological, biochemical and immunological features [7,8]. Several studies on BCG substrains have demonstrated changes

* Correspondence: lvidal@unam.mx

¹Programa de Inmunología Molecular Microbiana, Departamento de Microbiología y Parasitología, Facultad de Medicina, Universidad Nacional Autónoma de México, DF, México

Full list of author information is available at the end of the article

at the genetic level, and comparative analyses of *M. tuberculosis*, *M. bovis* and *M. bovis* BCG have identified region of difference (RD) and tandem duplication (DU) markers in these strains [9-12].

Regions of difference are DNA regions that are deleted in the *M. bovis* and *M. bovis* BCG genomes compared to *M. tuberculosis*. The RD1 region is involved in BCG attenuation [7,13]. It has been shown that deletion of this region in *M. tuberculosis* H37Rv leads to attenuation of the strain [14]; however, complementation of BCG Pasteur with RD1 does not fully restore virulence to wild-type levels [15]. BCG strains can be sub-classified according to the presence or absence of RD2 in early and late strains, respectively. Recently, Kozak *et al.* reported that BCG Pasteur, a strain that lacks RD2, exhibits decreased immunogenicity compared to BCG Russia, a strain that has retained RD2 [16]. Importantly, these two strains show no difference in their level of protection against pulmonary tuberculosis. Additionally, Castillo-Rodal *et al.* have shown that the RDs described to date do not correlate with the protective efficacy of BCG substrains in a murine model [17]. The differences observed among BCG strains suggest that additional attenuating mutations may be involved in the attenuation of individual BCG strains.

Analysis of the BCG Pasteur 1173P2 genome sequence has made it possible to construct a detailed genealogy of BCG vaccines. BCG substrains are classified into four groups (I-IV) based on RD and DU2 markers [9]. Furthermore, single-nucleotide polymorphisms (SNPs) that are unique to particular BCG substrains or shared among substrains have been identified. Some of these SNPs have functional implications for the affected genes. For example, a SNP in *mma3* (*BCG0692c*) is responsible for the lack of methoxymycolate production in late BCG substrains [18].

The evidence presented above supports further characterisation of BCG substrains to improve our understanding of the mechanisms and impact of attenuation to rational design of new vaccines and therapeutics for tuberculosis [2,19].

Even though it was one of the most widely used substrains for vaccination in Mexico, BCG Mexico 1931 has not been included in any previous comparative proteomic or genomic study of BCG strains. Characterisation of BCG Mexico 1931 will permit again its use for BCG vaccine production in Mexico. This BCG strain will be used to develop a new recombinant BCG vaccine. Recently, Hayashi *et al.* described the biochemical characteristics of 14 BCG strains (including a BCG Mexico substrain), as well as *M. bovis*, *M. tuberculosis*, *M. avium* and *M. smegmatis* strains. Interestingly, BCG Mexico presented a biochemical profile more similar to that of *M. bovis* than any other BCG strain [20].

Historical records show that the Pasteur Institute sent several shipments of BCG strains to Mexico between 1926 and 1927 (Pasteur Institute records, personal communication). In 1928, small-scale production of BCG vaccine began in Mexico. In 1949, a BCG vaccine production laboratory was opened, and the vaccine was distributed throughout Mexico and Latin America [21-23]. Since 1931, the BCG Mexico substrain has been maintained by the Laboratorios de Biológicos y Reactivos de Mexico, a state-owned company that produces biological agents in Mexico. The BCG Mexico substrain was used as the vaccine seed for many years [24]. In 1970, this strain was replaced with the BCG Danish 1331 strain for vaccine production [25]. In 1998, BCG vaccine production ended in Mexico; since then, the country has depended on imported vaccine. These changes in vaccine production have caused confusion regarding the identity of BCG Mexico. For this reason, we characterised three representative strains used for BCG vaccine production in Mexico, which are designated BCG Mexico 1931, 1988 and 1997 according to the production period in which they were used. In this report, we describe the genomic and proteomic features of BCG Mexico 1931.

Results and Discussion

RD and DU Profile of BCG Mexico strains

Our RD and DU profile analysis of BCG Mexico 1931 demonstrated the presence of the RD8, RD14, RD16 and RD Danish/Glaxo regions and the absence of the RD1, RD2 and N-RD18 regions, as well as a single copy of the insertion sequence IS6110. These properties are similar to those observed for BCG Phipps and BCG Tice (Table 1). In contrast, BCG Mexico 1988 and BCG Mexico 1997 exhibited identical RD and DU profiles, with the RD1, RD2 and RD Danish/Glaxo regions and one copy of IS6110 missing (Table 1). This profile is identical to that of BCG Danish. The absence of the RD Danish/Glaxo region, which is specific to BCG Danish, in BCG Mexico 1988 and 1997 confirms this result and is consistent with historical records indicating that BCG vaccine production in Mexico utilised the BCG Danish 1331 strain beginning in 1970.

The amplification pattern of DU regions in BCG Mexico 1931 indicated duplication of DU2-IV, in contrast to those of BCG Mexico 1988 and BCG Mexico 1997, which showed duplication of DU2-III (Table 1). These differences in RDs and DUs confirm that BCG Mexico 1931 is a different strain from BCG Mexico 1988 and 1997, which are related to BCG Danish.

The above results and subsequent sequencing of the BCG Mexico 1931 genome place this strain in DU2 group IV within the genealogy of BCG strains (Figure 1). These results differ from findings of previous studies

Table 1 Region of difference (RD) profiles in BCG Mexico strains

BCG Substrains	RD1	RD2	RD8	RD14	RD16	RD18	RDD/G	IS6110	DU2
Mexico seed lot 1931	-	-	+	+	+	-	+	+	IV
Phipps	-	-	+	+	+	-	+	+	IV
Tice	-	-	+	+	+	-	+	+	IV
Mexico seed lot 1988	-	-	+	+	+	+	-	+	III
Mexico vaccine 1997	-	-	+	+	+	+	-	+	III
Danish 1331	-	-	+	+	+	+	-	+	III

The symbol (-) indicates the loss of a region in the genome of a given strain, whereas the symbol (+) indicates the presence of the region. RD = region of difference; IS6110 = insertion sequence IS6110 (presence of a single copy); DU = tandem duplication.

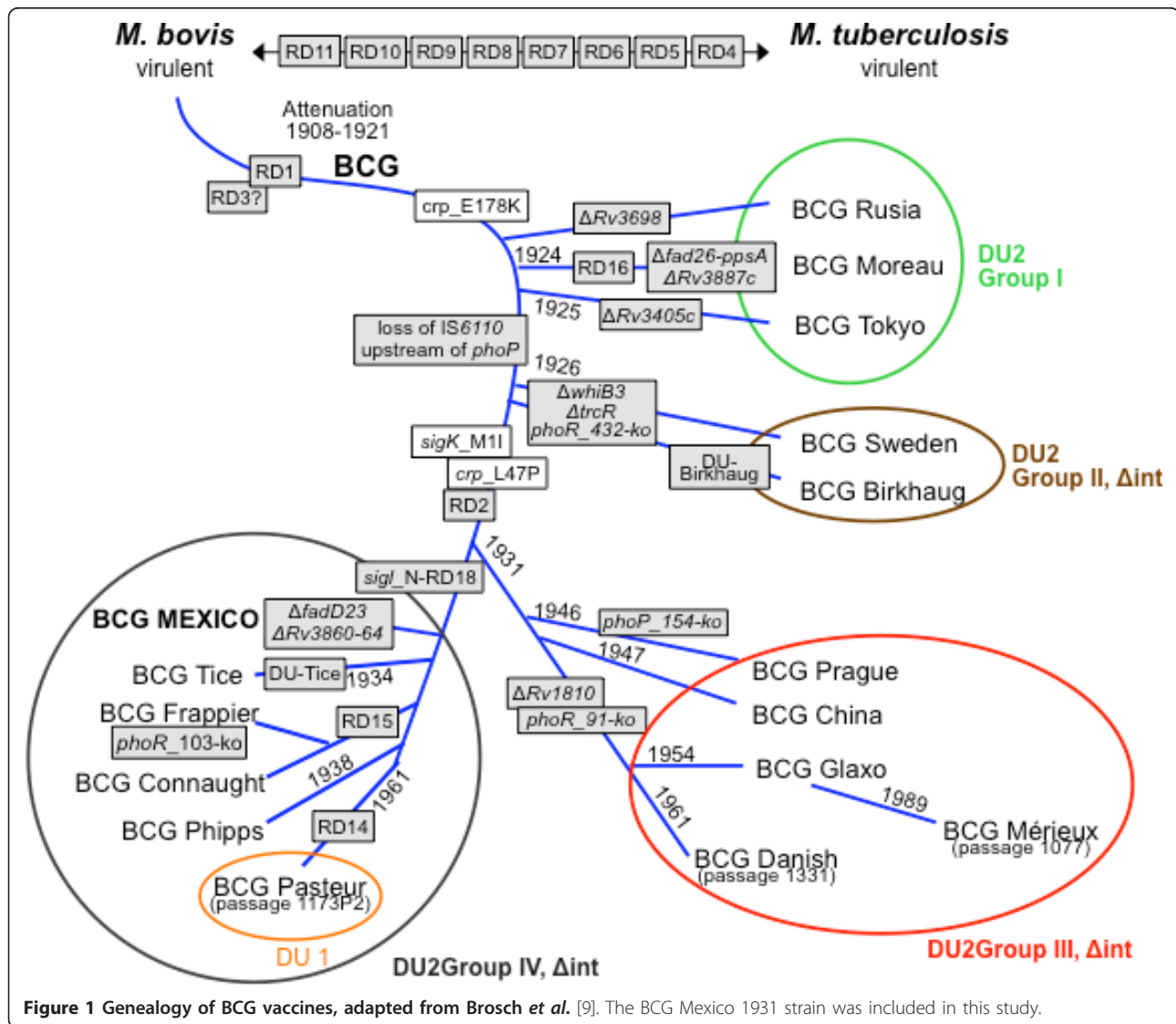


Figure 1 Genealogy of BCG vaccines, adapted from Brosch et al. [9]. The BCG Mexico 1931 strain was included in this study.

in which BCG Mexico was placed in DU2 group III as a strain derived from BCG Danish, which is an erroneous result that can be attributed to the different BCG vaccine production periods in Mexico [17,26].

Genome Sequence of BCG Mexico 1931

Mycobacterium bovis BCG Mexico 1931 has a circular chromosome of 4,350,386 bp with an overall G+C content of 65.7% [GenBank: CP002095]. The genome contains 3,904 genes that encode proteins (CDS), three genes that encode rRNAs and 45 genes that encode tRNAs. Additionally, 29 possible pseudogenes have been identified (Figure 2). The BCG Mexico 1931 genome is 20 Kb smaller than those of BCG Pasteur 1173P2

(4,374,522 bp) and BCG Tokyo 172 (4,371,711 bp). The differences between BCG Mexico 1931 and BCG Pasteur are due to lack of DU1, the presence of specific deletions and the presence of RD14 in BCG Mexico 1931. With respect to BCG Tokyo 172, the difference in genome size can be explained by the loss of RD2, N-RD18 and one copy of insertion sequence IS6110 as well as by differences in the size of DU2 (Figure 1).

The genome of BCG Mexico 1931 contains fewer genes (3,904) than those of BCG Pasteur (3,954) and BCG Tokyo (4,033) [9,27]. This variation is due to the presence of different RDs in each strain and to differences in the criteria used for annotation of hypothetical proteins not previously described in other BCG strains

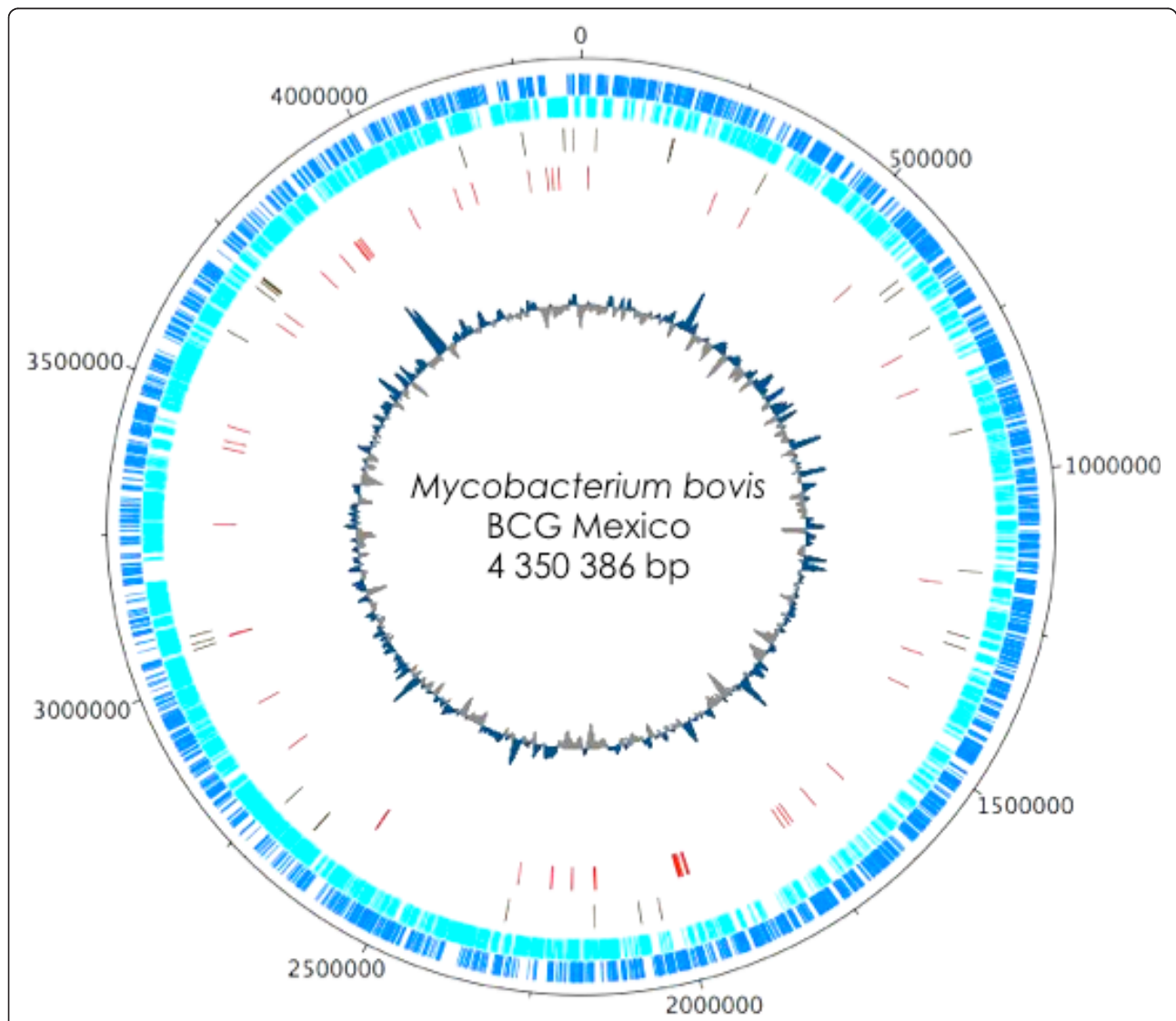


Figure 2 Circular map of the *M. bovis* BCG Mexico 1931 chromosome. The scale is in bases and is shown in the black outer circle. The dark blue circle shows forward-strand CDS, and the light blue circle shows reverse-strand CDS. The next two circles moving inward show pseudogenes (green) and difference regions (red). The innermost circle represents the G+C content.

but annotated in *M. tuberculosis* strains (GenBank database).

A comparison of the BCG Mexico 1931 genome sequence with those of BCG Pasteur and BCG Tokyo revealed 95 SNPs and three new RD regions, which were designated RDMex01 (53 bp), RDMex02 (655 bp) and RDMex03 (2,847 bp). These regions were deleted in the BCG Mexico 1931 sequence based on alignment with the BCG Pasteur and BCG Tokyo sequences. The deletions were confirmed by alignment with the *M. tuberculosis* and *M. bovis* sequences. These new RD regions led to the loss of three genes [*BCG3923*, *BCG3924* and *BCG3925c* (*whiB6*)] and the partial deletion of two genes [*BCG3926* and *BCG3889* (*fadD23*)] in BCG Mexico 1931 (Table 2). Interestingly, a PCR screen of these regions in nine BCG strains (Birkhaug, Connaught, Danish, Frappier, Moreau, Phipps, Tice, Tokyo and Sweden) showed that the RDMex02 and RDMex03 regions have been lost only in BCG Mexico 1931 and can therefore be used as molecular markers for this strain.

RDMex01 is an intergenic deletion located between the *BCG0767* (*rpsN1*) and *BCG0768* (*rpsH*) genes, which encode two subunits of the 30S ribosomal protein. The biological effect of this deletion is unknown.

RDMex02 is associated with deletion of 218 aa from *BCG3889* (*fadD23*), affecting a conserved region of the protein that includes two transmembrane domains. This gene encodes a probable fatty-acyl CoA ligase involved in lipid degradation. Lynett *et al.* have reported that this protein is involved in sulpholipid production and that disruption of the gene results in increased association between bacteria and macrophages [28]. Molina *et al.* found that BCG Mexico 1931 associates more strongly with macrophages (THP-1) compared to BCG Danish, BCG Moreau, BCG Phipps and BCG Tokyo172 [29].

Finally, RDMex03 was the largest deletion found in the BCG Mexico 1931 genome. It affected four genes: three genes encoding hypothetical proteins (*BCG3923*, *BCG3924* and *BCG3926*) and another gene encoding a putative transcriptional regulator [*BCG3925c* (*whiB6*)] belonging to the WhiB protein family (1-7). This family has been proposed to form part of a new redox system

in *M. tuberculosis* [30]. Interestingly, this deletion is situated in the extended RD1 region.

The new RDs described in BCG Mexico 1931 may contribute to understanding of the phenotypic differences between BCG Mexico 1931 and other BCG strains.

Our SNP analysis indicated the presence of 33 SNPs in BCG Mexico 1931 compared to BCG Pasteur and 77 SNPs in BCG Mexico 1931 compared to BCG Tokyo. Among these SNPs, at least 23 have been reported in two previous studies [27,31]. Additionally, in agreement with the SNP-based phylogeny constructed by García Pelayo *et al.*, BCG Mexico 1931 was grouped with BCG Tice in our analysis [31].

We found a total of 37 SNPs representing nonsynonymous mutations (nsSNPs), leading to amino acid substitutions (Tables 3 and 4). Four SNPs in this group were specific to BCG Mexico 1931 [*BCGMEX_0506c*, *BCGMEX_1957*, *BCGMEX_2390* and *BCG3741* (*ponA2*)]. On the basis of functional classes, the genes containing nsSNPs encoded hypothetical proteins (19%), proteins involved in intermediary metabolism and respiration (16%), proteins related to lipid metabolism (16%) and PE/PPE family proteins (16%) (Figure 3).

We found SNPs within *BCG0510c* (*pcaA*), *BCG0532* (*regX3*), *BCG0692c* (*mma3*), *BCG0484c* (*sigK*) and *BCG3734* (Table 3). The SNPs in the last three genes have been described in previous studies [7,31]. The SNP found in *BCG0692c* (*mma3*) causes an amino acid change with a concomitant loss of methoxymycolates in BCG strains obtained from the Pasteur Institute after 1927 [18]. This result is consistent with the findings of Hayashi *et al.*, who described the absence of these acids in BCG Mexico 1931 [26]. An SNP in the start codon of *BCG0484c* (*sigK*) is responsible for low expression of MPB70 and MPB83 in late BCG strains, including BCG Mexico 1931 [32]. Moreover, mutations in *BCG3734*, a CRP homologue global regulator, have been described as specific to BCG and are responsible for increased binding of CRP to its target DNA [33]. Mutations in *Rv0491* (*regX3*) and *Rv0470c* (*pcaA*) have been implicated in the virulence of *M. tuberculosis*. The *pcaA* gene encodes a mycolic acid cyclopropane synthetase and is

Table 2 New RD regions identified in BCG Mexico 1931

Deleted sequence	Location ¹	Length (bp)	Affected genes ¹
RDMex01 ²	844357...844410	53	Intergenic region between <i>rpsN1</i> and <i>rpsH</i>
RDMex02 (Δ <i>fad23</i>)	4272815...4273470	655	<i>fadD23</i>
RDMex03 (ΔRv3860-64)	4308821...4311668	2847	<i>BCG3923</i> <i>BCG3924</i> <i>whiB6</i> <i>BCG3926</i>

¹Locations and names of affected genes are given according to the BCG Pasteur 1173P2 genome; ²also deleted in BCG Danish 1331.

Table 3 Non-synonymous SNPs identified in BCG Mexico compared with BCG Pasteur 1173P2 and BCG Tokyo 172.

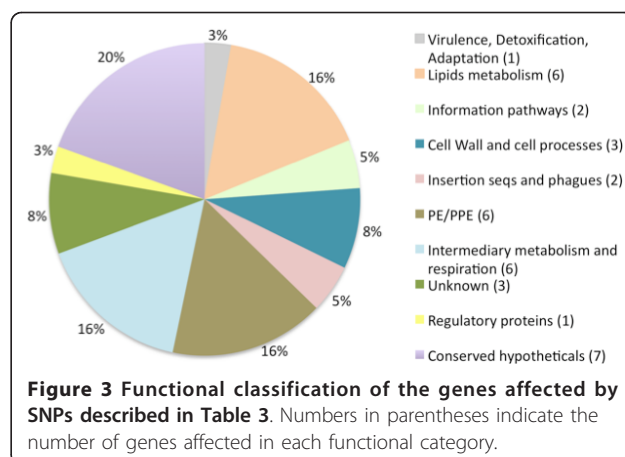
SNP	Location ¹	Gene	Change	In comparison with BCG
1	39256	BCGMEX_0037c	NS	Pasteur
2	89198	BCGMEX_0084	NS	Tokyo
3	338587	PE_PGRS4	NS	Tokyo
4	535543	sigK	NS	Tokyo
5	562401	pcaA	NS	Tokyo
6	582595	regX3	NS	Tokyo
7	585522	BCGMEX_0506c	NS	Tokyo and Pasteur
8	644544	BCGMEX_0568	NS	Tokyo
9	739067	mmaA3	NS	Tokyo
10	1165878	BCGMEX_1072c	NS	Tokyo
11	1294737	narJ	NS	Pasteur
12	1465719	BCGMEX_1346c	NS	Tokyo
13	1908146	PPE22	NS	Pasteur
14	1987564	BCGMEX_1787c	NA	Pasteur
15	2171779	BCGMEX_1957	NS	Tokyo and Pasteur
16	2261942	pkS12	NS	Pasteur
17	2268073	pkS12	NS	Pasteur
18	2599245	hrcA	NS	Tokyo
19	2625029	BCGMEX_2390	NA	Tokyo and Pasteur
20	2637892	PE_PGRS41	NS	Tokyo
21	2839572	BCGMEX_2587c	NS	Pasteur
22	3294834	ilvN	NS	Pasteur
23	3433045	PPE50	NS	Tokyo
24	3433175	PPE50	NS	Tokyo
25	3537078	BCGMEX_3256	NS	Tokyo
26	3592845	BCGMEX_3305c	NS	Tokyo
27	3674969	sdhB	NS	Tokyo
28	3785919	BCGMEX_3469	NS	Tokyo
29	4015927	hpt	NS	Tokyo
30	4059882	BCGMEX_3724	NA	Pasteur
31	4069095	BCGMEX_3734	NS	Tokyo
32	4075693	ponA2	NS	Tokyo and Pasteur
33	4087843	BCGMEX_3752	NS	Tokyo

¹Locations and names of affected genes follow BCG Mexico 1931; bold text indicates previously reported SNPs; NS = non-synonymous; NA = not annotated

Table 4 Insertions and pseudogenes identified in BCG Mexico 1931 compared with BCG Pasteur 1173P2 and BCG Tokyo 172

SNP	Location ¹	Gene	Change	In comparison with BCG
1	1671425	BCGMEX_1520	Insertion	Tokyo
2	2744951	PE_PGRS43b	Insertion	Tokyo
3	4060893	acs	Insertion	Tokyo
4	4252183	fadD23	Pseudogene	Pasteur

¹Locations and names of affected genes follow BCG Mexico 1931; bold text indicates previously reported SNPs.



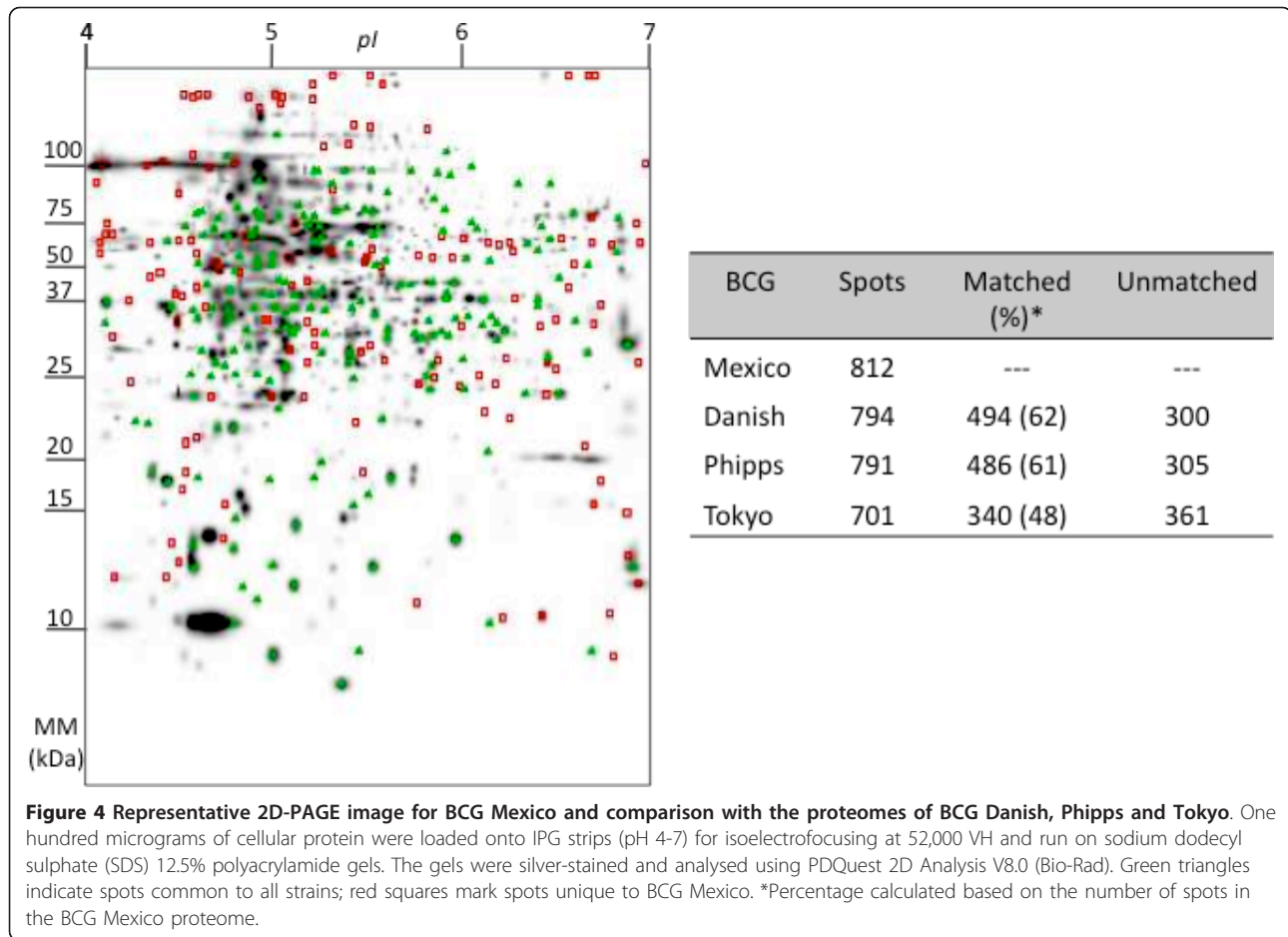
important for growth, persistence in macrophages and proinflammatory activity [34,35]. Additionally, *regX3* is part of a two-component system regulated by Pi (SenX3-RegX3) that is involved in the virulence of *M. tuberculosis* [36,37].

Interestingly, a specific nsSNP from BCG Mexico 1931 causes an amino acid change in *BCG3741* (*ponA2*). Mutations in this gene have been associated with increased sensitivity to heat shock (24 h at 45°C) and exposure to H₂O₂ compared to wild-type *M. tuberculosis*. Additionally, a *ponA2* mutant was found to exhibit lower survival in mice compared to wild-type *M. tuberculosis* [38].

We also identified six SNPs in *PE_PGRS4*, *PPE22*, *PE_PGRS41*, *PPE50* and *PE_PGRS43b* (Tables 3 and 4). These genes encode PE/PPE family proteins, which may play a role in the evasion of host immune responses, possibly via antigenic variation of mycobacteria [39]. In previous studies, it has been shown that the PPE22 protein elicits B cell responses, while PPE50 is required for mycobacterial growth *in vitro* [39,40]. Furthermore, we determined that *PE_PGRS54* (6,285 bp) and *PPE_PGRS55* (5,433 bp) correspond to a longer product in BCG Mexico compared with homologous sequences only for BCG Tokyo (6,153 and 5,088 bp, respectively). These results are consistent with data previously described [27]. Importantly, the functional implications of these size variations remain unknown.

Comparison of BCG proteomes

The protein contents of the cell fractions from four BCG substrains were analysed by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) using bacteria in mid-logarithmic phase. A total of 812, 794, 791 and 701 spots were visualised for BCG Mexico 1931, BCG Danish, BCG Phipps and BCG Tokyo, respectively

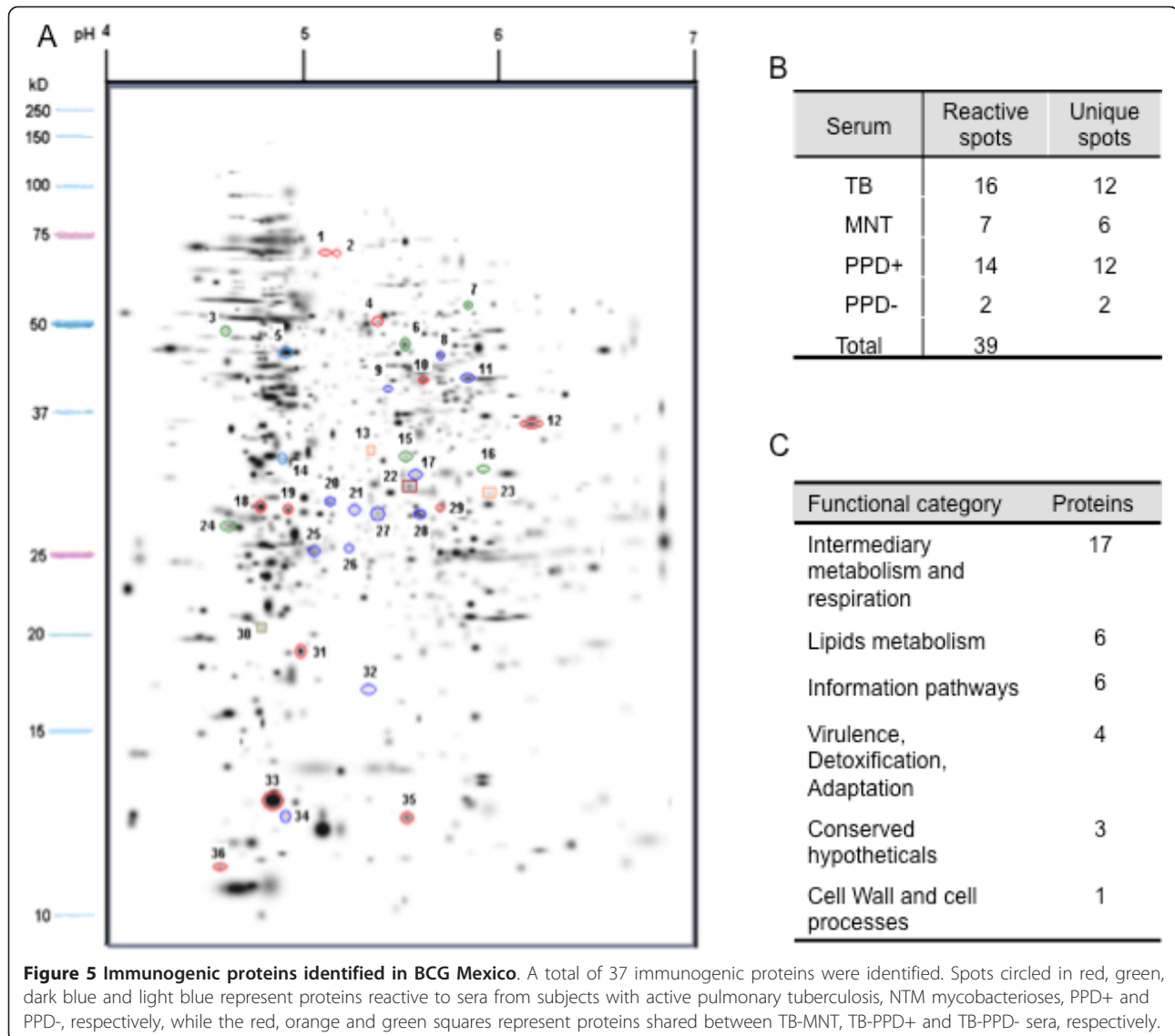


(Figure 4), with high reproducibility and low variation between duplicate experiments (0.4%). A comparative analysis of the proteomes of these BCG strains revealed 185 spots common to all strains and 136 that were unique to BCG Mexico 1931. Additionally, this analysis showed higher percentages of spots in common between BCG Mexico 1931 and BCG Danish (62%) or BCG Phipps (61%) than between BCG Mexico 1931 and BCG Tokyo (48%) (Figure 4).

Previous studies have shown that BCG strains (Connaught, Tice, Danish and Phipps) differ in their protein profiles [41-43]. Here, we observed that late strains (BCG Mexico 1931, Danish and Phipps) had a greater number of proteins in common compared to the early strain we studied (BCG Tokyo). This difference can be explained by mutations in transcriptional regulators such as *BCG3734* (*crp*) and *BCG0484c* (*sigK*) in late BCGs. Furthermore, the proteins unique to BCG Mexico 1931 may be useful for characterising this strain and explaining the causes of the observed phenotypic differences compared to other BCG strains.

Characterisation of the immune response by immune blotting

To identify immunogenic proteins in BCG Mexico 1931, we performed an immune blotting analysis. We detected 39 reactive spots in the immune proteome (Figure 5A and Additional file 1: Table S1). The largest numbers of reactive spots were obtained when using serum from subjects with active TB (16) or positive tuberculin skin test results (PPD+) (14); 12 of these spots were unique to each serum type (Figure 5B). This result indicates high variability in the proteins recognised by each type of serum. We identified 37 proteins by sequencing (Additional file 1: Table S1), the majority of which (17; 47%) corresponded to intermediary metabolism and respiration proteins (Figure 5C). Among the identified proteins, some have been previously described as virulence proteins in different strains of *M. tuberculosis*: phosphoenolpyruvate carboxykinase (*pckA*), isocitrate lyase (*icl*), 3-oxoacyl synthase II (*kasA*), *groEL*, TB27.3, the 85A and 85C antigens, alkyl-hydroperoxide reductase (*ahpC*) and heat shock protein HspX [44-47]. Rodriguez-Alvarez *et al.* have reported that GroEL and



AhpC are over-expressed in BCG Phipps [42]. To our knowledge, phosphoenolpyruvate carboxykinase, isocitrate lyase, 3-oxoacyl synthase II and AhpC are described as antigenic proteins for the first time in this report.

Conclusions

This study represents the first genomic and proteomic characterisation of BCG Mexico 1931. This substrain was used for BCG vaccine production in Mexico until 1970 and can now be used again for vaccine production and as a vector for the design of new second-generation vaccines against tuberculosis.

Initially, we determined the RD profiles of three BCG substrains representing different stages of vaccine production in Mexico. The RD profiles show that BCG

Mexico 1931 is different from BCG Mexico 1988 and 1997, which have the same profile as BCG Danish. These dates are consistent with historical records, which indicate that BCG vaccine was produced in Mexico from the BCG Danish strain after 1970.

Based on these results, the BCG Mexico 1931 substrain was used for genomic and proteomic characterisation. According to the RD profiles and genome sequence of BCG Mexico 1931, this substrain belongs to DU2 group IV within the genealogy of BCG vaccines.

Genetic studies of BCG substrains have provided new knowledge about the genes involved in the phenotypic differences observed for these strains (for example *mma3*, *fadD26*, *ppsA*, *phoP* and *whiB3*), making it possible to elucidate the basis of phenotypic variation between them. The results of this investigation and

previously published studies show that the genes with the largest numbers of mutations are related to two-component systems and transport (*esX*, *senX3-regX3*, *phoP-phoR*), regulatory proteins (*whiB* family, *crp*, *trcR*) and lipid metabolism (*fadD*, *ppsA*). Interestingly, most of these mutations may be involved in the phenotypic differences observed between BCG substrains. For example, the lack of production of some membrane lipids (PDIMs) in BCG Moreau is caused by deletions in *BCG2952* (*fadD26*) and *BCG2953* (*ppsA*) [48]. In this study, we identified specific regions in BCG Mexico 1931 that may be directly involved in the phenotypic characteristics of this strain and designated them RDMex02 and RDMex03. These regions affect the proteins FadD23 and WhiB6, which are related to the interaction between the bacteria and macrophages. In addition, we have identified SNPs in genes previously determined to be involved in the virulence of *Mycobacterium* strains. Further studies are needed to establish the contributions of these mutations and to assess the roles of the newly identified antigenic proteins in the BCG Mexico 1931 phenotype.

Methods

Bacterial strains and DNA isolation

Mycobacterium bovis BCG substrains were grown in Sauton medium for 15 days at 37°C, harvested by centrifugation and stored at -70°C until use. The Birkhaug, Connaught, Danish 1331, Frappier, Moreau, Phipps, Tice and Sweden BCG strains were kindly provided by M. Behr (McGill University Health Centre, Canada) and BCG Tokyo 172 was provided by M. Macías (Instituto Nacional de Pediatría, México). The Mexican Instituto Nacional de Higiene provided the BCG Mexico 1931, 1988 and 1997 substrains. Genomic DNA was extracted by the phenol/chloroform method following established protocols [49].

Identification of regions of difference (RD) and tandem duplications (DU) by PCR

The RD regions of the BCG Mexico substrains were profiled by multiplex PCR using primer sequences and conditions reported by the WHO (personal communication) and by Bedwell *et al.* [50]. These primers amplify the RD1 (ET1, ET2, ET3), RD2 (RD2l, RD2r), RD8 (RD8l, RD8r), RD14 (RD14l, RD14r), RD16 (RD16l, RD16r), N-RD18 (RD18L, RD18R, RD18wtR) and RD Danish/Glaxo (RD-D/G L, RD-D/G R) regions and an IS6110 insertion sequence (IS6110L, IS6110R, RD2 wtR). Similarly, a DU profile was obtained by PCR using primer sequences reported by Brosch *et al.* [9]. PCR amplification of the DU regions was performed in a volume of 30 µL with 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM MgCl₂, 0.1 mM dNTPs, 20 pmol of

each primer, 1 U of Platinum *Taq* DNA polymerase (Invitrogen, USA) and 50 ng of template DNA. The thermal profile consisted of one cycle at 94°C (5 min); 35 cycles at 94°C (1 min), 55°C (1 min) and 72°C (1 min); and one cycle at 72°C (5 min). Amplification products were analysed by horizontal electrophoresis on 3% (w/v) agarose gels stained with ethidium bromide.

BCG Mexico 1931 genome sequencing

The genome of *M. bovis* BCG Mexico 1931 was sequenced using 454 pyrosequencing, which was performed by the Sequencing Unit of CINVESTAV, Irapuato, Mexico. Additionally, a fosmid library with inserts of approximately 40 kb was constructed using the CopyControl™ pCCFOS™ system (Epicentre Technologies, USA), and the fos-end sequences of 250 clones were determined using the Sanger method (3730xl DNA Analyzer, Applied Biosystems, USA). Draft assemblies were based on 623,000 reads (36× coverage), and the Phred/Phrap/Consed software package was employed for sequence assembly and quality assessment [51] using the BCG Pasteur 1173P2 sequence as a reference [GenBank: AM408590]. To close gaps and resolve duplicated regions, the complete sequences of three fosmids (approximately 40 kb) and 110 PCR end reads were obtained. Annotation was performed using the RAST Server [52,53], Artemis [54,55] and BCGList [56].

BCG Mexico 1931 genome sequence analysis

The BCG Pasteur 1173P2, BCG Tokyo 172 [GenBank: AP010918] and BCG Mexico sequences were compared using Consed software [51] and the Basic Local Alignment and Search Tool (BLAST) [57]. The presence of the new regions of difference found in BCG Mexico 1931 (RDMex01, RDMex02 and RDMex03) in the Birkhaug, Connaught, Danish 1331, Frappier, Moreau, Phipps, Tice, Tokyo 172 and Sweden BCG strains was determined by PCR screening. The primer pairs used were 5'-GCCCAAACAGTTCGACGG-3' and 5'-CCA-GACATATGCGAGGAC-3' for RDMex01; 5'-GCGA-TAATGGGCGATGTCG-3' and 5'-CCGCGGTTGT TGAGTTCG-3' for RDMex02; and 5'-GCAGCAGT-GAACGCTTGG-3', 5'-CAGTGGAGCTGAAGGCAG-3' and 5'-GTTGCTTGGACGGCAATCG-3' for RDMex03. PCR amplifications were performed in a volume of 30 µL with 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM MgCl₂, 0.1 mM dNTPs, 20 pmol of each primer, 0.4 U Platinum *Taq* DNA polymerase (Invitrogen, USA) and 20 ng of template DNA. The thermal profile consisted of one cycle at 94°C (5 min); 35 cycles at 94°C (1 min), 55°C (1 min) and 72°C (1 min); and one cycle at 72°C (5 min). Amplification products were analysed by horizontal electrophoresis on 3% (w/v) agarose gels stained with ethidium bromide.

BCG Mexico 1931 proteome

M. bovis BCG Danish, BCG Mexico 1931, BCG Phipps and BCG Tokyo strains were grown in Middlebrook 7H9 medium (pH 7.2) for eight days at 37°C with shaking, harvested by centrifugation, washed and suspended in sterile water for lysis. Cellular proteins were obtained by sonication of mycobacteria (Ultrasonic Processor, Cole Parmer Corporation, USA) in the presence of a protease inhibitor (PMSF, 20 mM) at 4°C. The extracted proteins were quantified using a Bradford assay. For 2D-PAGE, approximately 100 mg of protein was solubilised, denatured, reduced in sample buffer [4% CHAPS, 2 M urea, 70 mM l-dithiothreitol (DTT), 0.001% bromophenol blue, and 0.1% 3-10 ampholyte] and used to rehydrate 11-cm pH 4-7 IPG strips (ReadyStrip™, IPG strips, Bio-Rad). Isoelectric focusing (IEF) was performed on a Multiphor II system (Amersham Biosciences, England) until reaching 52,000 VH at 17°C. The strips were equilibrated twice in a solution containing 4% urea, 30% glycerol (v/v), 50 mM Tris-HCl (pH 8.8), 2% SDS (w/v), and 0.002% bromophenol blue supplemented with 90 mM DTT for the first incubation (15 minutes) and 250 mM iodoacetamide (IAA) for the second incubation (15 minutes). Second-dimension electrophoresis was performed on a 12% polyacrylamide gel (Hoeffer SE-600, Amersham Biosciences, England) for approximately five hours with a voltage gradient of 50-200 V. Once fixed, the proteins were silver-stained, and gel images were captured in a digital format (Molecular Imager GS-800™ Calibrated Densitometer, Bio-Rad, USA). Gel analysis was performed using the program 2-D PDQuest Advance V.8.0 (Bio-Rad, USA). Duplicate gels with proteins obtained from independent cultures were included in the analysis. A master image gel was created using three replicates of each experiment and was used for comparison [42].

Immune blotting

To identify antigenic proteins in BCG Mexico 1931, we conducted an immune blot analysis from the 2D-PAGE gels. Proteins were transferred onto Hybond P polyvinylidene difluoride (PVDF) membranes (GE Healthcare, England) using a Trans-Blot SD Semi-Dry Transfer system (Bio-Rad, USA) for 1 h at 10 V. The membranes were blocked with Tris-buffered saline (TBS) containing 0.05% Tween-20 and 5% skim milk at 4°C overnight. The membranes were then incubated for 1 h at room temperature with sera from subjects with PPD+, PPD-, pulmonary TB or mycobacterioses caused by NTMs. The membranes were incubated with selected sera from each group having the highest IgG2 titres against each of the groups described above (data not shown). Then membranes were subsequently incubated with an alkaline phosphatase-conjugated mouse anti-human IgG2

antibody (1:5000; Zymed, Invitrogen, USA) for 1 h at room temperature. Immune detection was accomplished using the Inmobilon Western System (Millipore Co, USA). Chemiluminescence signals were measured using the Genius Plus system (TECAN, Switzerland).

Protein sequencing

Reactive proteins were sequenced using a 3200 QTRAP hybrid tandem mass spectrometer (3200 QTRAP, Applied Biosystems, USA) equipped with a nano-electrospray ion source (NanoSpray II) and a MicroIonSpray II head. Proteins were identified based on their MS/MS spectra datasets using the MASCOT search algorithm (Version 1.6b9, Matrix Science, London, UK). A BLAST search was conducted using the *M. tuberculosis* complex subset of the National Centre for Biotechnology Information (NCBI) non-redundant database (NCBI nr20070623).

Nucleotide and protein sequence accession numbers

The complete genome sequence of *Mycobacterium bovis* BCG Mexico 1931 has been deposited in the NCBI GenBank database under accession number CP002095. The GenBank accession numbers for the identified protein sequences are: JN034599 (*pckA*), JN034600 (*tig*), JN034601 (*sahH*), JN034602 (*atpD*), JN034603 (*icl*), JN034604 (*fumC*), JN034605 (*lpd*), JN034606 (*tuf*), JN034607 (*kasA*), JN034608 (*fadA*), JN034609 (*kasB*), JN034610 (*BCGMEX_2464c*), JN034611 (*thrC*), JN034612 (*groEL2*), JN034613 (*BCGMEX_2988*), JN034614 (*BCGMEX_3456c*), JN034615 (*TB31.7*), JN034616 (*tsf*), JN034617 (*nusG*), JN034618 (*fixB*), JN034619 (*BCGMEX_0855c*), JN034620 (*cysA2*), JN034621 (*sucD*), JN034622 (*cfp30B*), JN034623 (*echA8*), JN034624 (*fbpC*), JN034625 (*fbpA*), JN034626 (*nuoC*), JN034627 (*ahpC*), JN034628 (*clpP*), JN034629 (*ssb*), JN034630 (*rplL*), JN034631 (*hspX*) and JN034632 (*ndkA*).

Additional material

Additional file 1: TableS1. Detection data for antigenic proteins in BCG Mexico 1931.

List of abbreviations

BCG: Bacille Calmette-Guérin; DU: Tandem Duplications; NTM: Non-Tuberculous Mycobacteria; nsSNPs: Nonsynonymous SNPs; RD: Regions of Difference; SNPs: Single Nucleotide Polymorphisms; WHO: World Health Organization.

Acknowledgements and Funding

The authors are grateful to Enrique Merino for his contribution to the SNP analysis, Erick Amador-Gaytán and Jaqueline Larracilla-Camacho for their participation in proteomic data collection and Humberto Barrios for support in the genomic data analysis. This work was supported by the Consejo

Nacional de Ciencia y Tecnología (CONACyT) México grant SALUD-2007-C01-68729. PO is the recipient of a CONACyT- México doctoral scholarship (Reg. No.195403) at Programa de Maestría y Doctorado en Ciencias Bioquímicas, Facultad de Química, UNAM.

Author details

¹Programa de Inmunología Molecular Microbiana, Departamento de Microbiología y Parasitología, Facultad de Medicina, Universidad Nacional Autónoma de México, DF, México. ²Genómica Evolutiva, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México. ³Biológicos y Reactivos de México, Estado de México, México. ⁴Laboratorio de Péptidos y Proteínas, Departamento de Bioquímica, Facultad de Medicina, Universidad Nacional Autónoma de México, DF, México.

Authors' contributions

PO participated in the design and performance of the experiments, and writing of the paper; IGH and AAH performed experiments; GMH performed protein identification by sequencing; MAC, SPL and YLV participated in the study design and writing of the paper. All authors read and approved the final manuscript.

Received: 20 May 2011 Accepted: 8 October 2011

Published: 8 October 2011

References

1. WHO: WHO REPORT 2010: Global Tuberculosis Control. Switzerland; 2010.
2. WHO: BCG Vaccine. *Weekly Epidemiological Record* 2004, **4**.
3. Tuberculosis. [http://www.who.int/topics/tuberculosis/en/].
4. Fine PE: BCG vaccination against tuberculosis and leprosy. *Br Med Bull* 1988, **44**(3):691-703.
5. Trunz BB, Fine P, Dye C: Effect of BCG vaccination on childhood tuberculous meningitis and military tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness. *Lancet* 2006, **367**(9517):1173-1180.
6. Mendoza-Coronel E, Camacho-Sandoval R, Bonifaz LC, Lopez-Vidal Y: PD-L2 induction on dendritic cells exposed to *Mycobacterium avium* downregulates BCG-specific T cell response. *Tuberculosis (Edinb)* 2011, **91**(1):36-46.
7. Liu J, Tran V, Leung AS, Alexander DC, Zhu B: BCG vaccines: their mechanisms of attenuation and impact on safety and protective efficacy. *Hum Vaccin* 2009, **5**(2):70-78.
8. Oettinger T, Jorgensen M, Ladefoged A, Haslov K, Andersen P: Development of the *Mycobacterium bovis* BCG vaccine: review of the historical and biochemical evidence for a genealogical tree. *Tuber Lung Dis* 1999, **79**(4):243-250.
9. Brosch R, Gordon SV, Garnier T, Eiglmeier K, Frigui W, Valenti P, Dos Santos S, Duthoy S, Lacroix C, Garcia-Pelayo C, et al: Genome plasticity of BCG and impact on vaccine efficacy. *Proc Natl Acad Sci USA* 2007, **104**(13):5596-5601.
10. Brosch R, Gordon SV, Buchrieser C, Pym AS, Garnier T, Cole ST: Comparative genomics uncovers large tandem chromosomal duplications in *Mycobacterium bovis* BCG Pasteur. *Yeast* 2000, **17**(2):111-123.
11. Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, Small PM: Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 1999, **284**(5419):1520-1523.
12. Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK: Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J Bacteriol* 1996, **178**(5):1274-1282.
13. Hsu T, Hingley-Wilson SM, Chen B, Chen M, Dai AZ, Morin PM, Marks CB, Padiyar J, Goulding C, Gingery M, et al: The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *Proc Natl Acad Sci USA* 2003, **100**(21):12420-12425.
14. Lewis KN, Liao R, Guinn KM, Hickey MJ, Smith S, Behr MA, Sherman DR: Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guerin attenuation. *J Infect Dis* 2003, **187**(1):117-123.
15. Pym AS, Brodin P, Brosch R, Huerre M, Cole ST: Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*. *Mol Microbiol* 2002, **46**(3):709-717.
16. Kozak R, Behr MA: Divergence of immunologic and protective responses of different BCG strains in a murine model. *Vaccine* 2010, **29**(7):1519-1526.
17. Castillo-Rodal AI, Castanon-Arreola M, Hernandez-Pando R, Calva JJ, Sada-Diaz E, Lopez-Vidal Y: *Mycobacterium bovis* BCG substrains confer different levels of protection against *Mycobacterium tuberculosis* infection in a BALB/c model of progressive pulmonary tuberculosis. *Infect Immun* 2006, **74**(3):1718-1724.
18. Behr MA, Schroeder BG, Brinkman JN, Slayden RA, Barry CE: A point mutation in the *mma3* gene is responsible for impaired methoxymycolic acid production in *Mycobacterium bovis* BCG strains obtained after 1927. *J Bacteriol* 2000, **182**(12):3394-3399.
19. Walker KB, Brennan MJ, Ho MM, Eskola J, Thiry G, Sadoff J, Dobbelaer R, Grode L, Liu MA, Fruth U, et al: The second Geneva Consensus: Recommendations for novel live TB vaccines. *Vaccine* 2010, **28**(11):2259-2270.
20. Hayashi D, Takii T, Mukai T, Makino M, Yasuda E, Horita Y, Yamamoto R, Fujiwara A, Kanai K, Kondo M, et al: Biochemical characteristics among *Mycobacterium bovis* BCG substrains. *FEMS Microbiol Lett* 2010, **306**(2):103-109.
21. Sosa J: Métodos utilizados en el estudio de la vacuna BCG. *Salud Pública de México* 1967, **IX**(2):213-219.
22. P León A: S.S.A. Laboratorio de Producción de Vacuna B.C.G. *Salud Pública de México* 1963, **V**(6):1006-1011.
23. P León A, Jiménez M: La vacunación contra la tuberculosis en México. *Boletín de la oficina Sanitaria Panamericana* 1950, **29**(1):7-13.
24. Instituto-Nacional-de-Higiene: Cien años de lucha por la salud. Primera Edición edn. México: Grupo DESEA, S.A. de C.V; 1995.
25. Fernández d Castro J: Estado actual de la vacunación con BCG. *Gaceta Médica de México* 1976, **III**(4):271-279.
26. Hayashi D, Takii T, Fujiwara N, Fujita Y, Yano I, Yamamoto S, Kondo M, Yasuda E, Inagaki E, Kanai K, et al: Comparable studies of immunostimulating activities in vitro among *Mycobacterium bovis* bacillus Calmette-Guerin (BCG) substrains. *FEMS Immunol Med Microbiol* 2009, **56**(2):116-128.
27. Seki M, Honda I, Fujita I, Yano I, Yamamoto S, Koyama A: Whole genome sequence analysis of *Mycobacterium bovis* bacillus Calmette-Guerin (BCG) Tokyo 172: a comparative study of BCG vaccine substrains. *Vaccine* 2009, **27**(11):1710-1716.
28. Lynett J, Stokes RW: Selection of transposon mutants of *Mycobacterium tuberculosis* with increased macrophage infectivity identifies *fadD23* to be involved in sulfolipid production and association with macrophages. *Microbiology* 2007, **153**(Pt 9):3133-3140.
29. Molina-Olvera G: Sobrevivencia y multiplicidad de distintas subcepas de *Mycobacterium bovis* BCG en macrófagos de la línea celular THP-1. México: Universidad Nacional Autónoma de México; 2010.
30. Alam MS, Garg SK, Agrawal P: Studies on structural and functional divergence among seven *WhiB* proteins of *Mycobacterium tuberculosis* H37Rv. *FEBS J* 2009, **276**(1):76-93.
31. Garcia Pelayo MC, Uplekar S, Keniry A, Mendoza Lopez P, Garnier T, Nunez Garcia J, Boschirolu L, Zhou X, Parkhill J, Smith N, et al: A comprehensive survey of single nucleotide polymorphisms (SNPs) across *Mycobacterium bovis* strains and *M. bovis* BCG vaccine strains refines the genealogy and defines a minimal set of SNPs that separate virulent *M. bovis* strains and *M. bovis* BCG strains. *Infect Immun* 2009, **77**(5):2230-2238.
32. Charlet D, Mostowy S, Alexander D, Sit L, Wiker HG, Behr MA: Reduced expression of antigenic proteins MPB70 and MPB83 in *Mycobacterium bovis* BCG strains due to a start codon mutation in *sigK*. *Mol Microbiol* 2005, **56**(5):1302-1313.
33. Bai G, Gazdik MA, Schaak DD, McDonough KA: The *Mycobacterium bovis* BCG cyclic AMP receptor-like protein is a functional DNA binding protein in vitro and in vivo, but its activity differs from that of its *M. tuberculosis* ortholog, Rv3676. *Infect Immun* 2007, **75**(11):5509-5517.
34. Glickman MS, Cox JS, Jacobs WR Jr: A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of *Mycobacterium tuberculosis*. *Mol Cell* 2000, **5**(4):717-727.
35. Rao V, Fujiwara N, Porcelli SA, Glickman MS: *Mycobacterium tuberculosis* controls host innate immune activation through cyclopropane modification of a glycolipid effector molecule. *J Exp Med* 2005, **201**(4):535-543.

36. Parish T, Smith DA, Roberts G, Betts J, Stoker NG: **The senX3-regX3 two-component regulatory system of *Mycobacterium tuberculosis* is required for virulence.** *Microbiology* 2003, **149**(Pt 6):1423-1435.
37. Glover RT, Kriakov J, Garforth SJ, Baughn AD, Jacobs WR Jr: **The two-component regulatory system senX3-regX3 regulates phosphate-dependent gene expression in *Mycobacterium smegmatis*.** *J Bacteriol* 2007, **189**(15):5495-5503.
38. Vandal OH, Roberts JA, Odaira T, Schnappinger D, Nathan CF, Ehrst S: **Acid-susceptible mutants of *Mycobacterium tuberculosis* share hypersusceptibility to cell wall and oxidative stress and to the host environment.** *J Bacteriol* 2009, **191**(2):625-631.
39. Sampson SL: **Mycobacterial PE/PPE proteins at the host-pathogen interface.** *Clin Dev Immunol* 2011, **2011**:497203.
40. Sassetti CM, Boyd DH, Rubin EJ: **Genes required for mycobacterial growth defined by high density mutagenesis.** *Mol Microbiol* 2003, **48**(1):77-84.
41. Jungblut PR, Schaible UE, Mollenkopf HJ, Zimny-Arndt U, Raupach B, Mattow J, Halada P, Lamer S, Hagens K, Kaufmann SH: **Comparative proteome analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG strains: towards functional genomics of microbial pathogens.** *Mol Microbiol* 1999, **33**(6):1103-1117.
42. Rodriguez-Alvarez M, Mendoza-Hernandez G, Encarnacion S, Calva JJ, Lopez-Vidal Y: **Phenotypic differences between BCG vaccines at the proteome level.** *Tuberculosis (Edinb)* 2009, **89**(2):126-135.
43. Rodriguez-Alvarez M, Palomec-Nava ID, Mendoza-Hernandez G, Lopez-Vidal Y: **The secretome of a recombinant BCG substrain reveals differences in hypothetical proteins.** *Vaccine* 2010, **28**(23):3997-4001.
44. Liu K, Ba X, Yu J, Li J, Wei Q, Han G, Li G, Cui Y: **The phosphoenolpyruvate carboxykinase of *Mycobacterium tuberculosis* induces strong cell-mediated immune responses in mice.** *Mol Cell Biochem* 2006, **288**(1-2):65-71.
45. Liu K, Yu J, Russell DG: **pckA-deficient *Mycobacterium bovis* BCG shows attenuated virulence in mice and in macrophages.** *Microbiology* 2003, **149**(Pt 7):1829-1835.
46. Armitige LY, Jagannath C, Wanger AR, Norris SJ: **Disruption of the genes encoding antigen 85A and antigen 85B of *Mycobacterium tuberculosis* H37Rv: effect on growth in culture and in macrophages.** *Infect Immun* 2000, **68**(2):767-778.
47. Heym B, Stavropoulos E, Honore N, Domenech P, Saint-Joanis B, Wilson TM, Collins DM, Colston MJ, Cole ST: **Effects of overexpression of the alkyl hydroperoxide reductase AhpC on the virulence and isoniazid resistance of *Mycobacterium tuberculosis*.** *Infect Immun* 1997, **65**(4):1395-1401.
48. Leung AS, Tran V, Wu Z, Yu X, Alexander DC, Gao GF, Zhu B, Liu J: **Novel genome polymorphisms in BCG vaccine strains and impact on efficacy.** *BMC Genomics* 2008, **9**:413.
49. Sonnenberg JTBaMG: **Isolation of Genomic DNA from *Mycobacteria*.** In *Mycobacteria Protocols. Volume 101*. Edited by: Stoker TPANg. New Jersey: Human Press; 1998.
50. Bedwell J, Kairo SK, Behr MA, Bygraves JA: **Identification of substrains of BCG vaccine using multiplex PCR.** *Vaccine* 2001, **19**(15-16):2146-2151.
51. **Phred, Phrap, Consed.** [<http://www.phrap.org/phredphrapconsed.html>].
52. **Rapid Annotation using Subsystem Technology.** [<http://rast.nmpdr.org/>].
53. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, et al: **The RAST Server: rapid annotations using subsystems technology.** *BMC Genomics* 2008, **9**:75.
54. Berriman M, Rutherford K: **Viewing and annotating sequence data with Artemis.** *Brief Bioinform* 2003, **4**(2):124-132.
55. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B: **Artemis: sequence visualization and annotation.** *Bioinformatics* 2000, **16**(10):944-945.
56. **BCGList.** [<http://genolist.pasteur.fr/BCGList/>].
57. **Basic Local Alignment and Search Tool.** [<http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi>].

doi:10.1186/1471-2164-12-493

Cite this article as: Orduña et al.: Genomic and proteomic analyses of *Mycobacterium bovis* BCG Mexico 1931 reveal a diverse immunogenic repertoire against tuberculosis infection. *BMC Genomics* 2011 **12**:493.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



Spot	Accession number	Protein		Gene		Matched peptides	Score	Global Score ^{&}	Sequence coverage %
		Name	Function (FC*)	Locus BCG ^{o2}	Gene Name				
1	15607352	Phosphoenolpyruvate carboxykinase	Rate-limiting gluconeogenic enzyme (7)	0248	<i>pckA</i>	35-47/VVFTDGSEEEFQR	43	401	25
						68-81/NSYLALSDPSDVAR	61		
						86-92/TYICSAK	13		
						170-178/MGDDGFFVK	46		
						179-192/ALHSVGAPLEPGQK	13		
						179-192/ALHSVGAPLEPGQK	33		
						255-261/LISPENK	19		
						276-289/TNLAMLQPTIPGWR	31		
						290-301/AETLGDDIAWMR	12		
						378-389/ETETNAAHPNSR	6		
						433-455/DWQHGVFIGATLGSEQTAAAEKG	22		
						496-503/VFFVNWFR	38		
						510-520/FLWPGFGENSR	34		
						590-602/LPTGVKDEFDALK	49		
2	15607352	Phosphoenolpyruvate carboxykinase	Rate-limiting gluconeogenic enzyme (7)	0248	<i>pckA</i>	68-81 NSYLALSDPSDVAR	45	404	23
						170-178 MGDDGFFVK	24		
						179-192 ALHSVGAPLEPGQK	5		
						193-202 DVAWPCSETK	3		
						255-261 LISPENK	16		
						276-289 TNLAMLQPTIPGWR	27		
						290-301AETLGDDIAWMR	53		
						308-327 LYAVNPEFGFFGVAPGTNWK	21		
						423-432 TTVPLVTEAR	48		
						496-503 VFFVNWFR	51		
						510-520 FLWPGFGENSR	40		
						590-602 LPTGVKDEFDALK	45		
						590-602 LPTGVKDEFDALK	40		
						3	15609599		
15-30/INVEVPFAELEPDFQR	25								
41-48/LPGFRPGK	7								
61-75/EAMLDQIVNDALPSR	65								
195-212/LIAGLDDAVVGLSADESR	68								
195-212/LIAGLDDAVVGLSADESR	39								
218-235/LAAGEHAGQEAQVTVTVR	12								
264-271/ASLSDQVR	46								
275-283/RAQQAEQIR	3								
325-336/FNELLVEQGSSR	5								
337-344/AAFDAEAR	57								
383-398/QYGIEPQQLFGYLQER	19								
399-409/NQLPTMFADVR	31								
399-409 /NQLPTMFADVR	23								
4	215432212	S-adenosyl-L-homocysteine hydrolase	Thioester hydrolase which acting on ether bounds. Could be involved in methionine and seleno amino acid metabolisms (7)	3277	<i>sahH</i>	441-458/VSAGEAEEAEPADEGAAR	83	746	47
						2-8/TGNLVTK	17		
						9-16/NSLTPDVR	31		
						23-33/IADLSLADFGR	71		
						38-50/IAEHMPGLMSLR	13		
						52-60/EYAEVQPLK	30		
						149-172/SANMILDDGGDATMLVLRGMQYEK	9		
189-195/ VFLNLLR	45								

					208–213/ IAESVK	28			
					214–225/GVTEETTTGVL	68			
					226–248/LYQFAAAGDLA	11			
					259–267/HSLIDGINR	42			
					268–276/GTDALIGGK	45			
					342–350/DIIMLEHIK	31			
					342–350/DIIMLEHIK	13			
					354 375/DHAILGNIGHFD	73			
					354–375/DHAILGNIGHFD	17			
					381–39/VNVKPQVDLWTF	61			
					398–406/SIIVLSEGR	55			
					439–448/NDEYDNEVYR	42			
					460–471/IHVEALGGHLTK	17			
					460–471/IHVEALGGHLTK	34			
					475–494/EQAEYLGVD	45			
5	15608450	F0F1 ATP synthase subunit beta	Produces ATP from ADP in the presence of a proton gradient across the membrane (7)	1370	<i>atpD</i>	8–22/TDRPGKPGSSDTSGR	13	1290	68
					8–22/TDRPGKPGSSDTSGR	41			
					26–37/VTGPVVDVEFPR	19			
					26–37/VTGPVVDVEFPR	70			
					38–58/GSIPELFNALHAEITFESLAK	27			
					59–74/TLTLEVAQHLGDNLVR	40			
					59–74/ TLTLEVAQHLGDNLVR	44			
					75–86/TISLQPTDGLVR	63			
					87–95/GVEVIDTGR	67			
					96–106/SISVPVGEVVK	45			
					107–125/GHVFNALGDCLDEPGYGEK	17			
					107–125/GHVFNALGDCLDEPGYGEK	11			
					126–133/FEHWSIHR	33			
					134–144/KPPAFEELEPR	55			
					145–153/TEMLETGLK	45			
					145–153/TEMLETGLK	43			
					154–163/VVDLLTPYVR	63			
					167–177/ IALFGGAGVGK	63			
					178–187/TVLIQEMINR	53			
					178–187/TVLIQEMINR	48			
					191–204/NFGGTSVFAGVGER	27			
					191–204/NFGGTSVFAGVGER	78			
					207–222/EGNDLWVELAEANVLK	54			
					223–238/DTALVFGQMDEPPGTR	58			
					223–238/DTALVFGQMDEPPGTR	10			
					241–253/VALSALTMAEWFR	94			
					241–253/VALSALTMAEWFR	62			
					270–283/FTQAGSEVSTLLGR	87			
					284–304/MPSAVGYQPTLADEMGEQER	34			
					352–378/GIFPAVDPLASSSTILDPSVVGDEHYR	75			
					379–385/VAQEVIR	49			
					445–452/ETIEAFDR	54			
6	15839856	Isocitrate lyase	Involved in glyoxylate bypass (at the first step), an alternative to the tricarboxylic acid cycle (7)	0507	<i>icl</i>	456–477/GDFDHFVPEQAFFLIGLDDLAK	28	789	46
					1 22/MSVVGTPKSAEQIQQEWDTNPR	13			
					2–8/SVVGTPK	21			
					9–22/SAEQIQQEWDTNPR	39			
					29–50/TYSAEDVVALQGSVVEHTLAR	95			
					124–130/INNALQR	44			
					170–189/ALIAAGVAGSHWEDQLASEK	84			

					191-197/CGHLGGK	22			
					198-207/VLIPQHIR	11			
					198-207/VLIPQHIR	41			
					198-207/VLIPQHIR	38			
					214-228/LAADVADVPTVVIAR	50			
					229-244/TDAEAATLITSDVDER	69			
					245-253/DQPFITGER	38			
					263-271/NGIEPCAR	20			
					263-271/NGIEPCAR	33			
					274-295/AYAPFADLIWMETGTPDLEAAR	23			
					296-302/QFSEAVK	42			
					322-331/KHLDDATIAK	33			
					323-331/HLDDATIAK	8			
					323-331/HLDDATIAK	9			
					335-342/ELAAMGFK	42			
					380-386/EFAAEER	60			
					396-404/EVGAGYFDR	46			
7	15608238	Fumarate hydratase	Involved in the tricarboxylic acid cycle (7)	1158c	<i>fumC</i>	2-11/AVDADSANYR	77	549	37
						12-21/IEHDTMGEVR	35		
						12-21/IEHDTMGEVR	21		
						35-44/AVENFPISGR	50		
						59-75/GACAQVNSDLGLLAPEK	10		
						123-173GGVTLHPNDDVNMSQSSNDTFPTATHIAATEAAVA	12		
						HLIPALQQLHDALAAK			
						174-182/ALDWHTVVK	55		
						186-205/THLMDAVPVTLGQEFSGYAR	44		
						216-220/ACLPR	38		
						260-271/TAANSFEAQAAR	82		
						260-271/TAANSFEAQAAR	6		
						291-296/IANDIR	45		
						449-456/LSIEDLDR	60		
						457-465/RLDVLAMAK	58		
						458-465/LDVLAMAK	30		
						466-474/AEQLDSDRL	42		
8	15607603	Dihydrolipoamide dehydrogenase	Involved in energy metabolism (7)	0502	<i>lpd</i>	2-21/THYDVVVVLGAGPGGYVAAIR	23	331	26
						22-35/AAQLGLSTAIVEPK	50		
						55-64/NAELVHIFTK	40		
						148-168/LVPGTSLSANVVTYEEQILSR	9		
						206-216/ALPNEDADVSK	56		
						281-288/AGVALTDR	34		
						290-298/AIGVDDYMR	36		
						290-298/AIGVDDYMR	20		
						367-376/NEGYDVVVAK	35		
						377-384/FPFTANAK	25		
						385-396/AHGVGDPSGFVK	27		
9	15607825	Elongation factor Tu	This protein promotes the GTP-dependent binding of aminoacyl-tRNA to the a-site of ribosomes during protein biosynthesis (2)	0734	<i>tuf</i>	9-25/TKPHVNIGTIGHVDHGK	34	1020	70
						9-25/TKPHVNIGTIGHVDHGK	40		
						9-25/TKPHVNIGTIGHVDHGK	30		
						26-34/TTLTAAITK	47		
						35-47/VLHDKFPDLNETK	8		
						35-47/VLHDKFPDLNETK	33		
						40-47/FPDLNETK	40		
						48-59/AFDQIDNAPEER	71		
						62-76/GITINIAHVEYQTDK	65		

						62-76/GITINIAHVEYQTDK	37		
						78-92/HYAHVDAPGHADYIK	56		
						78-92/HYAHVDAPGHADYIK	48		
						78-92/HYAHVDAPGHADYIK	67		
						120-126/EHVLLAR	49		
						127-139/QVGVPYILVALNK	60		
						140-157/ADAVDDEELLELVEMEV	45		
						140-157/ADAVDDEELLELVEMEV	52		
						140-157/ADAVDDEELLELVEMEV	29		
						140-157/ADAVDDEELLELVEMEV	55		
						158-173/ELLAAQEFDEDAPVVR	65		
						186-206/WVASVEELMNAVDESIPDPVR	27		
						207-225/ETDKPFLMPVEDVFTITGR	19		
						207-225/ETDKPFLMPVEDVFTITGR	75		
						207-225/ETDKPFLMPVEDVFTITGR	49		
						236-255/GVINVNEEVEIVGIRPSTTK	42		
						256-265/TTVTGVEMFR	64		
						256-265/TTVTGVEMFR	39		
						267-282/LLDQQQAGDNVGLLLR	112		
						292-316/GQVVTKPGTTTPHTEFEGQVYILSK	21		
						322-336/HTPFFNNYRPFYFR	22		
						322-336/HTPFFNNYRPFYFR	14		
						365-376/LIQPVAMDEGLR	63		
						365 - 376/LIQPVAMDEGLR	45		
						377-384/FAIREGGR	8		
10	15609382	3-oxoacyl-(acyl carrier protein) synthase II	Involved in fatty acid biosynthesis (mycolic acids synthesis) (1)	2262	<i>kasA</i>	35-53/GLLAGESGIHALEDEFVTK	49	306	38
						80-85/MSYVQR	34		
						108-121/FAVVVGTGLGGAER	48		
						122-135/IVESYDLMNAGGPR	8		
						162-183/AGVMTPVSACSSGSEIAIAHAWR	9		
						258-264/GAKPLAR	14		
						265-286/LLGAGITSDAFHMVAPAADGVR	26		
						328-340/VAGCDQAAVYAPK	41		
						341-363/SALGHSIGAVGALESVLTVLTLR	46		
11	15607999	Acetyl-CoA acetyltransferase	Function unknown, but involvement in lipid degradation (1)	0911	<i>fadA</i>	396-415/YAVNNSFGFGGHVVALAFGR	35	482	51
						2-12/SEEFYEAIR	76		
						20-40/NGSLHEVKPLSLVVGLIDELR	61		
						72-90/AAVLASGMPVTSGGVQLNR	20		
						72-90/AAVLASGMPVTSGGVQLNR	39		
						91-105/FCASGLEAVNTAAQK	40		
						108-124/SGWDDLVLGGVESMSR	22		
						165-173/EDVDAYALR	58		
						178-189/AEEAWSGGYFAK	28		
						190-195/SVVPVR	31		
						196-215/DQNGLLILDHDEHMRPDTTK	5		
						221-241/LKPAFEGLAALGGFDDVALQK	44		
						242-247/YHWVEK	32		
						248-276/INHVHTGGNSSGIVDGAALVMIGSAAAGK	39		
						277-283/LQGLTPR	39		
						309-313/KVLDR	15		
						340-347/DLNIPDEK	36		
12	15609383	3-oxoacyl-(acyl carrier protein) synthase II	Involved in fatty acid biosynthesis (mycolic acids synthesis) (1)	2263	<i>kasB</i>	55-59/LLLDR	45	918	63
						65-79/TLDDPFVEEFDLPVR	58		
						80-94/IGGHLLEEFDHQLTR	16		

						80-94/IGGHLLLEFDHQLTR	18		
						80-94/IGGHLLLEFDHQLTR	55		
						100-105/MGYLQR	46		
						100-105/MGYLQR	42		
						106-112/MSTVLSR	48		
						106-112/MSTVLSR	46		
						113-127/RLWENAGSPEVDTNR	44		
						156-164/AVSPLTVQK	45		
						165-178/YMPNGAAAAVGLER	72		
						165-178/YMPNGAAAAVGLER	54		
						182-200/AGVMTPVSACASGAEAIAR	26		
						182-200/AGVMTPVSACASGAEAIAR	70		
						201-221/AWQQIVLGEADAAICGGVETR	51		
						222-234/IEAVPIAGFAQMR	62		
						222-234/IEAVPIAGFAQMR	30		
						235-253/IVMSTNNDPAGACRPFDR	75		
						235-253/IVMSTNNDPAGACRPFDR	16		
						235-253/IVMSTNNDPAGACRPFDR	59		
						256-276/DGFVFGEGGALLIETEEHAK	27		
						286-307/IMGASITSDGFHMVAPDPNGER	49		
						286-307/IMGASITSDGFHMVAPDPNGER	17		
						308-314/AGHAITR	30		
						308-314/AGHAITR	12		
						308-314/AGHAITR	30		
						315-345/AIQLAGLAPGDIDHVNAHATGTQVGDIAEGR	64		
						315-345/AIQLAGLAPGDIDHVNAHATGTQVGDIAEGR	35		
						346-362/AINNALGGNRPVYAPK	16		
						363-385/SALGHSVAVGAVESILTVLALR	75		
						363-385/SALGHSVAVGAVESILTVLALR	37		
13	15609591	2-oxoglutarate ferredoxin oxidoreductase subunit beta	Unknown; probably involved in cellular metabolism (7)	2474c	---	32-43/NAGVPTTDQPPK	19	219	26
						94-109/FPYYLETYGFHSIHGR	38		
						110-121/APAIATGLALAR	32		
						182-208/STPMGSLDHPFNPVSLALGAEATFVGR	24		
						317-331/LSDQNLDTVLGIFR	18		
						317-331/LSDQNLDTVLGIFR	56		
						351-367/NAAPSGTAALQSLLHGR	53		
						351-367/NAAPSGTAALQSLLHGR	22		
14	15608435	Threonine synthase	Involved in threonine biosynthesis (7)	1355	<i>thrC</i>	50-58/QTGCTIHLK	21	384	37
						59-69/VEGLNPTGSFK	45		
						72-86/GMTMAVTDALAHGQR	33		
						107-119/AGITCAVLIPQGK	64		
						125-134/LAQAVMHGAK	54		
						125-134/LAQAVMHGAK	18		
						207-219/GYTEYHQLGLIDK	20		
						252-269/IGSPASWTSVEAQQQSK	21		
						272-289/FLAASDEEILAAHYHLVAR	44		
						290-308/VEGVFVEPASAASIAGLLK	37		
						309-317/AIDGWVVAR	47		
15	15607581	Chaperonin GroEL	Prevents misfolding and promotes the refolding and proper assembly of unfolded polypeptides generated under stress conditions (0)	0479	<i>groEL2</i>	68-74/IGAELVK	9	964	49
						79-100/KTDDVAGDGTATVLAQALVR	24		
						105-116/NVAAGANPLGLK	22		
						238-264/VIGAGKPLIIAEDVEGEALSTLVVVK	73		
						276-282/APGFGDR	45		
						320-325/KVVVTK	48		

						320-325/KVVVTK	30		
						321-343/VVVTKDETTIVEGAGDTDAIAGR	41		
						326-343/DETTIVEGAGDTDAIAGR	68		
						349-360/QEIENSDSDYDR	49		
						349-362/QEIENSDSDYDREK	37		
						370-378/LAGGVAVIK	69		
						370-378/LAGGVAVIK	13		
						379-388/AGAATEVELK	70		
						394-399/IEDAVR	38		
						403-428/AAVEEGIVAGGGVTLQAAPTLDELK	11		
						429-441/LEGDEATGANIVK	96		
						442-449/VALEAPLK	55		
						450-465/QIAFNGLPEPGVVAEK	84		
						468-495/NLPAGHGLNAQTGVYEDLLAAGVADPVK	72		
						499-524/SALQNAASIAGLFLTTEAVVADKPEK	34		
						499-524/SALQNAASIAGLFLTTEAVVADKPEK	11		
						499-524/SALQNAASIAGLFLTTEAVVADKPEK	50		
						527-540/ASVPGGGDMGGMDF	16		
16	15610108	Oxidoreductase	Unknown; probably involved in cellular metabolism (7)	2993	---	39-49/AVSAALEIGCR	66	719	64
						50-65/LIDTAYAYGNEAAVGR	126		
						66-74/AIAASGVAR	68		
						75-82/EELFVTTK	43		
						83-92/LATPDQGFTR	54		
						83-92/LATPDQGFTR	269		
						104-123/LGLDYVDLYLIHWPAPPVVK	15		
						124-135/YVDAWGGMIQSR	61		
						124-135/YVDAWGGMIQSR	64		
						182-201/KANAQHTVVTQSYCPLALGR	93		
						182-201/KANAQHTVVTQSYCPLALGR	7		
						202-217/LLDNPTVTSIASEYVK	74		
						218-225/TPAQVLLR	61		
						226-237/WNLQLGNVVR	56		
						244-271/IASNFVDFELAAEHMDALGGLNDGTR	12		
						244-271/IASNFVDFELAAEHMDALGGLNDGTR	38		
						272-282/VREDPLTYAGT	53		
17	15610525	Dehydrogenase	Unknown; probably involved in cellular metabolism (7)	3458c	---	2-22/AIDPNSIGAVTEPMLFEWTR	65	949	66
						70-86/VGTFNPAALLHGSQGIR	56		
						70-86/VGTFNPAALLHGSQGIR	91		
						87-96/LHAPLPAAGK	43		
						97-109/LSVVTEVADIQDK	58		
						97-113/LSVVTEVADIQDKGEGK	68		
						114-120/NAIVVLR	45		
						123-142/GCDPESGSLVAETLTTLVLR	17		
						123-142/GCDPESGSLVAETLTTLVLR	89		
						123-142/GCDPESGSLVAETLTTLVLR	69		
						123-142/GQGGFGGAR	27		
						123-142/GQGGFGGAR	46		
						152-163/GERPAAPEFPDR	48		
						152-168/GERPAAPEFPDRHPDAR	15		
						152-168/GERPAAPEFPDRHPDAR	43		
						169-174/IDMPTR	42		
						169-174/IDMPTR	41		
						175-182/EDQALIYR	62		
						220-239/ALVAELGGGVAANITSIAAR	84		

					220–239/ALVAELGGGVAANITSIAAR	3			
					220–239/ALVAELGGGVAANITSIAAR	74			
					240–256/FTKPVFPGETLSTVIWR	35			
					240–256/FTKPVFPGETLSTVIWR	84			
					266–277/TEVAGSDGAEAR	65			
					278–290/VVLDGAVEYVAG	25			
18	15609760	Hypothetical protein Rv2623	Unknown (10)	2650	TB31.7	2–25/SSGNSSLGIIVGIDSPAAQVAVR	10	975	79
						2–25/SSGNSSLGIIVGIDSPAAQVAVR	84		
						30–34/DAELR	16		
						35–62/KIPLTLVHAVSPEVATWLEVPLPPGVLR	30		
						35–62/KIPLTLVHAVSPEVATWLEVPLPPGVLR	66		
						63–69/WQQDHGR	26		
						63–69/WQQDHGR	18		
						70–77/HLIDDALK	21		
						70–77/HLIDDALK	62		
						70–77/HLIDDALK	58		
						70–77/HLIDDALK	65		
						78–85/VVEQASLR	61		
						86–109/AGPPTVHSEIVPAAAVPTLVDMSK	41		
						86–109/AGPPTVHSEIVPAAAVPTLVDMSK	45		
						86–109/AGPPTVHSEIVPAAAVPTLVDMSK	67		
						86–109/AGPPTVHSEIVPAAAVPTLVDMSK	20		
						86–109/AGPPTVHSEIVPAAAVPTLVDMSK	21		
						86–109/AGPPTVHSEIVPAAAVPTLVDMSK	54		
						110–123/DAVLMVVGCLGSGR	49		
						110–123/DAVLMVVGCLGSGR	72		
						110–123/DAVLMVVGCLGSGR	36		
						110–123/DAVLMVVGCLGSGR	45		
						110–123/DAVLMVVGCLGSGR	76		
						128–138/LLGSVSSGLLR	60		
						128–138/LLGSVSSGLLR	85		
						139–185/HAHCPVVIHDEDSVMPHPQQAPVLVGDGSSAS	62		
						ELATAIAFDEASR			
						139–185/HAHCPVVIHDEDSVMPHPQQAPVLVGDGSSAS	50		
						ELATAIAFDEASR			
						224–230/LAGWQER	47		
						231–238/YPNVAITR	61		
						231–238/YPNVAITR	56		
						248–252/QLVQR	39		
						253–264/SEEAQLVVVGSR	62		
						253–264/SEEAQLVVVGSR	72		
						253–264/SEEAQLVVVGSR	78		
						267–286/GGYAGMLVGSVGETVAQLAR	74		
						267–286/GGYAGMLVGSVGETVAQLAR	95		
						267–286/GGYAGMLVGSVGETVAQLAR	83		
						287–293/TPVIVAR	48		
19	31794065	Elongation factor Ts	GTP complex up to the GTP hydrolysis stage on the ribosome (2)	2910c	<i>tsf</i>	2–10/ANFTAADVK	40	406	45
						25–36/NALAETDGDGDFK	21		
						116–127/TVEQAIAELSAK	71		
						137–150/VAIFDGTVEAYLHR	47		
						152–166/SADLPPAVGVLEAYR	50		
						167–184/GDDAAAAHAVALQIAALR	81		
						211–221/AEGKPEQALPK	21		
						211–221/AEGKPEQALPK	12		

						227-232/LNGFFK	19		
						233-247/DAVLLEQASVSDNKK	25		
						251-262/ALLDVAGVMVTR	35		
20	15607779	Transcription antitermination protein NusG	Influences transcription termination and antitermination. Acts as a component of the transcription complex (2)	0688	<i>nusG</i>	44-61/SKPGDWYVVHSYAGYENK	25	416	30
						64-69/ANLETR	20		
						103-112/KVLPGYILVR	42		
						104-112/VLPGYILVR	67		
						113-124/MDLTDDSWAAVR	97		
						150-154/FLLPR	41		
						217-224/VLVSIFGR	67		
						225-237/ETPVELTFGQVSK	63		
21	31794206	Electron transfer flavoprotein subunit alpha	The electron transfer flavoprotein serves as a specific electron acceptor for other dehydrogenases (7)	3051c	<i>fixB</i>	2-16/AEVLVLEHAEGALK	54	788	80
						2-16/AEVLVLEHAEGALK	51		
						2-16/AEVLVLEHAEGALK	59		
						2-17/AEVLVLEHAEGALKK	46		
						2-17/AEVLVLEHAEGALKK	4		
						2-17/AEVLVLEHAEGALKK	63		
						17-27/KVSAELITAAR	91		
						18-27/VSaelitaar	72		
						28-51/ALGEPAAVVVGVPGTAAPLVDGLK	57		
						58-68/IYVAESDLVDK	13		
						69-99/YLITPAVDVLAGLAESSAPAGVLIATADGK	75		
						69-99/YLITPAVDVLAGLAESSAPAGVLIATADGK	55		
						109-121/IGSGLLDVVDVR	56		
						109-121/IGSGLLDVVDVR	99		
						152-180/AGAVEAEPAAAGAGEQVSVEVAAAENAAR	45		
						185-206/EPAVAGDRPELTEATIVVAGGR	45		
						207-232/GVGSaENFSVVEALADSLGAAVGASR	17		
						207-232/GVGSaENFSVVEALADSLGAAVGASR	99		
						207-232/GVGSaENFSVVEALADSLGAAVGASR	120		
						233-251/AAVDSGYYPGQFQVGQTGK	71		
						233-251/AAVDSGYYPGQFQVGQTGK	60		
						252-270/TVSPQLYIALGISGAIQHR	25		
						252-270/TVSPQLYIALGISGAIQHR	15		
						252-270/TVSPQLYIALGISGAIQHR	114		
						278-304/TIVAVNKDEEAPIFEIADYGVVGD LFK	14		
						278-304/TIVAVNKDEEAPIFEIADYGVVGD LFK	46		
						305-314/VAPQLTEVIK	56		
22	15607971	Hypothetical protein Rv0831c	Unknown (10)	0884c	---	25-37/HPTTDSL TESANR	20	708	61
						25-37/HPTTDSL TESANR	66		
						25-37/HPTTDSL TESANR	44		
						41-51/HLLINDLPIER	68		
						41-51/HLLINDLPIER	50		
						52-73/QAQDVSWGMTAPGGAPT VADR	23		
						52-73/QAQDVSWGMTAPGGAPT VADR	26		
						81-89/DNTTAASLK	41		
						90-101/NQAI VVETTAYR	62		
						102-112/SFEAFTDVVMR	63		
						102-112/SFEAFTDVVMR	55		
						118-128/AQVSSIVGLER	62		
						133-138/FVLEIR	32		
						139-146/VPAGVDGR	57		
						147-162/ITWSNWIDEQLLGPQR	38		
						147-162/ITWSNWIDEQLLGPQR	68		

					163-180/FTPGGLVLTEWQGAAYR	76			
					163-180/FTPGGLVLTEWQGAAYR	68			
					163-186/FTPGGLVLTEWQGAAYRELQPGK	12			
					192-207/YGPGMGQALDPNYHLR	9			
					192-207/YGPGMGQALDPNYHLR	45			
					192-207/YGPGMGQALDPNYHLR	23			
					264-270/LKDELLR	41			
23	15607955	Thiosulfate sulfurtransferase CysA2	May be a sulfotransferase involved in the formation of thiosulfate (7)	0867c	<i>cysA2</i>	4-21/CDVLSADWAESNLHAPK	37	722	58
					37-44/DHIAGAIK	61			
					49-56/TDLQDPVK	39			
					49-57/TDLQDPVKR	37			
					49-57/TDLQDPVKR	41			
					49-57/TDLQDPVKR	43			
					58-67/DFVDAQQFSK	77			
					68-72/LLSER	29			
					100-105/LYGHEK	24			
					108-113/LLDGGR	38			
					146-158/AFRDEVLAAINVK	46			
					149-158/DEVLAAINVK	40			
					165-172/SPDEFSGK	48			
					173-200/ILAPAHLPQEQSQRPGHIPGAINVPWSR	27			
					201-209/AANEDGTFK	43			
					210-216/SDEELAK	28			
					217-227/LYADAGLDNSK	67			
					228-234/ETIAYCR	14			
					239-247/SSHTWFLR	33			
					248-256/ELLGHQNVK	44			
24	31794065	Elongation factor Ts	Associates with the EF-TU-GDP complex and induces the exchange of GDP to GTP, it remains bound to the aminoacyl-tRNA (2)	2910c	<i>tsf</i>	1-10/MANFTAADVK	12	777	61
					2-10/ANFTAADVK	45			
					12-24/LRELTGAGMLACK	3			
					25-36/NALAEITDGDFFDK	53			
					56-65/ATAEGLVAAK	40			
					56-65/ATAEGLVAAK	71			
					116-127/TVEQAI AELSAK	75			
					137-150/VAIFDGTVEAYLHR	52			
					137-150/VAIFDGTVEAYLHR	44			
					152-166/SADLPPAVGV LVEYR	49			
					152-166/SADLPPAVGV LVEYR	79			
					167-184/GDDAAAAHAVALQIAALR	74			
					191-202/DDVPEDIVASER	47			
					203-210/RIAEETAR	25			
					204-210/IAEETAR	20			
					211-221/AEGKPEQALPK	32			
					211-221/AEGKPEQALPK	33			
					227-232/LNGFFK	19			
					227-232/LNGFFK	44			
					233-246/DAVLLEQASVSDNK	20			
					233-247/DAVLLEQASVSDNKK	34			
					233-247/DAVLLEQASVSDNKK	45			
					251-262/ALLDVAGVMVTR	86			
25	15608092	Succinyl-CoA synthetase subunit alpha	Involved in tricarboxylic acid cycle (7)	1006	<i>sucD</i>	14-30/VIVQGITGSEATVHTAR	86	671	60
					14-30/VIVQGITGSEATVHTAR	74			
					34-45/AGTQIVGGVNAR	37			
					46-56/KAGTTVTHEDK	20			

					63-75/LPVFGSVAEAMEK	94			
					63-75/LPVFGSVAEAMEK	73			
					76-91/TGADVSIIFVPPTFAK	36			
					133-167/IIGPNCPGIISPGQSLAGITPANITGPGPIGLVSK	55			
					215-229/LIVMIGEIGGDAEER	42			
					215-229/LIVMIGEIGGDAEER	50			
					230-235/AADFIK	36			
					236-255/TNVSKPVVGYVAGFTAPEGK	33			
					236-255/TNVSKPVVGYVAGFTAPEGK	57			
					256-274/TMGHAGAIVSGSSGTAAAK	26			
					256-274/TMGHAGAIVSGSSGTAAAK	55			
					256-274/TMGHAGAIVSGSSGTAAAK	11			
					256-274/TMGHAGAIVSGSSGTAAAK	62			
					275-284/QEALAAAGVK	77			
					288-297/TPSATAALAR	65			
26	15607717	Hypothetical protein Rv0577	Unknown (10)	0622	<i>TB27.3 (cfp30B)</i>	9-26/QGTPNWVDLQTTDQSAK	16	236	37
						93-112/VVPGGGQVMMPAFDIGDAGR	7		
						93-112/VVPGGGQVMMPAFDIGDAGR	35		
						113-131/MSFITDPTGAAVGLWQANR	37		
						113-131/MSFITDPTGAAVGLWQANR	11		
						113-131/MSFITDPTGAAVGLWQANR	15		
						221-240/AAAAGGQVIAEPADIPSVGR	77		
						241-261/FAVLSDPQGAIFSVLKPAPQQ	24		
						241-261/FAVLSDPQGAIFSVLKPAPQQ	70		
27	15607581	Chaperonin GroEL	Prevents misfolding and promotes the refolding and proper assembly of unfolded polypeptides generated under stress conditions (0)	0479	<i>groEL2</i>	4-13/TIAYDEEARR	42	872	43
						4-13/TIAYDEEARR	23		
						18-27/GLNALADAVK	50		
						18-27/GLNALADAVK	56		
						28-33/VTLGPK	26		
						36-41/NVVLEK	31		
						42-57/KWGAPTITNDGVSIK	37		
						42-57/KWGAPTITNDGVSIK	44		
						43-57/WGAPTITNDGVSIK	67		
						58-67/EIELEDPEYK	40		
						68-74/IGAELVK	59		
						79-100/KTDDVAGDGTTTATVLAQALVR	88		
						79-100/KTDDVAGDGTTTATVLAQALVR	34		
						79-100/KTDDVAGDGTTTATVLAQALVR	44		
						80-100/TDDVAGDGTTTATVLAQALVR	73		
						80-100/TDDVAGDGTTTATVLAQALVR	56		
						126-132/VTETLLK	24		
						141-166/EQIAATAAISAGDQSIGDLIAEAMDK	34		
						196-208/GYISGYFVTDPER	67		
						209-224/QEAVLEDPYILLVSSK	58		
						379-388/AGAATEVELK	25		
						429-441/LEGDEATGANIVK	34		
						442-449/VALEAPLK	45		
						468-495/NLPAGHGLNAQTGVYEDLLAAGVADPVK	49		
						499-524/SALQNAASIAGLFLTTEAVVADKPEK	43		
28	15608210	Enoyl-CoA hydratase	Could possibly oxidizes fatty acids using specific components (1)	1128c	<i>echA8</i>	2-10/TYETILVER	71	637	49
						60-68/AFAAGADIK	62		
						127-133/FGQPEIK	37		
						134-145/LGVLPGMGGSQR	50		
						134-145/LGVLPGMGGSQR	45		

						155-163/AMDLILTGR	56		
						155-163/AMDLILTGR	45		
						164-172/TMDAAEAER	34		
						179-190/VVPADDLLTEAR	79		
						191-205/ATATTISQMSASAAR	70		
						191-205/ATATTISQMSASAAR	76		
						214-227/AFESSLSEGLLYER	59		
						214-227/AFESSLSEGLLYER	103		
						228-249/RLFHSAFATEDQSEGMAAFIEK	11		
						229-249/LFHSAFATEDQSEGMAAFIEK	53		
						229-249/LFHSAFATEDQSEGMAAFIEK	26		
						250-257/RAPQFTHR	35		
						250-257/RAPQFTHR	29		
						251-257/APQFTHR	48		
29	48828	Antigen 85C	Proteins of the antigen 85 complex are responsible for the high affinity of mycobacteria to fibronectin (1)	0163c	<i>fbpC</i>	70-87/VQFQGGGPHAVYLLDGLR	20	242	24
						70-87/VQFQGGGPHAVYLLDGLR	45		
						141-147/WETFLTR	36		
						235-243/NDPMVQIPR	56		
						251-271/IWVYCGNGTPSDLGGDNIPAK	19		
						272-279/FLEGLTLR	54		
						320-340/ADIQHVLNGATPPAAPAAPAA	35		
30	254233296	Secreted antigen 85-A fbpA	Involved in cell wall mycoloylation. Proteins of the antigen 85 complex are responsible for the high affinity of mycobacteria to fibronectin (1)	3866c	<i>fbpA</i>	84-103/VQFQSGGANSPALYLLDGLR	72	347	28
						157-173/WETFLTSELPGWLQANR	29		
						236-243/ASDMWGPK	48		
						244-250/EDPAWQR	31		
						251-259/NDPLLNVGK	30		
						260-266/LIANNTR	34		
						260-266/LIANNTR	30		
						267-287/VWVYCGNGKPSDLGGNNLPAK	65		
						288-294/FLEGFVR	24		
						288-294/FLEGFVR	37		
						356-363/LAPNRHNI	10		
31	15610283	NADH dehydrogenase subunit C	Involved in aerobic/anaerobic respiration (7)	3170	<i>nuoC</i>	67-72/LAEALR	43	389	44
						76-84/VEFEDAVEK	54		
						85-97/VVYRDELTLHVR	16		
						90-97/DELTLHVR	42		
						108-115/LRDEPELR	35		
						134-148/ELHAVYPLQSITHNR	53		
						179-196/ETYDFFGIIFDGHPALTR	40		
						179-196/ETYDFFGIIFDGHPALTR	40		
						197-209/IEMPDDWQGHQPQR	23		
						211-222/DYPLGGIPVEYK	52		
						223-232/GAQIPPPDER	31		
32	15609565	Alkyl hydroperoxide reductase subunit C	Involved in oxidative stress response (0)	2447	<i>ahpC</i>	2-25/PLLTIGDQFPAYQLTALIGDLSK	15	604	81
						30-46/QPGDYFTTITSDEHPGK	37		
						30-46/QPGDYFTTITSDEHPGK	56		
						49-55/VVFFWPK	42		
						56-70/DFTFVCPTIEIAAFSK	45		
						56-70/DFTFVCPTIEIAAFSK	13		
						71-78/LNDEFEDR	53		
						105-115/TLPFPMLSDIK	62		
						105-115/TLPFPMLSDIK	40		
						117-133/ELSQAAGVLNADGVADR	30		
						117-133/ELSQAAGVLNADGVADR	50		

						134–156/VTFIVDPNNEIQFVSATAGSVGR	77		
						157–163/NVDEVLK	30		
						164–179/VLDALQSDELCAACNWR	48		
						164–179/VLDALQSDELCAACNWR	75		
						180–192/KGDPTLDAGELLK	61		
						181–192/GDPTLDAGELLK	43		
33	126030155	Chain A, The 3.0 A Resolution Structure Of Caseinolytic Clp Protease 1	CLP may be responsible for a fairly general and central housekeeping function rather than for the degradation of specific substrates (7)	2481c	<i>clpP</i>	10–24/SNSQGLSLTDSVYER	53	373	49
						25–29/LLSER	28		
						30–44/IIFLGSEVNDEIANR	87		
						113–120/RYPALPHAR	27		
						114–120/YALPHAR	52		
						121–148/ILMHQPLGGVTGSAADIAIQAEQFAVIK	26		
						121–148/ILMHQPLGGVTGSAADIAIQAEQFAVIK	34		
						154–165/LNAEFTGQPIER	75		
						175–193/WFTAAEALEYGFVDHIITR	17		
34	15607196	Single-stranded DNA-binding protein	This protein is essential for replication of the chromosome. It is also involved in DNA recombination and repair (2)	0085	<i>ssb</i>	21–38/FTPSGAAVANFTVASTPR	33	366	39
						21–38/FTPSGAAVANFTVASTPR	92		
						49–56/DGEALFLR	74		
						49–56/DGEALFLR	54		
						57–61/CNIWR	18		
						62–73/EAAENVAESLTR	29		
						62–73/EAAENVAESLTR	63		
						77–82/VIVSGR	46		
						97–111/TVIEVEVDEIGPSLR	76		
35	15607792	50S ribosomal protein L7/L12	Involved in translation mechanisms: seems to be the binding site for several of the factors involved in protein synthesis and appear to be essential for accurate translation (2)	0701	<i>rplL</i>	4–14/LSTDELLDAFK	28	367	42
						4–14/LSTDELLDAFK	45		
						4–14/LSTDELLDAFK	89		
						15–26/EMTLELSDFVK	68		
						15–26/EMTLELSDFVK	47		
						74–79/KIGVIK	36		
						95–107/DLVDGAPKPLEK	37		
						95–107/DLVDGAPKPLEK	22		
						118–130/AKLEAAGATVTVK	69		
						118–130/AKLEAAGATVTVK	50		
						120–130/LEAAGATVTVK	67		
36	15609168	Heat shock protein hspX	Stress protein induced by anoxia. Has a proposed role in maintenance of long term viability during latent, asymptomatic infections, and a proposed role in replication during initial infection (0)	2050c	<i>hspX</i>	42–47/LEDEMK	26	288	56
						55–71/AELPGVDPDKDVIDMVR	7		
						55–71/AELPGVDPDKDVIDMVR	11		
						65–71/DVDIMVR	8		
						72–78/DGQLTIK	23		
						72–78/DGQLTIK	47		
						91–100/SEFAYGSFVR	35		
						101–114/TVSLPVGAEDEDDIK	59		
						101–119/TVSLPVGAEDEDDIKATYDK	28		
						120–136/GILTVSVAVSEGKPTK	25		
						120–136/GILTVSVAVSEGKPTK	38		
						137–141/HIQIR	37		
						137–141/HIQIR	31		
37	15609582	Nucleoside diphosphate kinase	Major role in the synthesis of nucleoside triphosphates other than ATP (7)	2465c	<i>ndkA</i>	5–16/TLVLIKPDGIER	56	304	52
						5–16/TLVLIKPDGIER	47		
						17–25/QLIGEIISR	57		
						30–39/GLTIAALQLR	42		
						87–104/QLAGGTDPVQAAAPGTIR	35		
						105–127/GDFALETQFNLVHGSDSAESAQR	120		

* FC= functional category. 0: virulence, detoxification, adaptation; 1: lipid metabolism; 2: information pathways; 3: cell wall and cell processes; 4: stable RNAs; 5: insertion sequences and phages; 6: Pe/PPE; 7: intermediary metabolism and respiration; 8: unknown; 9: regulatory protein; 10: conserved hypothetical (from BCGList World-Wide Web Server <http://genolist.pasteur.fr/BCGList/>).

Ω Annotation from BCGList World-Wide Web Server (<http://genolist.pasteur.fr/BCGList/>).

& According to Mascot Search Results, protein scores >25 are significant ($p < 0.05$)