



Universidad Autónoma de Querétaro
Facultad de Ciencias Naturales

Filogenia y genómica comparativa de cepas de
Mycobacterium bovis de México

Tesis

Que como parte de los requisitos para obtener el Grado de
Doctora en Ciencias Biológicas

Presenta

Claudia Angélica Perea Razo

Dirigido por:

Dr. Germinal Jorge Cantó Alarcón

Co-director:

Dr. Sergio Iván Román Ponce

Febrero 2020



Universidad Autónoma de Querétaro
Facultad de Ciencias Naturales
Doctorado en Ciencias Biológicas

Filogenia y genómica comparativa de cepas de *Mycobacterium bovis* de
México

Que como parte de los requisitos para obtener el Grado de
Doctora en Ciencias Biológicas

Presenta

Claudia Angélica Perea Razo

Dirigido por:

Dr. Germinal Jorge Cantó Alarcón

Co-dirigido por:

Dr. Sergio Iván Román Ponce

Dr. Germinal Jorge Cantó Alarcón
Presidente

Dr. Sergio Iván Román Ponce
Secretario

Dr. Feliciano Milián Suazo
Vocal

Dra. Elba Rodríguez Hernández
Vocal

Dr. Carlos Agustín Vega y Murguía
Vocal

Centro Universitario, Querétaro, Qro.
Febrero 2020
México

DEDICATORIAS

A mi madre por creer en mí, siempre, de manera incondicional. Por echarme porras cuando me invadía la duda. Por consentirme. Por su confianza infinita. Por hacerme ver la realidad. Y más que nada, por su excepcional ejemplo.

A mi padre, por el gran ser humano que fue, su memoria me impulsa a valorar cada logro, fracaso, felicidad, tristeza, y todo aquello que la vida me pone en el camino.

A la cabrita loca por escuchar mis quejas, frustraciones, decepciones, dudas, y ante todo responder en positivo y darme ánimos para transformar cualquier inconveniente en el empuje de seguir adelante.

A Pau y Sofi, por recordarme siempre lo que es realmente importante.

A mi Gigin, por el pequeño gigante maestro que fue, sin miedo a nada. Y a Roxy, por ser mi apoyo incondicional, desvelándose conmigo y levantar mis ánimos solo con su mirada.

A mí, por mi esfuerzo, por mi paciencia, por mi determinación, por mi integridad, por mi creatividad, por mi ímpetu y por mi crecimiento.

AGRADECIMIENTOS

En primera instancia, le merezco agradecimiento a mi universidad, ya que ha sido casa de mi formación como profesional desde la licenciatura, a través de la maestría y por último en el doctorado. Aquí he encontrado personas que se convirtieron en verdaderos modelos a seguir, incluyendo profesores, compañeros, técnicos, colegas, administrativos, etc. Me ha retado tanto profesionalmente, como personalmente, contribuyendo de manera significativa a mi futuro. He tenido la fortuna de ver la transformación que ha vivido la facultad, y he tenido el honor de poder contribuir a su desarrollo. Gracias por educarme en la verdad y en el honor.

De la misma manera, le expongo gran agradecimiento al Dr. Cantó, por aventurarse con mi terquedad y ayudarme a darle forma a mis ideas y corregir mis tartamudeos. Agradezco la libertad que me dio para explorar mi potencial, pues así he forjado la seguridad en mí capacidad para desempeñarme más allá de esta casa de estudios. Ha sido un pilar en mi formación.

Gracias sinceras al Dr. Milián por su apoyo durante todo este proceso, por proporcionarme puentes de comunicación que ahora son esenciales para lo que sigue después.

Agradezco igualmente a mis sinodales: Dra. Elba Rodríguez, Dr. Sergio Román y Dr. Carlos Vega, por adentrarse en este proyecto conmigo, que, aunque a veces parecía no tener pies o cabeza, sus aportaciones lograron guiarme y conducir el proyecto hasta su resolución.

A la Dra. Suelee Robbe-Austerman y Tod Stuber por su gran generosidad en cuanto a apoyo técnico y capacitación, por su confianza, y por su amable hospitalidad durante mi estancia en su laboratorio.

Gracias al Dr. Eduardo Ruíz, porque sin sus talleres y cursos de filogenética, filogeografía y evolución molecular, no hubiera sido posible mi avance en mi proyecto. Al Dr. Víctor González, por permitirme acudir a su unidad de investigación para capacitarme en los procesos bioinformáticos que utilicé en mi proyecto, así como a Luis Lozano quien personalmente me instruyó en todos los programas, ambos con gran apertura y amabilidad. Al Wellcome Trust Sanger Institute por la beca otorgada que me permitió asistir a un curso internacional de análisis de genomas de patógenos.

Por supuesto, a CONACyT, por la beca otorgada para realizar mis estudios de doctorado.

Gracias sinceras a todo el apoyo técnico, administrativo, emocional, y/o económico por parte de compañeros, maestros, etc., su apoyo fue verdaderamente valioso.

ÍNDICE

Dedicatorias.....	i
Agradecimientos.....	ii
Índice.....	iv
Índice de Cuadros.....	vi
Índice de Figuras.....	vii
Resumen.....	ix
Abstract.....	x
I. INTRODUCCIÓN.....	1
II. ANTECEDENTES.....	3
2.2. Situación actual de la Tuberculosis Bovina en México.....	3
2.3. Epidemiología de <i>Mycobacterium bovis</i>	7
2.4. Genotipificación de <i>Mycobacterium bovis</i>	10
2.5. El genoma de <i>Mycobacterium bovis</i>	13
2.6. Evolución de <i>Mycobacterium bovis</i>	17
III. HIPÓTESIS.....	22
IV. OBJETIVOS.....	23
4.2. Objetivo general.....	23
4.3. Objetivos específicos.....	23
V. METODOLOGÍA.....	24
5.2. Cepas bacterianas.....	24
5.3. Aislamiento de <i>Mycobacterium bovis</i>	25
5.4. Espoligotipificación.....	26
5.5. Secuenciación genómica (WGS).....	26
5.4.2. Revisión de calidad de las secuencias.....	26
5.6. Identificación de los polimorfismos de un solo nucleótido (SNP).....	27
5.7. Análisis filogenético.....	28
5.8. Análisis del efecto de los SNP sobre la función proteica.....	28
5.9. Ensamblaje de los genomas.....	29

5.9.2. Revisión de calidad de los ensamblados.....	29
5.10. Anotación funcional.....	30
5.11. Análisis de pangenoma.....	31
5.12. Análisis estadísticos.....	31
VI. RESULTADOS Y DISCUSIÓN.....	33
6.2. Espoligotipificación y secuenciación genómica.....	33
6.3. Identificación y clasificación de las variantes en el genoma de <i>M. bovis</i>	40
6.4. Análisis del efecto de las variantes sobre la expresión génica.....	43
6.5. Análisis de pangenoma.....	52
VII. CONCLUSIONES.....	56
VIII. BIBLIOGRAFÍA.....	58

ÍNDICE DE CUADROS

Cuadro		Página
1	Prevalencia de tuberculosis bovina por estado a diciembre 2017.....	5
2	Genes definidos como esenciales <i>in vivo</i> , que contienen SNP no sinónimos en todas las cepas de <i>M. bovis</i> BCG.....	12
3	Comparación del genoma de <i>M. bovis</i> AF2122/97 con <i>M. tuberculosis</i> H37R y CDC1551.....	14
4	Definición y descripción de los Complejos Clonales de <i>M. bovis</i> Af1, Af2 y Eu1.....	18
5	Características de las cepas de <i>M. bovis</i> utilizadas en este estudio.....	24
6	Distribución de los grupos genéticos por estado y número de cepas correspondientes.....	36
7	Distribución de los genes que forman parte del pangenoma de <i>M. bovis</i> de México.....	51

ÍNDICE DE FIGURAS

Figura		Página
1	Situación actual de la tuberculosis bovina en México (adaptado de SENASICA, 2019).....	5
2	Representación esquemática del análisis evolutivo (espiligoforest) de los 39 espiligotipos encontrados.....	35
3	A. Árbol filogenético elaborado con base en las secuencias concatenadas de SNP obtenidos a partir de la secuenciación genómica de 320 cepas de <i>M. bovis</i> de México. B. Acercamiento a los Grupos H, I, J, mostrando que la mayoría de las cepas son de Baja California.....	37-38
4	Promedio de variantes y tasa de variación por espiligotipo.....	41
5	Clasificación de las variantes por tipo.....	42
6	Distribución de las variantes por tipo de secuencia afectada...	44
7	Proporción de las variantes por tipo de secuencia afectada y su posible efecto.....	46
8	Clasificación de las variantes por su efecto sobre el amino ácido: no sinónimas, (missense), sin sentido (nonsense), sinónimas (silent).....	48
9	Clasificación de las variantes por su efecto sobre la función proteica.....	49
10	Cantidad de variantes por el tipo de cambio de bases.....	50
11	Tasa de transiciones y transversiones totales.....	51
12	Frecuencia de genes y número de genomas.....	53
13	Esquina superior izquierda: Número de genes nuevos con respecto al pangenoma; Esquina superior derecha: Número de genes conservados con respecto al pangenoma; Esquina	53

inferior izquierda: Número total de genes que conforman el pangenoma; Esquina inferior derecha: Número de genes únicos encontrados en el pangenoma.....

14 Análisis de componentes principales y mapa de calor representan la similitud genética de los genomas las cepas de *M. bovis* de México..... 55

Dirección General de Bibliotecas UAQ

RESUMEN

Mycobacterium bovis es el agente causal de la tuberculosis bovina y es capaz de infectar una amplia variedad de huéspedes mamíferos incluyendo al humano. En México, la tuberculosis bovina es endémica y representa un importante problema para el sector pecuario, así como para la salud pública. En este estudio se analizaron 320 aislados de *M. bovis* obtenidos a partir de ganado bovino (cárnico y lechero) y humanos, de distintos estados del país: Baja California, Coahuila, Jalisco, Guanajuato, Aguascalientes, Querétaro, Hidalgo, Estado de México y Veracruz. Se realizó la genotipificación de los aislados mediante espoligotipificación y secuenciación genómica completa para evaluar la diversidad genética con respecto al huésped, fin zootécnico y región geográfica, con el fin de determinar características genómicas específicas. Se determinaron 39 espoligotipos, los más frecuentes fueron SB0673, SB0971, SB0140 y SB0145; aunque el origen y huésped no influyeron sobre la diversidad genética, las cepas de Baja California obtuvieron perfiles genéticos exclusivos de esa región. A partir de los perfiles de polimorfismos de un solo nucleótido (SNP) se obtuvieron 12 grupos genéticos, los cuales mostraron concordancia en cuanto a las relaciones genéticas obtenidas a partir de la espoligotipificación, validando de esta manera el uso actual de esta técnica frente a las tecnologías más modernas para el monitoreo epidemiológico de la enfermedad. Se observó un pangenoma abierto, con un total de 3,687 genes esenciales. La densidad promedio de SNP fue de 1 por cada 10,850 pares de bases, los cuales principalmente cayeron en regiones río arriba y río abajo de los genes, y en segundo lugar dentro de exones. Así mismo, las mutaciones no sinónimas superaron a las mutaciones sinónimas, en su mayoría con un efecto predicho “moderado” sobre la función proteica. La diversidad genética de *M. bovis* en México puede deberse a mutaciones puntuales que ha obtenido a través del tiempo y de un evento de transmisión a otro. Es necesario desarrollar y reforzar estrategias que eviten la continua diseminación del microorganismo para impulsar de manera más efectiva los esfuerzos de erradicación.

Palabras clave: SNP, *Mycobacterium bovis*, secuenciación de genoma completo, tuberculosis bovina, espoligotipificación.

ABSTRACT

Mycobacterium bovis is the causal agent of bovine tuberculosis, though it also causes disease in humans and is able to infect a wide range of mammal hosts. In Mexico, bovine tuberculosis is endemic and it represents an important problem for the livestock industry, as well as public health. This study included 320 *M. bovis* isolates from cattle (beef and dairy) and humans, from different states of the country: Baja California, Coahuila, Jalisco, Guanajuato, Aguascalientes, Querétaro, Hidalgo, Estado de México and Veracruz. Genotyping was performed through spoligotyping and whole genome sequencing to evaluate genetic diversity with respect to host, zotechnical purpose, and geographical region, in order to determine specific genomic characteristics. Thirty-nine spoligotypes were obtained, the most frequent being SB0673, SB0971, SB0140, and SB0145; even though origin and host had no influence on genetic diversity, the strains from Baja California showed genetic profiles exclusive of that region. Single nucleotide polymorphism (SNP) profiles grouped the strains into 12 main clades that were in agreement with the genetic relationships obtained through spoligotyping, which makes the use of this technique still valid against more modern technologies for the epidemiological surveillance of the disease. An open pangenome was seen, with a core genome of 3,687 genes. Average SNP density was 1 every 10,850 base pairs, and most SNP fell within regions upstream and downstream from genes, followed by within exons. The amount of non-synonymous mutations surpassed synonymous mutations, and most of them had a predicted “moderate” effect on protein function. *M. bovis* genetic diversity in Mexico has been shaped in part by point mutations that the bacteria has gained through time and by each transmission event. It is necessary to develop and strengthen strategies that prevent the continuous dissemination of the microorganism in order to boost efforts for the eradication of the disease.

Key words: SNP, *Mycobacterium bovis*, whole genome sequencing, bovine tuberculosis, spoligotyping.

I. INTRODUCCIÓN

Mycobacterium bovis es la causa más común de tuberculosis en el ganado bovino y es uno de los miembros del Complejo *Mycobacterium tuberculosis* (MTBC), el cual incluye varias especies y subespecies que ocasionan una patología similar en una gran variedad de mamíferos (Smith et al., 2006). El huésped preferido de *M. bovis* es el ganado bovino, aunque ha sido aislado frecuentemente del humano y otros mamíferos (Smith et al., 2006). Consecuentemente, en la actividad pecuaria, ésta enfermedad ocasiona pérdidas económicas significantes, ya sea por la muerte de los animales y las mermas en producción, como por el decomiso de órganos y canales en rastro, el desecho involuntario de ganado, las limitantes de movilización, comercialización y el costo de programas de control y erradicación (Milián et al., 2013). Además, puesto que es una zoonosis importante, *M. bovis* representa un riesgo importante en salud pública, ya que la Organización Mundial de la Salud reporta a la tuberculosis como la segunda causa de muerte por un agente infeccioso a nivel mundial (OMS, 2015).

Comparado a otras especies bacterianas, *Mycobacterium bovis* contiene relativamente poca diversidad genética; sin embargo, existe evidencia creciente que la variación que existe entre cepas puede ser biológicamente significativa. Las cepas de campo de *Mycobacterium bovis* tienen perfiles de expresión génica variados, así como distintos genes eliminados de su cromosoma (Garnier et al., 2003). Algunos genotipos de *M. bovis* que están genéticamente relacionados incluso han aparentado estar asociados a regiones geográficas específicas. Esta observación puede indicar diferencias biológicas subyacentes entre las cepas de campo, incluyendo la adaptación a una serie de huéspedes específicos o respuesta a la vacunación (Muller et al., 2008).

El grado de enfermedad está influenciado por varios factores que pueden no estar asociados al patógeno, como el antecedente genético del huésped infectado, enfermedades concomitantes y estado inmunológico, entre otras (Arce et al., 2008). Identificar factores de virulencia relacionados a la micobacteria y asociarlos a ciertos

fenotipos requiere de un método para categorizar o clasificar a las cepas de *M. bovis* dentro de grupos que compartan de manera significativa características genotípicas y fenotípicas. Las técnicas de filogenética molecular facilitan este tipo de estudios al agrupar cepas genéticamente relacionadas y de esta manera disponer de un marco evolutivo de referencia para investigar polimorfismos que conlleven una potencial relevancia biológica (Zenteno-Cuevas et al., 2014).

El genoma de *M. bovis* está altamente conservado, lo que dificulta el análisis filogenético mediante la secuenciación de loci múltiples de genes constitutivos, por lo que lo hace impráctico y poco informativo (Wright et al., 2013). En lugar de esto, *M. bovis* ha sido genotipificado mediante la identificación de elementos de inserción, como la IS6110, elementos génicos repetitivos en las regiones de repetición directa (espoligotipificación), los microsatelites variables (análisis de las unidades repetitivas micobacterianas, o MIRU), entre otros (Fernandes-Ramos et al., 2014). Estas técnicas han sido exitosas al identificar grupos de aislados que tienen un ancestro en común. Sin embargo, estos métodos aún no han sido evaluados contra algún estándar de oro único, lo cual hace difícil su significancia en cuanto a la información filogenética que pueden proporcionar. Estos métodos tienen además la desventaja de que la tasa de sustitución de cada marcador filogenético no es uniforme para todos los marcadores (Smith et al., 2006). La genotipificación con base a SNP promete ser la técnica más precisa y con más alto grado de resolución hasta la fecha, resolviendo los patrones de evolución que ha sufrido *M. bovis* (Smith et al., 2012; García-Pelayo et al., 2009). Por lo tanto, en este estudio se utilizará la secuenciación genómica completa (WGS) de cepas de *M. bovis* de diferentes regiones de México para evaluar su diversidad genética con base en SNP y obtener más claridad en cuanto a la evolución y dispersión de este patógeno en el país.

II. ANTECEDENTES

2.2. Situación actual de la tuberculosis bovina en México

La tuberculosis bovina (TBb) es una enfermedad crónica bacteriana de los animales y el hombre causada por el bacilo *Mycobacterium bovis*. Esta enfermedad se caracteriza por la formación de granulomas nodulares conocidos como tubérculos. Aunque se define como una enfermedad crónica, en ocasiones la tuberculosis tiene un curso agudo y progresivo (OIE, 2012). Las lesiones en el ganado se encuentran principalmente en linfonodos de cabeza, tórax y cavidad abdominal, pulmones, hígado, pleura y peritoneo, aunque cualquier tipo de tejido es susceptible de ser afectado (Gormley y Corner, 2018).

La distribución geográfica de la tuberculosis bovina ha cambiado considerablemente en las últimas décadas. Antes de la introducción de medidas de control y la pasteurización de la leche en los países desarrollados, la tuberculosis estaba ampliamente distribuida en todo el mundo (Sanou et al., 2014). Actualmente, la prevalencia de la TBb en países desarrollados ha sido reducida a niveles mínimos (0.001%); varios países de Europa y América del Norte, así como Australia, están libres de la enfermedad o a punto de erradicarla totalmente del ganado. Sin embargo, el mantenimiento de la infección de *M. bovis* en especies silvestres ha comprometido considerablemente los intentos de erradicación en países como Irlanda, Nueva Zelanda, y los Estados Unidos (Thoen et al., 2009).

En México, el inventario nacional bovino es de casi 35 millones de cabezas, de las que alrededor del 92% es dedicada a la producción de carne y doble propósito y el 8% a la producción de leche (SIAP, 2018). Las razas de carne se explotan en zonas extensivas del norte del país, principalmente razas europeas y en las regiones tropicales, donde se explotan básicamente razas cebuínas. Las razas especializadas en la producción de leche, Holstein y Jersey, se ubican en las regiones templadas del centro del país, como el estado de Jalisco, y en algunas

regiones norteñas, concretamente en la zona de la Laguna, en los límites de los estados de Durango y Coahuila, las cuencas de Tijuana en Baja California, y las cuencas de Delicias y Ciudad Juárez en Chihuahua (Milián et al, 2013).

Antes de 1992 la prevalencia de la tuberculosis bovina era prácticamente desconocida en México; sin embargo, a partir del reconocimiento de regiones en Fase de Erradicación, en 1994, por la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (SAGARPA), como resultado de las primeras actividades de la campaña, en la cual el primer paso fue la tuberculinización de todos los animales mayores de 6 meses de todos los hatos del país, se establecieron regiones o estados con una prevalencia menor al 2% (Milián et al., 2013). Estas regiones de baja prevalencia han disminuido este parámetro de 0.5% a 0.01%, principalmente como respuesta a las exigencias del mercado de exportación de becerros. Actualmente, el 85.77% del territorio nacional tiene una prevalencia menor a 0.5% y se tienen 25 regiones o estados considerados como de baja prevalencia (SENASICA, 2019). Consecuentemente, el territorio nacional se encuentra clasificado en zona de control, donde existe una prevalencia promedio de 2.5% con excepción de las cuencas lecherras, y zona de erradicación, con una prevalencia promedio de 0.5% (Figura 1). Son 17 estados los que se encuentran en control: Aguascalientes, Baja California, Campeche, Ciudad de México, Durango, Guanajuato, Hidalgo, Jalisco, Estado de México, Michoacán, Nayarit, Oaxaca, Puebla, Querétaro, San Luis Potosí, Tlaxcala, y Zacatecas (SENASICA, 2018).



Figura 1. Situación actual de la tuberculosis bovina en México (adaptado de SENASICA, 2019).

Para el monitoreo y vigilancia del estatus sanitario del país en cuanto a la TBb, a través del Servicio Nacional de Sanidad, Inocuidad y Calidad Agroalimentaria (SENASICA), y la Dirección General de Salud Animal, México cuenta con la Campaña Nacional Contra la Tuberculosis Bovina, la cual consiste en el diagnóstico en campo, aplicación de cuarentenas en hatos infectados, eliminación e indemnización de animales reactivos a las pruebas diagnósticas, inspección en rastros para confirmar y detectar nuevos casos, control de la movilización, reconocimiento y protección de regiones de baja prevalencia, certificación de hatos libres de la enfermedad, y seguimiento epidemiológico. Actualmente, cada uno de los estados del país cuenta con una campaña estatal, la cual es dirigida por los Comités de Fomento y Protección Pecuaria Estatales, que son organizaciones estatales de productores, en coordinación y supervisión de las delegaciones de la SAGARPA (Milián et al., 2013).

Cuadro 1. Prevalencia de tuberculosis bovina por estado a diciembre del 2017 (adaptado de SENASICA, 2017).

Estado	Región	Prevalencia (%)
Aguascalientes	A	0.11
	B	4

Baja California	A	0
	B	3.86
Baja California Sur		0.17
Campeche	A	0.08
	B	9.65
Chiapas	A	0.055
	B	0.078
Chihuahua	A	0.049
	B1	1.74
	B2	6.05
	B3	16.1
Coahuila	A	0.04
	B1	4
	B2	1.89
Colima		0.18
Ciudad de México		Desconocida
Durango	A	0.08
	B	Desconocida
Estado de México	A	0.03
	B	Desconocida
Guanajuato	A	0.05
	A1	0.21
	B	Desconocida
Guerrero	A	0.092
	B	0.6
Hidalgo	A	0.006
	A1	0
	B	Desconocida
Jalisco	A1	0.04
	A2	0.13
	A4	0.45
	B1	0.36
	B	Desconocida
Michoacán	A	0.053
	B	Desconocida

Morelos		0.5
Nayarit	A	0.138
	B	1.59
Nuevo León	A	0.01
	B	3.6
Oaxaca	A	0.09
	B	Desconocida
Puebla	A1	0
	A2	0
	B	Desconocida
Querétaro	A	0.03
	B	Desconocida
Quintana Roo		0
San Luis Potosí	A1	0.017
	A2	0.087
	B	0.37
Sinaloa		0.38
Sonora	Norte	0.01
Tabasco	A	0.02
	B	0.27
Tamaulipas		0.017
Tlaxcala		Desconocida
Veracruz	A	0.049
	B	0.2
Yucatán		0.026
Zacatecas	Zac-Jal	0.12
	B	4.35

2.3. Epidemiología de *Mycobacterium bovis*

M. bovis posee uno de los más amplios espectros de hospedadores entre los patógenos conocidos, por lo que la epidemiología de esta infección es muy compleja y engloba interacciones entre el ser humano, animales domésticos y animales salvajes. Entre los posibles hospedadores de esta especie se encuentran el ser

humano, los primates no humanos, vaca, cabra, oveja, cerdo, jabalí, caballo, asno, gato, perro, ciervo, bisonte, búfalo, antílope, oryx, camello, llama, alpaca, reno, tejón, zarigüeya, liebre, conejo, hurón, feneco, elefante, jirafa, rinoceronte, tigre, león, foca, etc. En el ganado bovino, los resultados de infecciones experimentales por las vías respiratoria y digestiva, junto con la distribución de las lesiones encontradas en el ganado naturalmente infectado, muestran que aproximadamente un 80-90% del ganado se infecta por inhalación (Francis, 1947; Pritchard, 1988).

Varios estudios clásicos realizados con infecciones experimentales apoyan también esta teoría, y se ha comprobado que mientras son suficientes unos pocos bacilos para producir la infección por vía respiratoria, se necesitan varios miles de millones de ellos para hacerlo por vía oral (McFadyean, 1910; Chausé, 1913).

La lesión más frecuentemente encontrada es un foco bronconeumónico en pulmón y lesión en el ganglio linfático correspondiente (complejo primario). La forma en que se desarrollan las lesiones en los linfonódulos retrofaríngeos es aún desconocida. Thoen y Himes, en 1986, propusieron la teoría de que estas lesiones podrían deberse a la inhalación de bacilos que no penetran en el tracto respiratorio inferior y que empiezan a multiplicarse en la orofaringe, la cual drena en los retrofaríngeos.

Tradicionalmente, y de un modo erróneo, se ha considerado que sólo excretan *M. bovis* aquellos animales con lesiones en pulmón claras y evidentes. Al contrario, se ha comprobado que los animales en las primeras fases de la enfermedad y antes del desarrollo de lesiones visibles, contribuyen notablemente a la producción de aerosoles infectivos (McIlroy et al., 1986; Neill et al., 1988). Neill et al., en 1991 y 1992, comprobaron que existe una relación inversamente proporcional entre la dosis infectante y el tiempo que transcurre hasta la liberación por vía respiratoria de *M. bovis*, además de que en las infecciones naturales la eliminación de bacilos comienza a los 87 días después de la infección, lo que ocurre en algunos casos en ganado negativo a la tuberculina. Aunque en el pasado las vías de transmisión oral y genital pudieron tener importancia epidemiológica, desde

el establecimiento de los programas de erradicación la vía dominante de transmisión es la aerógena (Morris et al., 1994).

La vía de transmisión determina la ubicación y el espectro de las lesiones observadas en la tuberculosis bovina. La inhalación es la ruta de infección más común y ocasiona lesiones en la nasofaringe y el tracto respiratorio inferior, como los pulmones y ganglios linfáticos asociados. En contraste, la ingestión de *M. bovis* por medio de pastura, alimento o agua infectados usualmente ocasiona lesiones en los linfonodos mesentéricos, y en ocasiones en la pared intestinal. Otras rutas de infección como la transplacentaria, genital o intramamaria son infrecuentes debido a la situación epidemiológica de la mayoría de los países con programas de erradicación vigentes (Menzies y Neill, 2000). La transmisión transplacentaria hacia el feto probablemente resulta de una endometritis tuberculosa y lleva a un complejo primario en el hígado y/o los nódulos linfáticos portales. Los becerros nacidos con infecciones congénitas usualmente desarrollan la enfermedad generalizada en las primeras semanas o meses de vida. La transmisión sexual de la tuberculosis bovina puede ocurrir cuando la infección se establece en los órganos sexuales; la transmisión intramamaria ocurre de manera iatrogénica de vaca a vaca mediante el uso de tratamientos intramamarios. Por último, la transmisión alimentaria en becerros lactantes es frecuentemente consecuencia de la presencia de *M. bovis* en la leche materna (Vural y Tunca, 2001).

Ciertos estudios parecen demostrar que dentro del ganado bovino existen variaciones en la resistencia a la tuberculosis, de forma que las razas africanas y asiáticas (cebús) son más resistentes a la enfermedad que el ganado europeo (Carmichael, 1939, 1941). En el caso de los primates no humanos hay que señalar que son igualmente susceptibles a la infección por *M. bovis* y *M. tuberculosis*, de igual manera también se han notificado infecciones por *M. africanum*. La infección aparece raramente en animales en libertad, aunque existen comunicaciones de casos ocurridos por ingestión de carcasas procedentes de ganado tuberculoso (Tarara et al., 1985). Sin embargo, en cautiverio, la incidencia de tuberculosis en estos animales es extremadamente alta y se han comunicado varios brotes de

enfermedad en zoológicos (Wilson et al., 1984) y animalarios (Thoen et al., 1977), por lo que la vía de transmisión aerógena juega un papel importante.

2.4. Genotipificación de *Mycobacterium bovis*

Las técnicas de genotipificación de *M. bovis* incluyen, principalmente, los polimorfismos de la longitud de fragmentos de restricción (RFLP), espoligotipificación, unidades repetitivas interespaciadas micobacteriales – repetidos en tándem de número variable (MIRU-VNTR), y actualmente ha tomado importancia la genotipificación con base en polimorfismos de un solo nucleótido (SNP) a partir de la secuenciación del genoma completo.

La genotipificación mediante RFLP se refiere a secuencias específicas de nucleótidos en el ADN que son reconocidas y cortadas por las enzimas de restricción (también llamadas endonucleasas de restricción) y que varían entre micobacterias. Estas secuencias de restricción presentan usualmente patrones de distancia, longitud y disposición diferentes en el ADN de diferentes micobacterias de una población. Los RFLP son marcadores génicos del ADN y se pueden encontrar en regiones que codifican proteínas o exones, en los intrones o en el ADN que separa un gen de otro. Para que un fragmento de ADN pueda utilizarse como marcador genético, este debe ser polimórfico y tener más de un alelo. La técnica está basada en las variaciones generadas por el elemento de inserción IS6110 (transposición). La variación de las cepas está basada en el número de copias del IS6110 y en la posición donde se inserta en el ADN bacterial (Cousins et al., 1998).

La espoligotipificación explora la variación que existe en el locus DR del complejo *Mycobacterium tuberculosis* (MTB) (exclusiva de esta bacteria). El locus DR (Direct Repeat) es una región cromosómica que consiste de regiones repetidas de 36 nucleótidos muy conservados. Cada DR está separado por secuencias espaciadoras no repetitivas de 34 a 41 pb. Las cepas varían en el número de DRs y en la presencia o ausencia de los espaciadores particulares. La técnica utiliza 43

diferentes espaciadores para identificar los diferentes genotipos del complejo MTB (Kamerbeek et al., 1997).

Las MIRU-VNTR son secuencias repetitivas en tándem similares a minisatélites de eucariotas que han sido identificados dispersos en las regiones intergénicas del genoma del complejo MTB. Estos segmentos repetitivos o minisatélites contienen de 46-100 bp y difieren entre las cepas en su número de copias. Se ha demostrado que la tipificación por VNTR es muy útil para patógenos cuya población es altamente clonal como MTB. Las MIRU son VNTR localizadas en el genoma de *Mycobacterium bovis*, se han encontrado 41 posiciones (41 loci) de los cuales 15 son polimórficos en el número de repeticiones de una cepa a otra (Supply et al., 2006).

La secuenciación del ADN es un conjunto de métodos y técnicas bioquímicas cuya finalidad es la determinación del orden de los nucleótidos (A, C, G y T) en un oligonucleótido de ADN. Determinar la secuencia de ADN es útil en el estudio de la investigación básica de los procesos biológicos fundamentales. Además, se puede utilizar la secuenciación del ADN para conocer las mutaciones somáticas, como las sustituciones de bases, generadas entre distintos organismos (Stucki y Gagneux, 2013).

La diferenciación de variaciones genéticas se ha convertido en una herramienta indispensable para el estudio de la evolución, epidemiología y ecología de los organismos patógenos, así como para obtener una mejor percepción sobre las interacciones huésped-patógeno. *M. bovis* pertenece al MTBC, los cuales se caracterizan por tener una similitud de 99.9% en sus secuencias nucleotídicas, así como tener genes de 16S ARNr idénticos y una variación alélica restringida en sus genes estructurales. En la era pos genómica, los polimorfismos de un solo nucleótido (SNP) han servido como una herramienta robusta para delinear relaciones filogenéticas entre cepas de bacterias patógenas estrechamente relacionadas (Álvarez et al., 2009). Además de ser una fuente primaria de variación genética, los SNP son relativamente fáciles de analizar y proporcionan la información necesaria para realizar estudios a gran escala de genética de

poblaciones. Un estudio por parte de García-Pelayo et al., descubrió más de 700 SNP, mediante un análisis comparativo entre los genomas de la cepa virulenta AF2122/97 (del Reino Unido) y la cepa vacunal *M. bovis* BCG Pasteur, el cual ha redefinido la genealogía global de la cepas BCG y ha diferenciado aislados de *M. bovis* de los linajes Francés y Británico (Garnier et al., 2003).

Cuadro 2. Genes definidos como esenciales *in vivo* mediante hibridación del sitio del transposición, que contiene SNP no sinónimos en todas las cepas de *M. bovis* BCG (adaptado de García-Pelayo et al., 2009).

Gen	Secuencia codificante	Secuencia codificante BCG	SNP	Función
<i>nrp</i>	Rv101	BCG0134	L1365M	Sintetasa peptídica no ribosomal
<i>senX3</i>	Rv0490	BCG0531	F109S	Sensor de dos componentes
<i>kdpD</i>	Rv1028c	BCG1085c	P83S; N776D	Sensor de dos componentes
<i>murl</i>	Rv1338	BCG1400	R154L	Glutamato racemasa
<i>lysX</i>	Rv1640c	BCG1679c	D769E	Lisil-ARNt sintetasa
<i>pks12</i>	Rv2048c	BCG2067c	S2964R	Manosil-B-1-fosfoisoprenoide sintetasa
<i>fadE22</i>	Rv3061c	BCG3086c	K488E; S497C	Acetil coenzima A deshidrogenasa
Rv3335c	Rv3335c	BCG3406c	A86V	Proteina integral membranal
Rv3616c	Rv3616c	BCG3680c	A4V	Antígeno de secreción ESX-1

Los SNP representan diferencias de un solo nucleótido entre dos o más secuencias. El término "SNP" frecuentemente es intercambiado por "mutación", "polimorfismo", o "sustitución". En términos estrictos, un cambio en una sola base es denominado mutación puntual, y esto sucede por medio de errores durante la replicación del ADN o como consecuencia de daño al ADN. Cualquier mutación puede incrementarse en frecuencia debido a un desvío genético aleatorio. Cuando un alelo alcanza cierto grado de frecuencia entre la población, se le refiere como

polimorfismo. El umbral para definir una nueva variante como un "polimorfismo" en vez de "mutación" es arbitrario, y normalmente se establece en al menos el 1% de la población (Stucki y Gagneux, 2013).

Según su posición en el genoma, los SNP pueden ser codificantes o no codificantes; la mayoría (90-96%) de los SNP del complejo MTB se encuentran en regiones codificantes. Los SNP codificantes incluso se pueden dividir en sinónimos o no sinónimos, esto si dan lugar a cambios en la secuencia de amino ácidos correspondientes. Ya que los SNP no sinónimos suelen tener un mayor efecto sobre el estado biológico del organismo (ya sea de manera benéfica o maligna), se mantienen bajo una mayor presión de selección, a lo contrario de los SNP sinónimos que no ocasionan cambios en el organismo (Joshi, 2012; Stucki y Gagneux, 2013).

De cualquier forma, los SNP son eventos de rara ocurrencia dentro del complejo MTB, ya que aparecen aproximadamente cada 3 kb en una secuencia de ADN. Esta baja frecuencia y la limitada transferencia horizontal de genes dentro del complejo MTB resultan en disminuidos niveles de homoplasia (e.g. la ocurrencia independiente del mismo SNP en aislados que no están relacionados filogenéticamente). De esta manera, los SNP representan marcadores para inferir filogenias y para la clasificación de cepas. Además, los SNP tienen utilidad en la medición de distancias evolutivas entre aislados, por ejemplo, para estimar el tiempo de divergencia de un aislado de su antecesor genético. Así, mientras más aislados contengan un SNP en particular, más antiguo se considera dicho SNP (García-Pelayo et al., 2009).

2.5. El genoma de *Mycobacterium bovis*

El genoma de referencia *M. bovis* AF2122/97 es una cepa altamente virulenta obtenida en el Reino Unido en el año 1997 de una vaca infectada con lesiones caseosas en el pulmón y los linfonodos broncomediastínicos. La secuencia genómica contiene 4,345,492 pares de bases (pb), y se encuentra constituido en un cromosoma de forma circular, con un contenido promedio de G+C de 65.63%.

El genoma contiene 3,952 genes que codifican a proteínas, incluyendo un profago y 42 elementos de inserción (IS) (Garnier et al., 2013). El genoma de *M. bovis* presenta una similitud de más del 99.95% a nivel de nucleótidos con *M. tuberculosis*, pero su genoma es más pequeño debido a las mutaciones genéticas que ha sufrido *M. bovis*, principalmente por mecanismos de delección (Guta et al., 2014). El mecanismo de delección es un evento considerado único y unidireccional en las bacterias y es una mutación longitudinal, donde existe una ausencia anormal de un segmento de un cromosoma (Garnier et al., 2013). Se han revelado 11 delecciones del genoma de *M. bovis* con un tamaño aproximado de 1 a 12.7 kilobases (kb). No existen evidencias de otros mecanismos de mutación como translocación, duplicación e inversiones, tal vez por el lento crecimiento de la micobacteria. Sin embargo, existen polimorfismos de un solo nucleótido (SNP) distintivas del bacilo bovino (Golby et al., 2013). Encontrándose 2,437 SNP entre el genoma de *M. bovis* AF2122/97 y *M. tuberculosis* H37R, y 2,423 SNP comparado con *M. tuberculosis* CDC1551 (Acosta-Salinas et al., 2009; Garnier et al., 2003).

Cuadro 3. Comparación del genoma de *M. bovis* AF2122/97 con *M. tuberculosis* H37R y CDC1551.

Característica	<i>M. bovis</i> AF2122/97	<i>M. tuberculosis</i> H37R	<i>M. tuberculosis</i> CDC1551
Tamaño del genoma, pb	4,345,492	4,411,532	4,403,836
G+C %	65.6	65.6	65.6
Genes codificantes para proteínas*	3,951	3,995	4,249
SNP	-	2,437	2,423
Transición	-	1,649	1,630
Transversión	-	788	793
Delección (≥1 pb)	-	205	221
Inserción (≥1 pb)	-	117	245

Antes de que estuviera disponible la secuencia del genoma de *M. bovis*, la genómica comparativa del MTBC se realizaba utilizando métodos basados en la hibridación de ADN, los cuales se aprovechaban del alto grado de identidad que existe entre secuencias de este complejo. La secuencia de *M. bovis* contiene solamente un locus de TbD1, el cual se encuentra ausente en la mayoría de las cepas existentes de *M. tuberculosis* (Garnier et al., 2003).

Las paredes celulares de las bacterias patógenas muestran variaciones en las secuencias de proteínas y su composición macromolecular, lo cual se ve reflejado por la presión de selección de estas estructuras (Fernandes-Ramos et al., 2014). Por lo tanto, el mayor grado de variación entre las secuencias de los bacilos humano y bovino se encuentre en los genes codificantes para proteínas de secreción y de la pared celular. La variación en los genes codificantes de lipoproteínas se puede observar en *lppO*, *lppT*, *lppG*, y *lppM*, como deleciones o translocaciones, mientras que *M. bovis* tiene una copia duplicada de *lppA*. De manera similar, el gen *rpfA* en *M. bovis*, el cual es uno de los cinco miembros de una familia que codifica para proteínas de secreción que promueven la reactivación de bacilos latentes, muestra un deleción dentro del marco de lectura de 240 pb que da lugar a la síntesis de una proteína de menor tamaño (Brunton et al., 2015).

Existe una gran variación en los genes codificantes de las familias de proteínas PE-PGRS y PPE, las cuales se piensa que son antígenos de superficie y juegan un papel importante en la adhesión y la respuesta inmune (Majlessi et al., 2015). Otro grupo de antígenos ya conocidos que son afectados por las deleciones que ocurren en *M. bovis* son de la familia ESAT-6. Seis de las proteínas ESAT-6, codificadas por los genes Rv2346c, Rv2347, Rv3619c, Rv3620c, Rv3890c (Mb3919c) y Rv3905c (Mb3935c) en *M. tuberculosis*, están ausentes o alteradas en *M. bovis*, lo cual tiene impacto sobre el potencial antigénico de la micobacteria (Hauer et al., 2015).

Dentro de las diferencias más evidentes en cuanto a la variación en el secretoma está la expresión elevada de dos antígenos serodominantes, MPB70 y

MPB83, en el bacilo del bovino. MPB83 es una proteína glucolisada asociada a la pared celular, mientras que MPB70 es una proteína de secreción que puede llegar a constituir el 10% del filtrado proteico del cultivo de *M. bovis* (Ludanny et al., 2015). También se pueden observar diferencias en los genes codificantes para la síntesis (*pks*) y transporte (*mmpSL*) de policétidos y lípidos complejos con fracciones de policétidos. Estos lípidos son factores importantes en la inducción de patologías en el huésped que mejoran el medio ambiente para los patógenos. Los genes *pks1*, *mmpL13* y Mb1695c (un transportador macrólido putativo adyacente al conjunto *pks10/7/8/17/9/11*), pueden ser traducidos a productos funcionales en *M. bovis*, pero son interrumpidos en *M. tuberculosis*. Por el contrario, los genes *pks6*, *mmpL1* y *mmpL9* están trastornados en *M. bovis*. Se ha observado a nivel funcional que *pks1* codifica para la biosíntesis del glicolípido fenólico principal de *M. bovis*, y las cepas que muestran alteración de éste no logran la producción de dicho lípido (Allen et al., 2013).

Al análisis de deleciones en *M. bovis* ha revelado que la alteración del gen *treY* no está presente en todas las cepas de *M. bovis*, lo cual convierte a este gen como un posible marcador molecular para análisis filogenéticos. Además, la interrupción del ortólogo del gen Rv1373 (Mb1407/8) da lugar a una insuficiencia de lípidos sulfatados en la membrana, ya que la enzima para la cual codifica es una sulfotransferasa glucolípídica. Estas diferencias conllevan efectos importantes sobre el fenotipo de la micobacteria y su interacción con el huésped (Gutiérrez-Reyes et al., 2012).

La deleción de información es una tendencia dominante en el genoma de *M. bovis*. Los genes involucrados en transporte y estructuras de la superficie celular (*pstB*, *ugpA*, *mc3A-F*, *lppO*, *lpqG*, *lprM*, *pks6*, *mmpL1*, *mmpL9*, Rv1510, Rv1508, Rv1371), el metabolismo de los ácidos grasos (*fadE22*, *echA1*), biosíntesis de cofactores (*moaE*, *moaC2*), detoxificación (*ephA*, *ephF*, *alkA*) y el metabolismo de intermediarios (*epiA*, *gmdA*) son pseudogenes o han sido eliminados. De igual manera, *M. bovis* ha perdido el sistema *AtsA* para el reciclaje de sulfato. La pérdida de esta función está vinculada a la insuficiencia de glucolípidos sulfatados. Sin

embargo, como *M. bovis* puede llevar a cabo recombinación homóloga, es probable que efectos polares puedan suprimir los defectos (Cadmus et al., 2011).

Existe la posibilidad de que los eventos de delección de genes del genoma de *M. bovis* representan “agujeros negros”, i.e., la pérdida de genes que son perjudiciales para la patogenicidad de un nicho específico. Sin embargo, los análisis llevados a cabo hasta la fecha sugieren que, aunque el proceso de adaptación no necesitó la presencia de ciertos genes, las alteraciones en la expresión de algunos genes y los elementos expuestos de la membrana celular pudieron jugar un papel muy importante. De esta manera, conocer la secuencia completa de genoma puede tener gran impacto sobre el entendimiento de la evolución, adaptación al huésped y patobiología de la tuberculosis, además de que en un futuro será esencial para la elaboración de vacunas y reactivos de diagnóstico para combatir la enfermedad (Garnier et al., 2003).

2.6. Evolución de *Mycobacterium bovis*

El complejo de bacterias MTB, hasta donde se conoce, es completamente clonal, sin ninguna evidencia de que exista transferencia de material genético entre linajes. Esta singular estructura poblacional es compartida solamente con algunos otros organismos y normalmente se le considera como una característica derivada (Smith, 2012). La mayoría de los organismos contienen secuencias genómicas derivadas de dos o más individuos – cada uno con una historia filogenética distinta. Las tasas muy bajas de homoplasias (mutaciones idénticas en linajes diferentes) y la congruencia de diferentes marcadores moleculares encontrada en las reconstrucciones filogenéticas de secuencias nucleotídicas del MTBC son consecuencia de la clonalidad de estos microorganismos (Smith, 2012). Se espera que sea de esta manera en un organismo clonal puesto que las homoplasias solamente pueden ser generadas a partir de recombinación de material genético entre diferentes linajes, o mediante mutaciones independientes repetidas en el mismo locus. La muy baja diversidad en las secuencias dentro del MTBC sugiere

que las homoplasias ocasionadas por mutaciones repetidas son improbables a menos que el locus esté evolucionando bajo alguna forma de selección direccional (por ejemplo, los genes que adquieren resistencia a los antibióticos), o si un locus tiene una tasa de mutación más alta que la tasa de mutación nucleotídica general (transposición de la IS6110), o incluso si un proceso de mutación se presta para generar alelos similares en diferentes linajes (loci de VNTRs) (García-Pelayo et al., 2009).

Puesto que el cromosoma de organismos completamente clonales evoluciona como un único grupo vinculado, la historia filogenética de estos indica que coalescen hacia un solo organismo (una célula) que existió en un lugar en un punto del tiempo. Aunque la tuberculosis ha sido reportada en cada continente donde hay ganado, es posible asumir que este patógeno se originó en un solo lugar, en algún punto del tiempo, y desde entonces se ha distribuido a través del mundo (Smith, 2012). A la fecha se han descrito cuatro complejos clonales de *M. bovis* mediante la presencia de deleciones específicas, cada una con espoligotipos específicos; mientras que dos se distribuyen principalmente en África (Af1 y Af2), Eu1 y Eu2 tienen una distribución más amplia, aunque de estos, la distribución de Eu2 es menor y se piensa que tuvo su origen en la Península Ibérica (Cuadro 4).

Cuadro 4. Definición y descripción de los Complejos Clonales de *M. bovis* Af1, Af2 y Eu1.

Africano 1 (Af1)	
Definición	Presencia de la deleción RDAf1 (5.3kb entre Mb0586c y Mb0590c)
Marcador del espoligotipo	Ausencia del espaciador 30
Secuencia del espoligotipo	11011111011111111111111101111111100000
Referencia	Muller et al., 2009
Distribución	Alta frecuencia en Africa sub-Sahara del Oeste Central (Mali, Camerún, Chad, Nigeria)
Africano 2 (Af2)	
Definición	Presencia de la deleción RDAf2 (14.1kb entre MB0599 y Mb0610)

complejos clonales a los que pertenecen es simple puesto que los complejos clonales identificados hasta ahora son mutuamente exclusivos – un miembro de un complejo clonal no puede ser miembro de otro complejo clonal al mismo tiempo (Smith, 2012). La definición precisa de un complejo clonal puede ser un poco arbitraria, pero se ha procurado identificar aquellos complejos clonales de *M. bovis* que se encuentran con alta frecuencia en varios países, y por lo tanto son de gran importancia epidemiológica.

Se ha sugerido que *M. tuberculosis* (el patógeno del humano) surgió de *M. bovis* durante el periodo en que el ganado fue domesticado por el hombre, aproximadamente hace 10-15,000 años, cuando el bacilo tuberculoso del bovino fue transmitido a la población humana (Diamond, 2002). Esta conclusión se basó en la observación de cepas existentes de *M. bovis* en una amplia gama de hospederos, infectando mamíferos salvajes y domesticados, así como al humano, mientras que la infección natural por *M. tuberculosis* se encuentra aparentemente restringida al humano. Utilizando la información de los genomas de las cepas de humano y bovino fue posible darle respuesta a estas cuestiones, y mediante la información genética confirmar o rechazar las ideas previas sobre la evolución de *M. tuberculosis*. Una combinación de micro- y macro-ensayos además de genómica comparativa, han descrito un conjunto de deleciones en *M. bovis* en comparación con *M. tuberculosis*. Puesto que no existe la transferencia horizontal de material genético entre cepas del complejo *M. tuberculosis* (MTBC), las deleciones pueden servir como marcadores unidireccionales para la construcción de las filogenias. La comparación de las deleciones de *M. bovis* contra las especies del MTBC (*M. tuberculosis*, *M. africanum*, *M. microti*, *M. bovis*, etc.) permitió refutar la idea de que *M. tuberculosis* en humanos fue ocasionado por una selección de una forma adaptada-a-humanos de *M. bovis* que evolucionó hacia *M. tuberculosis* (Mostowy et al., 2002). Mediante la inclusión de las mutaciones de un solo nucleótido y las deleciones de las unidades de los espoligotipos, Hewinson et al. (2006) lograron expandir el análisis filogenético de cepas del MTBC y concluyeron que este grupo de organismos serían mejor descritos como una serie de ecotipos adaptados a

ciertos huéspedes, cada uno con una preferencia de huésped que representa los diferentes nichos (Brosch et al., 2002; Smith et al., 2006).

Dirección General de Bibliotecas UAQ

III. HIPÓTESIS

La genotipificación de cepas de *M. bovis* de México con base en secuenciación genómica y SNP diferenciará con mayor grado de resolución los grupos genéticos establecidos a partir de espoligotipificación, lo cual representará de manera más precisa la diversidad genética de este patógeno asociado a bovinos de diferentes regiones del país.

Dirección General de Bibliotecas UAO

IV. OBJETIVOS

4.2. Objetivo general

Caracterizar la variabilidad genética de cepas de *M. bovis* de México mediante secuenciación de genoma completo.

4.3. Objetivos específicos

1. Secuenciar los genomas de 320 cepas de *Mycobacterium bovis* de diferentes estados de México.
2. Determinar los SNP asociados a estas cepas.
3. Realizar un análisis filogenético de las cepas.
4. Analizar el uso de secuenciación genómica completa (WGS) y la espiligotipificación para el estudio de la epidemiología de la tuberculosis bovina.
5. Realizar un análisis *in silico* de los efectos de los SNP sobre la expresión génica.
6. Analizar la diversidad genómica del conjunto de cepas incluidas en este estudio.

V. METODOLOGÍA

El presente estudio es de carácter descriptivo, y consistió en dos etapas. En una primera etapa se realizó un análisis de las relaciones genéticas entre las cepas y la evaluación de la diversidad genética. La segunda etapa consistió en el análisis genómico, el cual incluyó una predicción *in silico* de los efectos de los SNP, así como una evaluación de la diversidad genómica del conjunto de cepas.

5.2. Cepas bacterianas

El estudio incluyó un total de 320 cepas de *M. bovis* provenientes de distintos estados de México: Aguascalientes (28), Baja California (25), Coahuila (10), Guanajuato (8), Hidalgo (12), Jalisco (33), Estado de México (30), y Querétaro (174). Las cepas se obtuvieron de bovinos a partir de monitoreo en rastros, a través de la inspección de las canales e identificación de lesiones macroscópicas en distintos tejidos (nódulos linfáticos, pulmón, hígado, etc.), mientras que dos cepas se obtuvieron a partir de muestras de expectoración de pacientes con tuberculosis pulmonar. Esta colección de cepas se creó a lo largo de varios años, por lo que se incluyen cepas de 1997 (128), 2009 (83), 2010 (33), 2012 (31), 2013 (34), 2014 (1) y 2015 (10). El Cuadro 3 expone a manera de resumen las características de las cepas.

Cuadro 5. Características de las cepas de *M. bovis* utilizadas en este estudio.

Año	Cantidad	Proporción (%)
1997	128	40
2009	83	26
2010	33	10
2012	31	10
2013	34	11
2014	1	0

	2015	10	3
Estado			
	Aguascalientes	28	9
	Baja California	25	8
	Coahuila	10	3
	Guanajuato	8	3
	Hidalgo	12	4
	Jalisco	33	10
	Estado de México	30	9
	Querétaro	174	54
Huésped			
	Bovino	318	99
	Leche		75
	Carne		15
	Desconocido		230
	Humano	2	1

5.3. Aislamiento de *Mycobacterium bovis*

El aislamiento de la micobacteria se obtuvo por medio de cultivo, el cual se realizó mediante el método de Petroff modificado (Organización Panamericana de la Salud, 2008), en medio Stonebrink adicionado con piruvato de sodio como fuente de carbono (especializado para *M. bovis*). Los cultivos se mantuvieron en horno de incubación a 37°C durante un periodo de 3 a 6 semanas. La extracción de ADN se realizó mediante la técnica CTAB (Anexo 1), la cual incluye lisozima-proteinasa K y fenol-cloroformo para la obtención y purificación del ADN (Rodríguez-Hernández et al., 2014), con purezas obtenidas en un rango de 1.6 a 1.9, y concentraciones promedio de 200 ng/μl. Para la identificación de *M. bovis* se llevó a cabo la técnica de PCR descrita por Nava-Vargas (2015) (Anexo 2).

5.4. Espoligotipificación

La genotipificación mediante el uso de espoligotipos se llevó a cabo *in silico* a partir de la secuenciación genómica, con base en la presencia o ausencia de los fragmentos de repetidos directos (DR) utilizando el programa RD-Analyzer (Faksri et al., 2016). El programa utiliza algoritmos de búsqueda para predecir regiones de delección en el genoma a partir de las lecturas obtenidas de la secuenciación. Utiliza los mecanismos de alineamiento y mapeo de los programas BWA-MEM y SAMtools (v 0.1.19) (Li y Durbin, 2009; Li et al., 2009).

5.5. Secuenciación genómica (WGS)

Las bibliotecas de secuenciación se prepararon de acuerdo a las instrucciones del fabricante, utilizando el kit de preparación de ADN Nextera XT (Illumina, San Diego, CA). Estas se normalizaron utilizando el kit comercial Agencourt AMPure XP (Beckman Coulter, Brea, CA), de acuerdo a las instrucciones del fabricante. La secuenciación se llevó a cabo con el equipo MiSeq utilizando el kit MiSeq Reagent Kit versión 2 (Illumina). Las corridas se realizaron de acuerdo a las recomendaciones del fabricante para producir lecturas pareadas de 2 x 250. La cobertura que obtenida fue de al menos 30x. Todas las secuencias (SRA) se encuentran disponibles en la base de datos del Archivo de Secuencias de Lecturas (SRA) del Centro Nacional para la Información Biotecnológica (NCBI) (<https://www.ncbi.nlm.nih.gov/>), con folio de proyecto PRJNA449507.

5.5.2. Revisión de calidad de las secuencias

Los archivos FASTQ obtenidos a partir de la secuenciación se revisaron mediante el programa FASTQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) para evaluar la calidad de las secuencias. Con base en los parámetros de “Estadísticas Básicas” (longitud

de la secuencia de 20 a 250 pb; %GC=65) y “Calidad de la secuencia por cada base” (puntaje phred 26 a 40), se rechazaron las secuencias que no cumplieron con dichos parámetros. Adicionalmente, para poder recuperar mayor cantidad de secuencias, se eliminaron los fragmentos de mala calidad mediante el programa TRIMGALORE (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), utilizando los parámetros por defecto para lecturas pareadas con adaptadores de Illumina.

5.6. Identificación de los polimorfismos de un solo nucleótido (SNP)

Las secuencias obtenidas a partir del MiSeq se alinearon al genoma de referencia de *M. bovis* AF2122/97 (Número de acceso de NCBI NC_02945), mediante el programa Burrows-Wheeler Aligner v0.7 (BWA) (Li y Durbin, 2009), utilizando la configuración por defecto. El archivo SAM resultante se convirtió a un archivo BAM clasificado e indexado utilizando Samtools 0.1.18 (Li et al., 2009). Los duplicados se identificaron y se removieron utilizando el programa MarkDuplicates de Picard Tools 1.81 (<http://picard.sourceforge.net>). El programa Genome Analysis Tool Kit (GATK) (McKenna et al., 2010) se utilizó para recalibrar los puntajes de la calidad de las bases, el re-alineamiento de indels, y el descubrimiento de SNP, con los parámetros estándar de filtro fijo o mediante la re-calibración del puntaje de la calidad de variantes de acuerdo a las recomendaciones de Buenas Prácticas de GATK (DePristo et al., 2011; Van der Auwera et al., 2013). Se aplicaron dos filtros adicionales para remover SNP en el archivo de variantes (VCF) respectivo. Primero, los SNP con puntajes de calidad <150 y AC=1 se removieron. En segundo lugar, los SNP que cayeron en regiones repetidas (familias de PPE, PE_PGRES, elementos transponibles), o que se encontraron en loci con baja densidad de lectura se también removieron (similar a lo descrito por Bryant et al., 2013; Ford et al., 2011; Kato-Maeda et al., 2013). Todos los SNP restantes de cada muestra se compilaron y alinearon contra la referencia usando la posición de la referencia. La Tubería de Análisis de SNP está disponible en línea (https://github.com/stuber/SNP_analysis).

Los alineamientos y SNP se validaron y visualizaron utilizando el programa Integrated Genomics Viewer (Robinson et al., 2011; Thorvaldsdóttir et al., 2013).

5.7. Análisis filogenético

Para cada cepa, los SNP identificados se concatenaron en una sola secuencia de caracteres. Estas secuencias se alinearon y utilizando el programa RAxML (Stamatakis, 2014), se realizó la reconstrucción de la filogenia de las cepas. Se utilizó un modelo GTR+I y el algoritmo de máxima verosimilitud para la elaboración del árbol filogenético.

5.8. Análisis del efecto de los SNP sobre la función proteica

Primero, se obtuvieron los SNP mediante el programa BRESEQ (Deatherage et al., 2014). Se realizó un alineamiento de cada cepa mexicana contra la cepa de referencia *M. bovis* AF2122/97 (NCBI Accession NC_002945.4). Las lecturas fueron mapeadas al genoma de referencia y se identificaron mutaciones puntuales en la cepa mexicana con respecto a *M. bovis* AF2122/97. Los archivos de entrada fueron las lecturas pareadas (fastq.gz), la secuencia de referencia NC_002945.4, con parámetros por defecto. Los archivos de salida incluyen una página .html donde se pueden observar visualmente, a manera de cuadros, las mutaciones encontradas, su posición respecto al genoma de referencia, el tipo de mutación, y si esta se encuentra dentro de un gen.

Dentro de los archivos de salida también está incluido un archivo vcf (variant call file), el cual es una compilación de todas las mutaciones predichas para cada cepa en particular, y toda la información correspondiente (cromosoma, posición, identificador, secuencia de referencia, secuencia blanco, calidad, etc.).

Para el análisis del efecto sobre la función proteica, se utilizó el archivo .vcf de cada cepa como entrada para el programa SNPEFF (Cingolani et al., 2012). Éste requiere los archivos VCF obtenidos a partir de la identificación de los SNP para

realizar el análisis de predicción. El programa categoriza los efectos de los SNP como sinónimos o no sinónimos, pérdida o ganancia de codón de inicio, pérdida o ganancia de codón de paro, así como referente a la posición (5'UTR, 3'UTR, intergénico, río arriba, río abajo, etc.).

5.9. Ensamblaje de los genomas

El ensamblaje de los genomas consistió en obtener fragmentos contíguos (contigs) para cada cepa, los cuales son segmentos de ADN superpuestos que juntos representan una región consenso de ADN (Aguilar-Bultet y Falquet, 2015). Para realizar el ensamblaje de los genomas, se utilizaron los programas VELVET (Zerbino y Birney, 2008) y SPAdes (Bankevich et al., 2012). Como archivos de entrada se utilizaron las secuencias pareadas fastq de Illumina correspondientes a cada cepa. Para SPAdes, se utilizaron los parámetros "--careful", con cámaras 33, 55, 77, 99, 111, 121, 127. Para VELVET, primero se utilizó velvet para obtener el óptimo número de cámaras a utilizar con un tamaño mínimo de 33 y máximo de 249, con lecturas cortas pareadas fastq ("--fastq --s 33 --e 249 --x 22 -shortPaired -separate"). Una vez obtenido el tamaño de cámara óptimo, se corrió velveth con los parámetros "-fastq -shortPaired -separate" y velvetg con los parámetros --cov_cutoff auto "--exp_cov auto --ins_length 350" para designar el porcentaje de cobertura y el tamaño de los insertos.

5.9.2. Revisión de calidad de los ensamblajes

Debido a que es necesario utilizar ensamblajes de buena calidad para realizar los análisis genómicos posteriores, se utilizó el programa QUAST (Gurevich et al., 2013) para evaluar la calidad de los ensamblajes realizados por VELVET y SPAdes. Como archivo de entrada se utilizaron los contigs obtenidos a partir del ensamblaje en formato .fasta, y los parámetros por defecto del programa. Como resultado el programa evalúa los contigs y produce métricas descriptivas que incluyen el número

total de contigs, la longitud del contig más largo, la longitud total del genoma, el contenido de GC (guaninas-citosinas) en porcentaje, y los valores de N50, NG50, N75, NG75 y L50, principalmente. De acuerdo a los valores de estos parámetros que ya se han reportado previamente, se descarta o no un ensamblaje para los siguientes análisis.

5.10. Anotación funcional de los genomas

Una vez evaluados los ensamblajes y descartados los ensamblajes de mala calidad, se procedió a realizar la anotación funcional de los genomas. Esto consiste en definir la función de los genes que se encuentran en una secuencia genómica. Generalmente se utilizan algoritmos que consisten en una búsqueda por homología de las secuencias encontradas en el genoma blanco, y compararlas contra una base de datos curada y establecida donde se encuentran ya descritos genes para algún organismo en particular. Para este estudio se utilizaron los programas PROKKA (Seemann, 2014) y RAST (Aziz et al., 2008) para llevar a cabo este proceso.

PROKKA toma como archivos de entrada los contigs de los genomas ensamblados, en formato .fasta. Para mayor especificidad, se utilizaron los parámetros --genus *Mycobacterium* --species *bovis* --kingdom Bacteria. Los archivos de salida .fna, .faa y .tbl proporcionan la secuencia original de nucleótidos para cada contig, la secuencia traducida a amino ácidos para cada gen, así como un cuadro que contiene todas las secuencias codificantes encontradas, respectivamente. RAST, adicionalmente, categoriza los genes de acuerdo a su función y los agrupa en subsistemas funcionales y familias de proteínas. Los parámetros utilizados para RAST fueron "Dominio: Bacteria", "Género: *Mycobacterium*", "Especie: *bovis*", "Código genético: 11". Se utilizó el esquema de anotación "Classic RAST" y automáticamente corregir errores. Para los demás parámetros se eligieron los valores por defecto.

5.11. Análisis del pangenoma

La diversidad genómica de este conjunto de cepas de *M. bovis*, se evaluó mediante un análisis de pangenoma. El pangenoma describe la colección de todos los genes en una especie (o grupo de microorganismos) en particular, ya que las bacterias pueden presentar una gran variación de contenido génico entre cepas estrechamente relacionadas. El pangenoma incluye el genoma central (o “core”) que consiste en todos los genes presentes en todas las cepas de un determinado clado, y el genoma accesorio que consiste en todos los genes dispensables, o no esenciales, que se encuentran solamente en algunas de las cepas, así como genes específicos de una cepa en particular (McInerney et al., 2017). Para esto se utilizó el programa ROARY (Page et al., 2015). Los archivos de entrada utilizados fueron los .gff obtenidos a partir del programa PROKKA. Los parámetros de funcionamiento fueron `-e -mafft -i 90`, indicando que se realice un alineamiento rápido mediante MAFFT y se utilice un porcentaje de identidad mínima de 90.

5.12. Análisis estadísticos

Se realizaron análisis multivariados para evaluar las diferencias entre el contenido génico de las cepas, así como con base a las variantes (SNPs) obtenidas para cada una de ellas. Se analizó la relación entre las cepas mediante un análisis de coordenadas principales (PCoA), para lo cual se utilizó el programa R Studio, y la paquetería `vegan` para la visualización del análisis. Se elaboró una matriz de distancias con los SNPs, utilizando 1 como presencia y 0 como ausencia, y el método de Jaccard. También se elaboró un mapa de calor donde se representan las distancias entre cada una de las cepas. Para validar la significancia estadística de estos resultados, se realizó un análisis de permanova. Para explorar la presencia de algún patrón o estructura con respecto a las variables de respuesta (contenido génico por subsistemas), se realizó un análisis de componentes principales, con las paqueterías `ggplot2`, `ggpubr`, `factoextra`, y `FactoMineR`, utilizando la función `prcomp`,

y una matriz de correlaciones con base en la base de datos del contenido génico de cada cepa, categorizados por subsistemas de acuerdo a FIGfams (<http://seed-viewer.theseed.org/seedviewer.cgi?page=FigFamViewer>).

Dirección General de Bibliotecas UAQ

VI. RESULTADOS Y DISCUSIÓN

En este trabajo se secuenció el genoma completo de 320 cepas de *M. bovis* de ocho estados de México. De estas, 318 fueron de bovino (216 de ganado lechero y 12 de ganado cárnico – no fue posible confirmar esta información para las 90 cepas restantes) y dos de humano.

6.2. Espoligotipificación y secuenciación genómica

El análisis de las 320 cepas de *M. bovis* reveló 39 espoligotipos, de los cuales 11 se clasificaron como “desconocidos” (UNK) puesto que no se encuentran aún registrados en la base de datos internacional Mbovis.org. El espoligotipo más frecuente fue SB0673, seguido por SB0971, SB0140 y SB0145. Casi todos los espoligotipos se encontraron distribuidos ampliamente; sin embargo, el espoligotipo SB0145 fue el único cuya distribución abarcó los nueve estados. Dos espoligotipos, SB1058 y SB1531, fueron exclusivos de Baja California. La mayor diversidad se encontró en el estado de Querétaro, con 24 de los 39 (61%) espoligotipos. De manera contraria, Baja California tuvo la menor diversidad, con sólo dos espoligotipos. El análisis evolutivo de los espoligotipos identificó al SB0140 como el genotipo ancestral, mismo que dio origen a tres de los espoligotipos más frecuentes SB0673, SB0971 y SB0145 (Figura 2). Estos, a su vez, dieron origen a otros espoligotipos de menor frecuencia. Sin embargo, de acuerdo al espoligoforest, los espoligotipos SB0130 y SB0121 dan lugar a dos grupos genéticos diferentes cada uno, cuya relación genética no se pudo resolver con respecto al resto de las cepas; y para cinco espoligotipos (SB1213, UNK3, UNK4, UNK5 y UNK8) no se hayó relación genética alguna con respecto a los demás. Adicionalmente, el espoligotipo SB0673 resultó ser el más prolífico, ya que es el origen de nueve distintos espoligotipos. Reyes et al. (2008) explican que cuando un nodo en un spoliogforest es muy grande, como el que se ve para el SB0673 en la Figura 2, esto quiere decir que es posible que ese espoligotipo se haya transmitido de manera extensiva

durante un largo periodo de tiempo; además, cuando un nodo tiene un gran número de descendientes, esto sugiere que ese espoligotipo ha tenido un largo periodo de tiempo a través del cual pudo generar mutaciones. El espoligotipo SB0673 ha sido reportado en México en varios estudios, sugiriendo que éste ha sido dominante durante un largo periodo de tiempo en el país (Rodwell et al., 2010; Gutiérrez-Reyes et al., 2012; Zumárraga et al., 2013; Vazquez-Chacon et al., 2015; Milián-Suazo et al., 2016; Cobos-Marín et al., 2005).

Dirección General de Bibliotecas UNAM

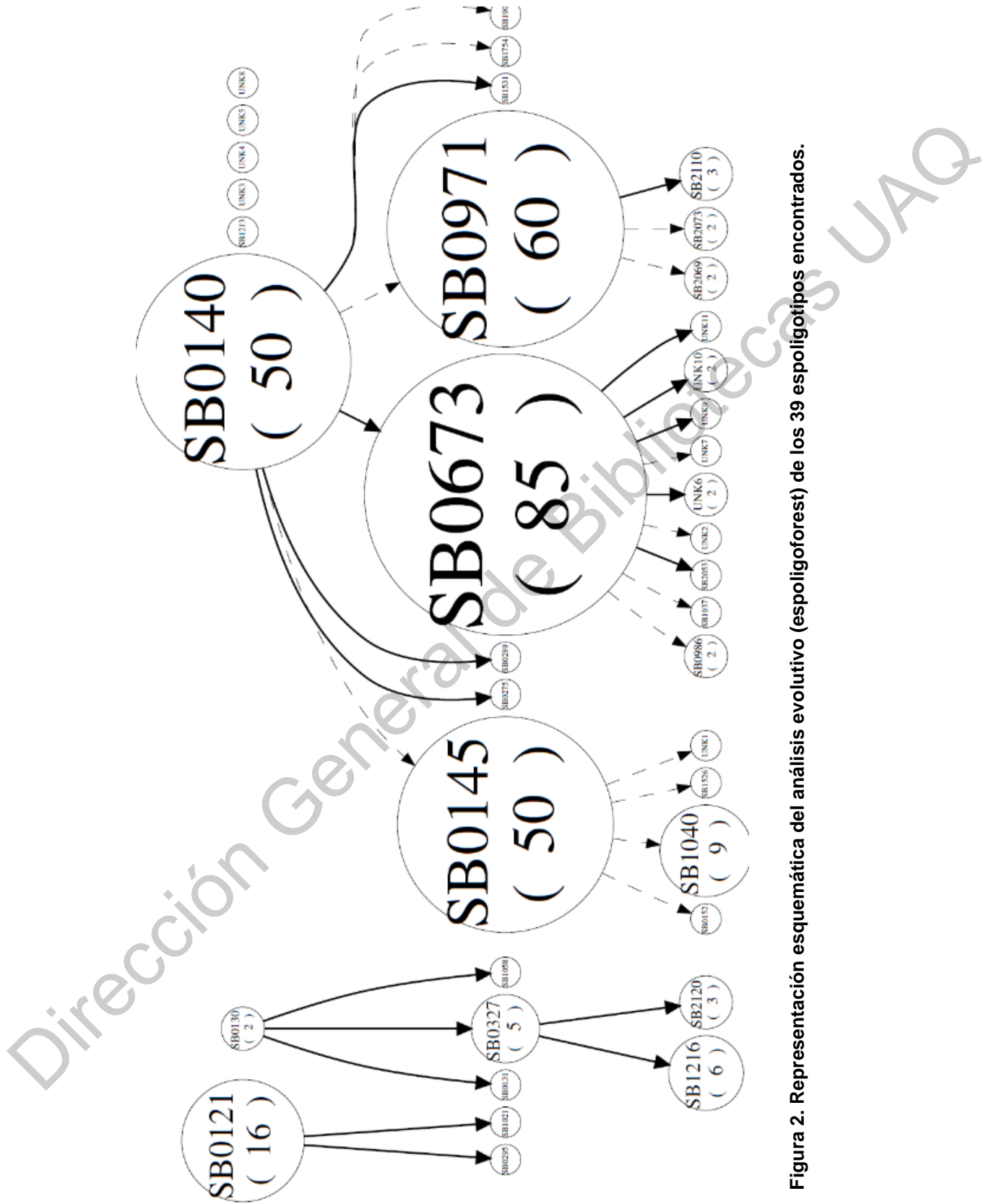


Figura 2. Representación esquemática del análisis evolutivo (espoligoforest) de los 39 espoligotipos encontrados.

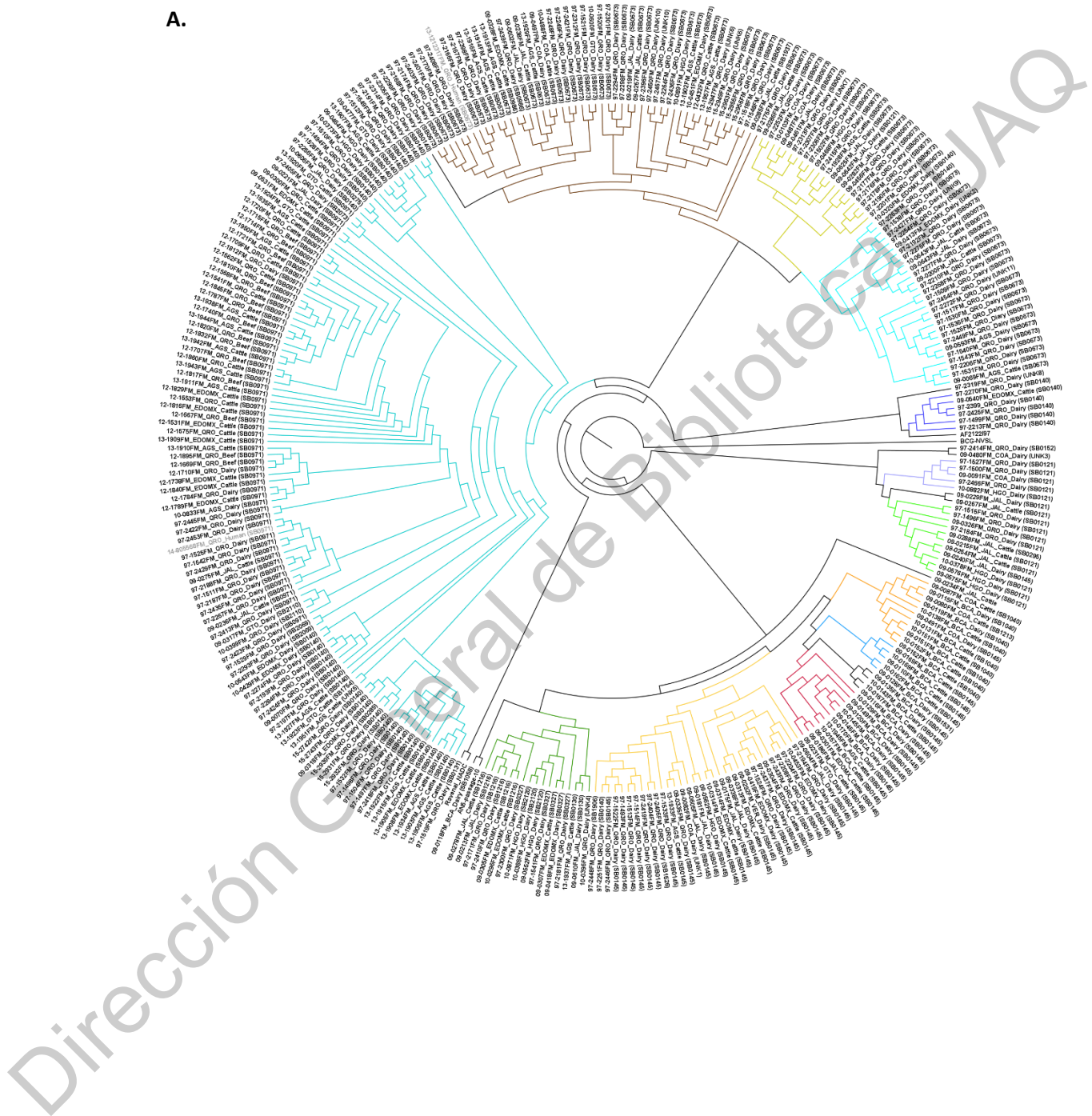
Con base al análisis a partir de los SNP, se lograron recuperar 12 grupos genéticos (Grupo SNP) principales, A a L (Cuadro 6 y Figura 3). Los Grupos E y D incluyeron la mayor cantidad de cepas, 51 y 107 respectivamente. El Grupo D incluyó cepas provenientes de ganado lechero y cárnico, mientras que las cepas provenientes de humanos se encontraron dentro de los Grupos E y D. Los Grupos D y E tuvieron la mayor distribución, encontrados en 7 de los 9 estados incluidos en el estudio. El Grupo I fue exclusivo de Baja California, y nuevamente, Querétaro tuvo la mayor diversidad genética con 9 de los 12 (75%) grupos genéticos. Sin embargo, Veracruz solamente tuvo el Grupo G.

Cuadro 6. Distribución de los grupos genéticos por estado y número de cepas correspondientes a cada uno.

ORIGEN	AGS	BCA	COA	EDOMX	GTO	HGO	JAL	QRO	VER
CEPAS	25	19	9	29	8	12	30	170	1
GRUPO SNP	B	H	C	A	D	C	B	A	G
	C	I	D	B	E	D	C	B	
	D	J	G	D	H	E	D	C	
	E		J	E		F	E	D	
	G		L	F		G	F	E	
	H			G		K	G	F	
				H		L	H	G	
							K	K	
								L	

En la filogenia no se logra observar ningún tipo de agrupamiento por huésped (ganado de leche, ganado de carne o humano) o región geográfica, con excepción de Baja California, cuyas cepas forman un clado específico formado por los grupos H, I y J (Figura 3).

A.



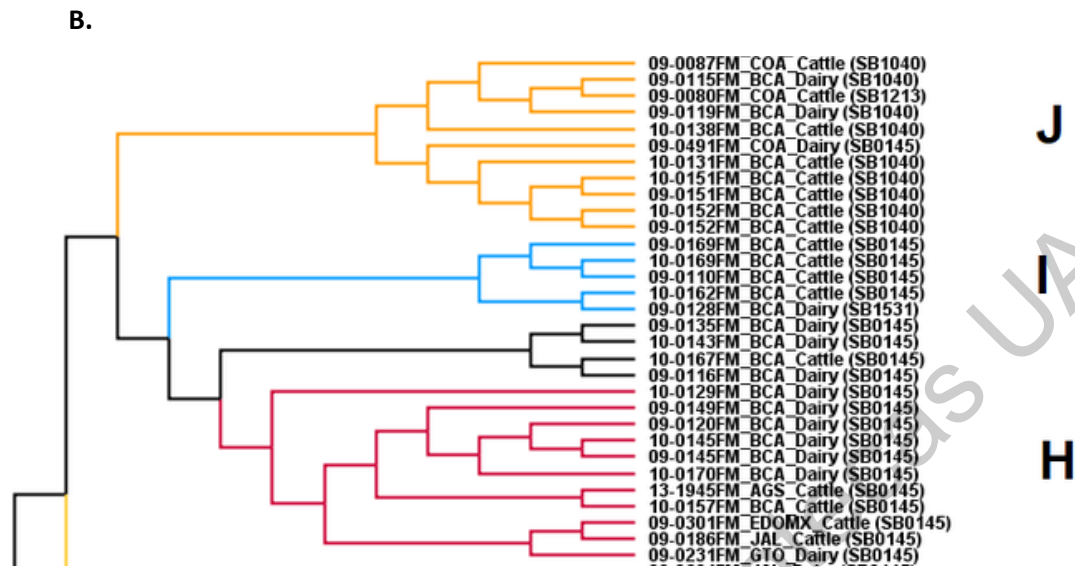


Figura 3. A. Árbol filogenético elaborado a partir de las secuencias concatenadas de SNP obtenidas por medio de la secuenciación genómica de 320 cepas de *M. bovis* de México. B. Acercamiento a los Grupos H, I, J, lo cual muestra que la mayoría de las cepas son de Baja California.

Estudios previos en México (Perumalla et al., 1996; Milián-Suazo et al., 2000; Milián-Suazo et al., 2010; Nava-Vargas et al., 2016; Milián-Suazo et al., 2016) han obtenido resultados similares con respecto al grado de diversidad genética, reportando por encima de 10 espoligotipos diferentes. Los espoligotipos obtenidos en este estudio coinciden con lo reportado por Sandoval-Azuara et al. (2017) y Milián-Suazo et al. (2002), donde los espoligotipos SB0673, SB0971 y SB0140 mostraron la mayor frecuencia. Por el contrario, en este estudio se encontró que tres grupos genéticos fueron exclusivos de Baja California (Grupos H-I-J); esta regionalización de cepas no se había reportado en estudios previos. Una posible explicación para este fenómeno es que estas cepas hayan sido introducidas a Baja California desde Estados Unidos, ya que el SB0145 ha sido reportado como dominante en ese país (Milián-Suazo et al., 2016), que se encuentra inmediatamente al norte y mantiene relaciones comerciales para la exportación de vaquillas hacia Baja California. Otra explicación puede ser un posible aislamiento y

consecuente evolución de las cepas a través del tiempo en esta región. Como se observa en la Figura 2, el espoligotipo SB0145 se originó a partir del espoligotipo SB0140; si la restricción en la movilización de ganado evitó que hubiera flujo de cepas entre Baja California y el resto del país, ya que Sonora es libre de la enfermedad y por lo tanto el movimiento de ganado a través de este estado es muy controlado, esto pudo actuar como una barrera aislante para la enfermedad en esta región.

La Campaña Nacional Contra la Tuberculosis Bovina establece restricciones en la movilización de animales de acuerdo al estatus sanitario de su zona de origen, prohibiendo el paso de animales infectados a zonas libres de enfermedad. Actualmente, Sonora mantiene un estatus zoonosanitario de Acreditado Modificado Avanzado, el cual es considerado como la casi nula presencia de TBb por la Secretaría de Agricultura de los EE.UU. (USDA); mientras que BCA, JAL, GTO, QRO, EDOMX, HGO y AGS presentan el estatus de nivel más bajo como No Acreditado. Esto sugiere que Sonora ha actuado como una barrera geográfica que ha permitido la evolución aislada de las cepas del norte con respecto a las cepas del centro, ocasionando la diferenciación genética visible en la filogenia obtenida en este estudio. Este tipo de regionalización de cepas también fue reportada por Hauer et al. (2015), donde zonas geográficas específicas mostraron genotipos exclusivos de esas zonas.

El grado de diversidad genética de las cepas de *M. bovis* de México obtenido, aunado a la amplia distribución de la mayoría de los grupos genéticos, sugiere que la infección no ha sido contenida/controlada de manera eficiente en la zona centro del país. En un estudio realizado con *M. tuberculosis* (Gagneux, 2018), se ha visto que el grado de polimorfismo que existe en una población de microorganismos es relativa al grado de infección “reactivada” contra infección “transmitida”, es decir, un bajo nivel de polimorfismo sugiere una transmisión continua, mientras que un alto nivel de polimorfismo (alta diversidad genética) sugiere infección latente. La tuberculosis bovina en México es una infección persistente en el ganado bovino, y

esto puede dar lugar a la evolución de las cepas, lo que a su vez puede ocasionar una creciente diversidad genética del patógeno.

Dentro del alcance de este trabajo, la secuenciación genómica sería útil para la identificación de zonas de alerta que actúan como fuentes de infección, sin embargo, el tamaño de muestra en este caso resultó ineficiente para obtener una clara respuesta al respecto. Otros estudios (Tsao et al., 2014; Price-Carter et al., 2018; Hauer et al., 2019) utilizaron cerca de 1,000 aislados para realizar los análisis y obtuvieron mejores resultados. Sin embargo, un estudio en África (Ghebremariam et al., 2018), con tan sólo 31 aislados lograron caracterizar y diferenciar las cepas regionales. México ha importado ganado de varios lugares del mundo, incluyendo Canadá, Estados Unidos, Reino Unido, etc., lo cual ha contribuido a la introducción de la tuberculosis bovina, y por lo cual encontrar una sola fuente de infección puede ser improbable. Es necesario realizar un análisis más robusto con una mayor cantidad de aislados para poder resolver esta problemática en México.

6.3. Identificación y clasificación de las variantes en el genoma de *M. bovis*

Se utilizaron 282 cepas de *M. bovis* para clasificar los SNP y evaluar el efecto predicho que estas variaciones pudieran tener sobre la expresión génica. Se identificaron un promedio de 431 SNP, con un mínimo de 96 y un máximo de 775 SNP. Considerando que el tamaño del genoma de referencia de *M. bovis* (AF2122/97, NC002945.4) es de 4,345,492 pb, la tasa de variación promedio fue de 1 cada 10,850 pb (5,607 – 45,265 pb) (Figura 4). Con respecto a esto, se observó una tendencia por espoligotipo, ya que las cepas con mayor densidad de SNP (≥ 600) son espoligotipo SB0121, seguido de SB0145, SB0152, SB0295 y UNK3. De manera contraria, las cepas con menor cantidad de SNP se encontraron dentro de los espoligotipos SB0673, SB0971 y SB0140. Los SNP se fijan a través del tiempo y con cada evento de transmisión (Robbe-Austerman, 2017), por lo que es posible deducir que los espoligotipos de mayor frecuencia, y que por ende son transmitidos en mayor grado, tengan la mayor cantidad de SNP. Sin embargo, una célula con

mayor cantidad de mutaciones es menos viable que una célula con pocas o cero mutaciones (Lasserre et al., 2018). Por otro lado, el espoligotipo SB0140 se considera un espoligotipo ancestral, y puesto que los espoligotipos SB0145 y SB0152 provienen de este, y considerando que *M. bovis* es un organismo clonal, es posible que la cantidad de SNP se haya acumulado a lo largo del tiempo.

Dirección General de Bibliotecas UAQ

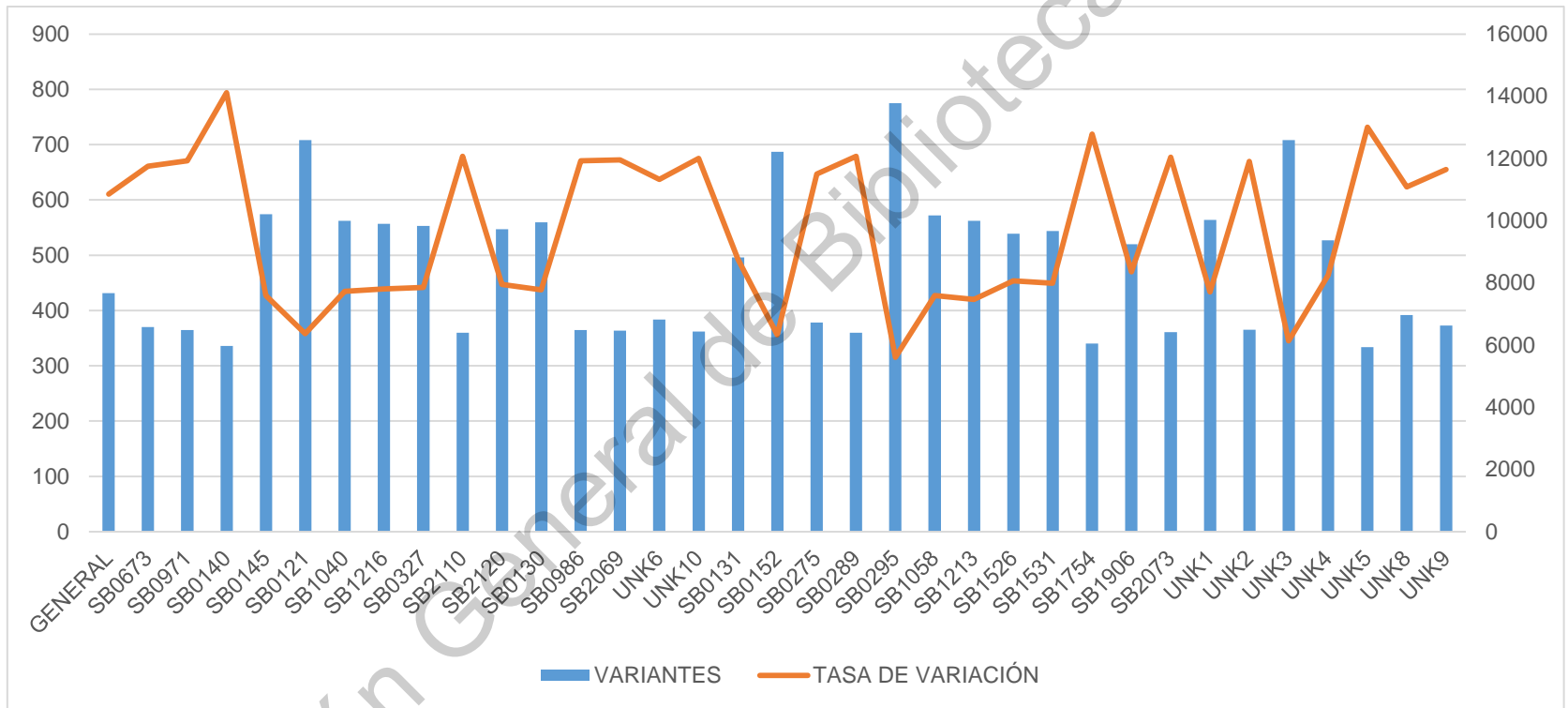


Figura 4. Promedio de variantes identificadas y tasa de variación por espigotipo.

Del total de variantes encontradas, la mayor proporción correspondió a variantes de tipo SNP (85%), sin embargo, también se encontraron variantes de tipo inserción (5%), deleción (5%) y polimorfismos de múltiples nucleótidos (2%) (Figura 5). Se ha reportado que las deleciones e inserciones son factores importantes en la remodelación del genoma micobacteriano (Ho et al., 2000; Mostowy et al., 2005), pero también existe gran influencia de las sustituciones sinónimas sobre la selección genética dentro de una especie (Wang et al., 2013).

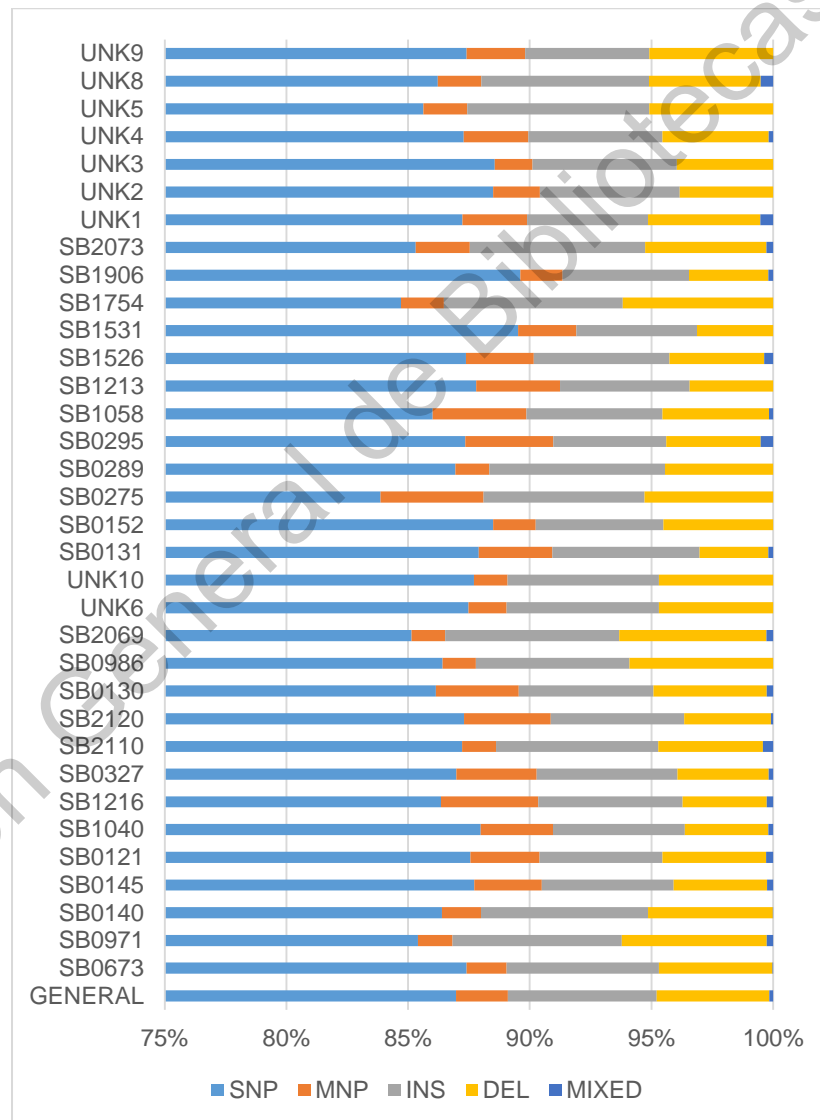


Figura 5. Clasificación de las variantes por tipo.

6.4. Análisis del efecto de las variantes sobre la expresión génica

El efecto de una variante sobre la expresión génica depende del tipo de secuencia que afecta. En este caso, se identificaron siete categorías principales: regiones río arriba (UPSTREAM) y río abajo (DOWNSTREAM) de un gen, regiones intergénicas (INTERGENIC), exones (EXON), dentro de genes (GENE), transcritos (TRANSCRIPT) y sitios de corte y empalme (SPLICE SITE REGION) (Figura 6). Los análisis indicaron que la mayor cantidad de las variantes cayó dentro de zonas río arriba y río debajo de las secuencias codificantes, lo cual representa un posible cambio en la regulación de la expresión génica ya que en estas zonas se encuentran factores de transcripción que regulan los genes adyacentes (García-Pelayo et al., 2009). En mediana proporción se encontró la región de los exones, que representa la secuencia codificante dentro de un gen; eso puede impactar de manera significativa la estructura de la proteína y por ende su función, a menos que el SNP sea sinónimo y no ocasione cambio en el amino ácido correspondiente (Joshi et al., 2012). Y por último, y con la menor proporción, se encontró la región intergénica (entre los genes), para lo cual el riesgo de que una mutación afecte la expresión es mucho menor. Los sitios con menor frecuencia de variaciones fueron los genes, sitios de corte y empalme, y los transcritos. El género *Mycobacterium* mantiene hasta un 95% de similitud genómica, lo cual indica que existe un alto grado de conservación en los genes, esto puede explicar la baja frecuencia de variantes en regiones de alto impacto como estas.

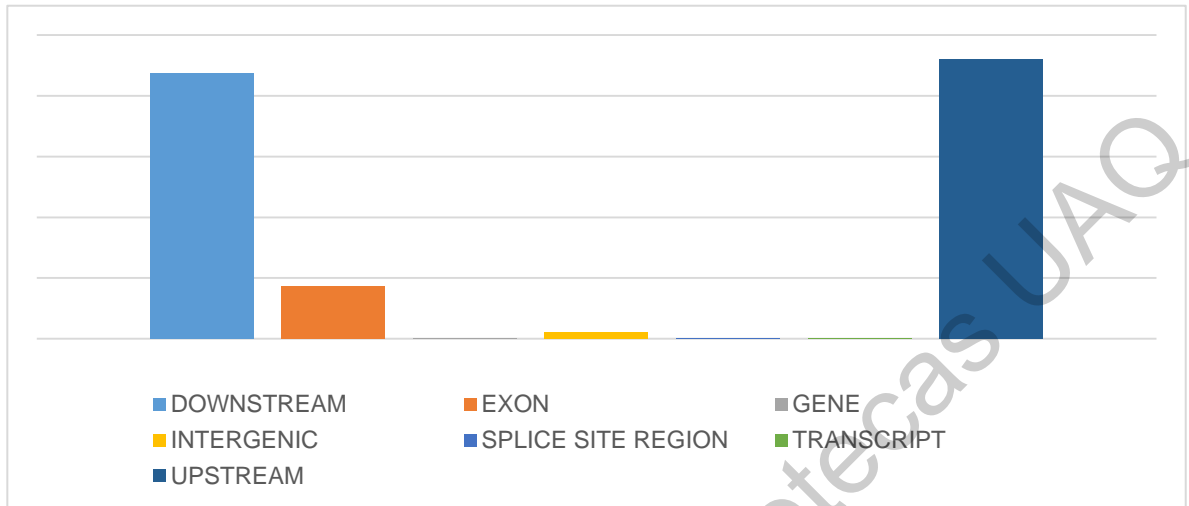


Figura 6. Distribución de las variantes por tipo de secuencia afectada.

Las regiones dentro las cuales pueden encontrarse las variantes pueden ser aún más específicas con respecto al tipo de daño que ocasionan, y se denominan: eliminación del transcrito (transcript_ablation), variante sinónima (synonymous variant), variante no sinónima (nonsynonymous variant), paro retenido (stop_retained_variant), paro perdido (stop_lost), paro adquirido (stop_gained), inicio perdido (start_lost), region de splicing (splice_region_variant), transcrito no codificante (non_coding_transcript_exon_variant), codón de inicio no canónico (non_canonical_start_codon), variante antisentido (missense_variant), variante de intron (intron_variant), region intergénica (intergenic_region), codón iniciador (initiator_codon_variant), fusión génica (gene_fusion), variante de cambio de marco de lectura (frameshift_variant), eliminación de secuencia (feature_ablation), variante río abajo (downstream_gene_variant), inserción disruptiva en marco de lectura (disruptive_inframe_insertion), deleción disruptiva en marco de lectura (disruptive_inframe_deletion), inserción conservativa en marco de lectura (disruptive_inframe_insertion), deleción conservativa en marco de lectura (disruptive_inframe_deletion), y fusión génica bidireccional (bidirectional_gene_fusion) (Figura 7). Nuevamente se puede observar que las

variantes dentro de zonas río arriba y río abajo son las que se encuentran en mayor proporción. En segundo lugar, se identificaron las variantes antisentido, las cuales son variantes no sinónimas y ocasionan algún cambio en el amino ácido correspondiente. En tercer lugar, se encuentran las variantes sinónimas, que no presentan un riesgo significativo puesto que no afecta de forma funcional. A esta le siguen las variantes intergénicas y variantes de cambio de marco de lectura; mientras que las variantes intergénicas pueden no representar peligro alguno para la expresión de las proteínas, un cambio en el marco de lectura sí puede ocasionar la producción defectuosa de alguna proteína. Esto aunado a las variantes antisentido ya mencionadas, puede ser indicativo de cambios en el fenotipo de la micobacteria (Hauer et al., 2019). Aquí, esto se puede observar distribuido de manera general, sin alguna tendencia específica para un espoligotipo en particular.

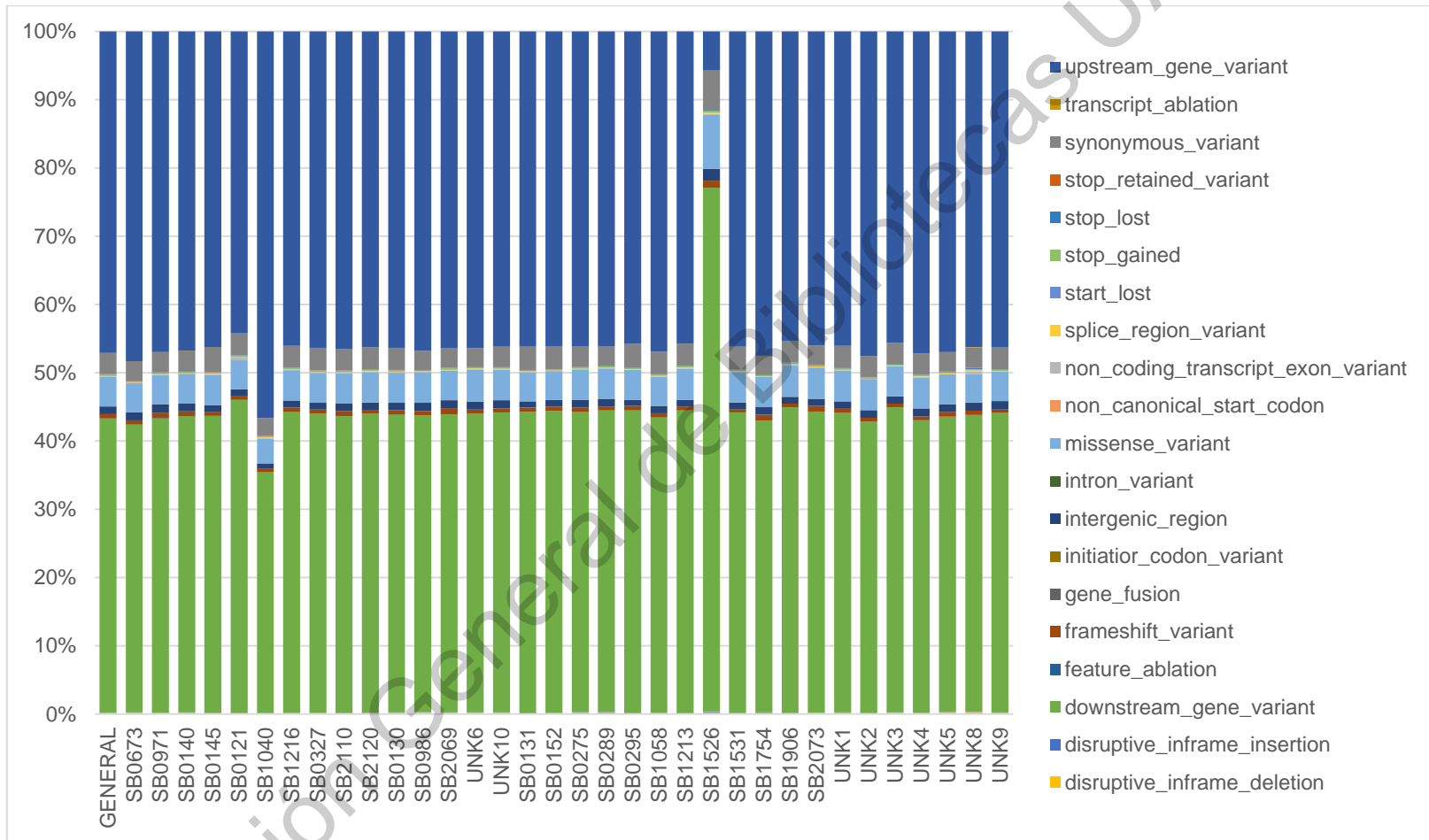


Figura 7. Proporción de las variantes por tipo de secuencia afectada y su posible efecto.

Como una primera aproximación al efecto de las variantes sobre la función génica, estas se clasificaron como antisentido (variantes no sinónimas que ocasionan cambio en el amino ácido correspondiente), sin sentido (variantes que ocasionan un codón de paro prematuro) y silenciosas (variantes sinónimas que no ocasionan cambio en el amino ácido correspondiente) (Figura 8). De manera general en todas las cepas, las variantes de mayor proporción fueron las antisentido, o no sinónimas. Lasserre et al. (2018) y García-Pelayo et al. (2009) reportaron mayor proporción de las variantes sinónimas; sin embargo, Hauer et al. (2019) y Joshi et al. (2012), también encontraron una mayor proporción de variantes no sinónimas con respecto a las sinónimas. La tasa de mutaciones antisentido a mutaciones silenciosas (sustituciones no sinónimas/sustituciones sinónimas), en promedio, fue mayor a 2. García-Pelayo et al. (2009) reportó una tasa de 0.57, mientras que Hauer et al. (2019) 1.77. Cuando esta tasa es igual a 1, se dice que existe selección neutral, cuando es entre 0 y 1 la selección es negativa, pero si es mayor a 1, entonces se considera selección positiva. Consecuentemente, esto sugiere que las regiones donde caen estas variantes se están heredando a la siguiente generación (García-Pelayo et al., 2009). Las variantes sin sentido se encuentran en menor proporción, estas alteraciones pueden ocasionar codones de paro prematuros que ocasionan la interrupción de la formación de una proteína. Por lo tanto, este tipo de variantes no son deseadas ya que no contribuyen a la supervivencia de la célula.

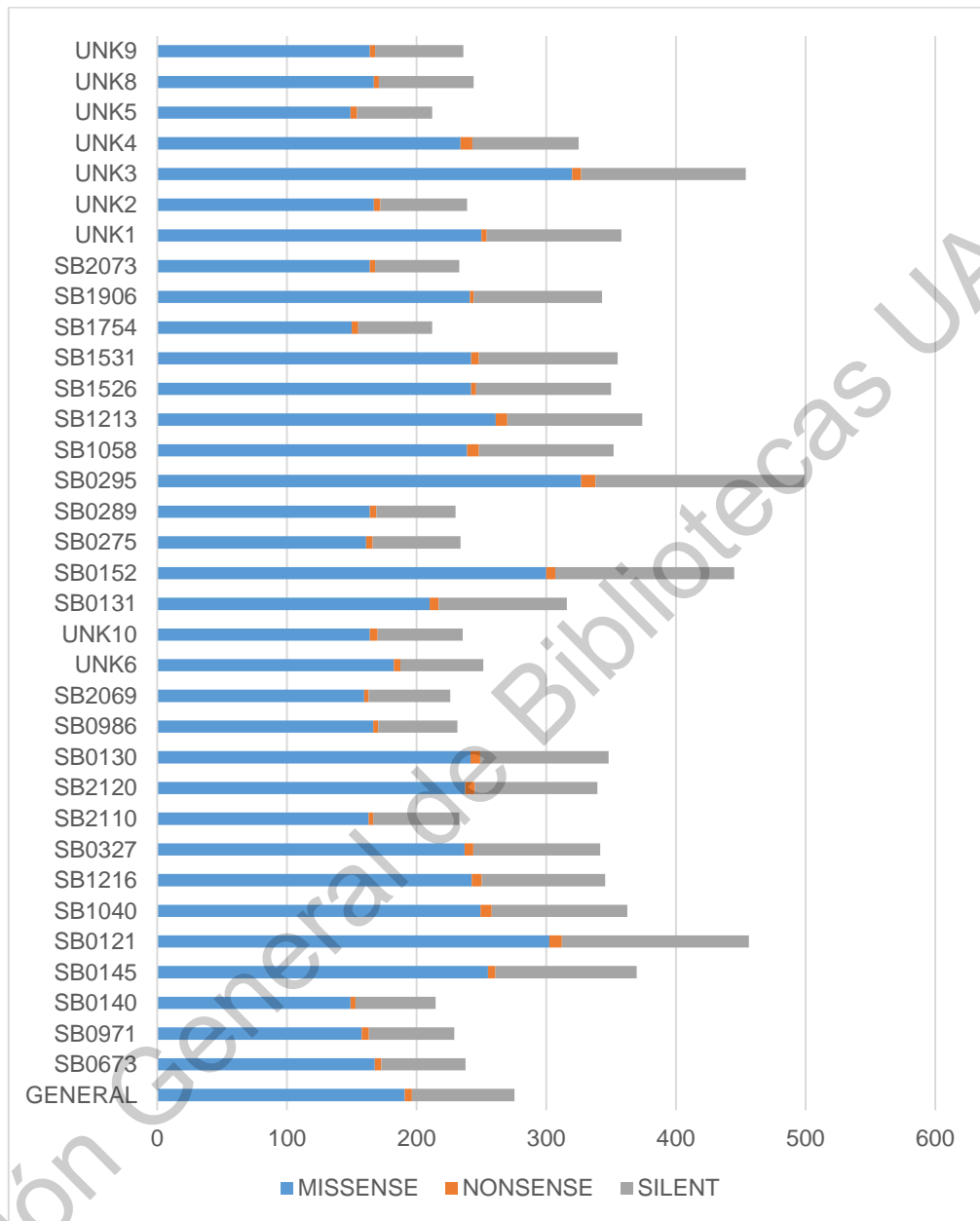


Figura 8. Clasificación de las variantes por su efecto sobre el amino ácido: no sinónimas (missense), sin sentido (nonsense), sinónimas (silent).

Adicionalmente, con respecto al efecto predicho sobre la expresión génica, este se clasificó como bajo (LOW), moderado (MODERATE) y alto (HIGH) (Figura 9). Se encontró que para la mayor cantidad de variantes se predijo un daño moderado, en segundo lugar se encontraron las variantes de daño bajo, y por último las variantes de alto impacto. Esto es lo contrario a aquello reportado por Thacker et al. (2015), en donde la mayoría de las variantes reportadas fueron de bajo impacto. Hauer et al. (2019) también reportaron una pequeña cantidad de SNP asociados a genes de virulencia donde su impacto puede estar ocasionando un cambio en el fenotipo de la micobacteria. Un estudio más detallado realizado por Abdelaal et al. (2019), reportó la asociación de polimorfismos genómicos a la aparición de virulencia en una cepa de Egipto.

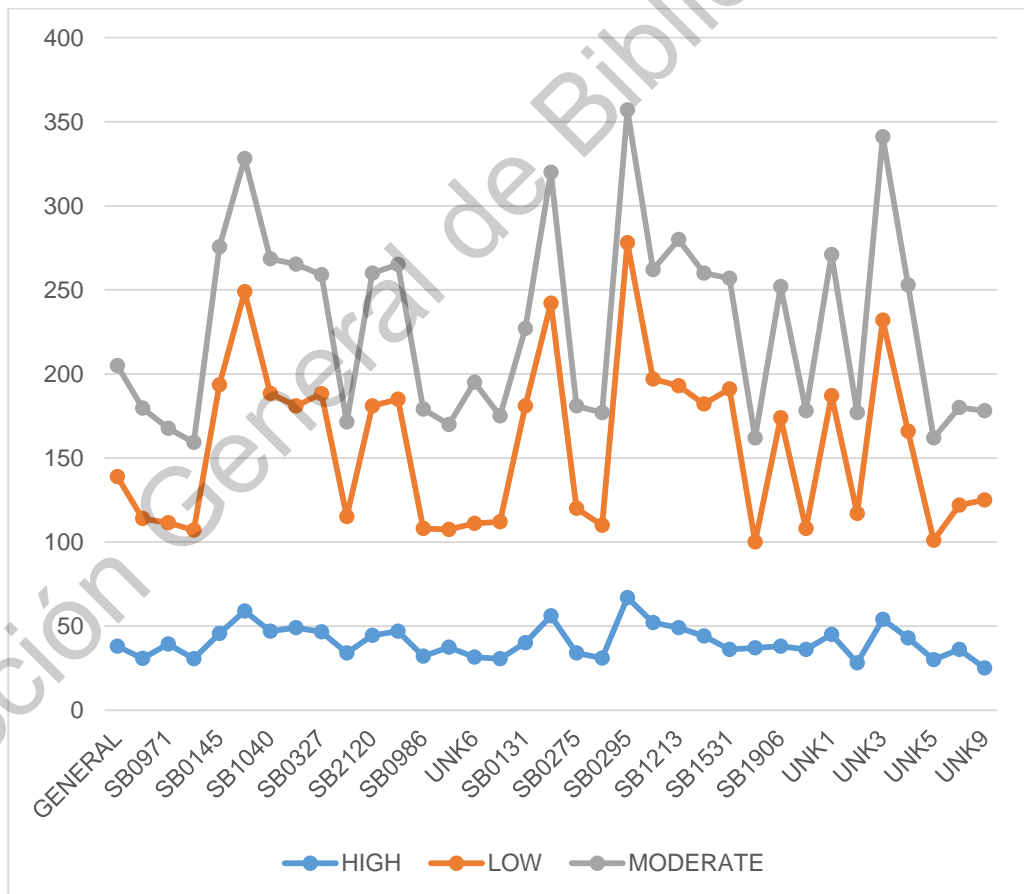


Figura 9. Clasificación de las variantes por su efecto sobre la función proteica.

En un análisis más específico, se calculó la variabilidad del cambio de bases (adenina-A, timina-T, citosina-C, guanina-G) ocasionado por los SNP (Figura 10). Los cambios de tipo A→G, C→T, G→A, y T→C fueron los más frecuentes, mientras que los cambios de tipo A→C, C→A, G→C, G→T, T→A, y T→G ocurrieron con baja frecuencia.

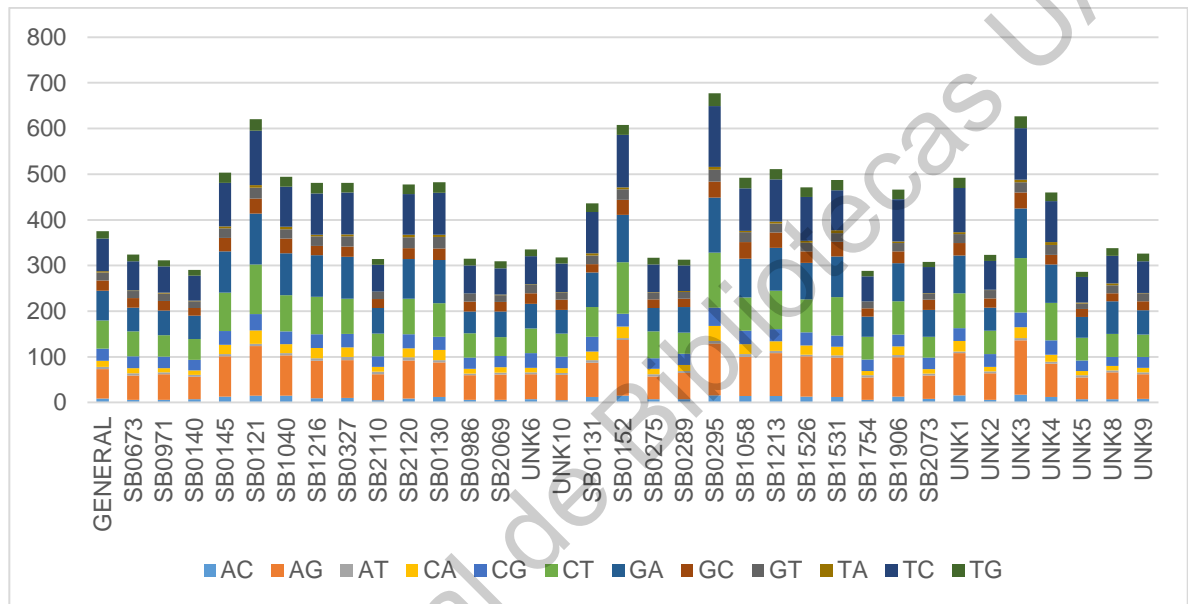


Figura 10. Cantidad de variantes por el tipo de cambio de bases.

Así mismo, se evaluó la tasa de transiciones sobre transversiones (TsTv). Una transversión, que es el cambio de una purina por una pirimidina, o viceversa, es considerada como de mayor gravedad en comparación con una transición, que es el cambio entre pirimidinas o purinas. Esto se debe a la diferencia en cuanto a la estructura química de las bases nitrogenadas. Consecuentemente, las transversiones pueden ocasionar inestabilidad en la estructura del ADN ya que la complementariedad de las bases puede verse afectada (Karp, 2019). En este estudio se logró observar que ciertos esplogitipos presentan tasas de TsTv más altas, tales como SB0121, SB0145, SB0152, SB0295, y UNK3 (Figura 11). La tasa TsTv general obtenida fue de 2.32, con un rango entre 2.02 y 2.82.

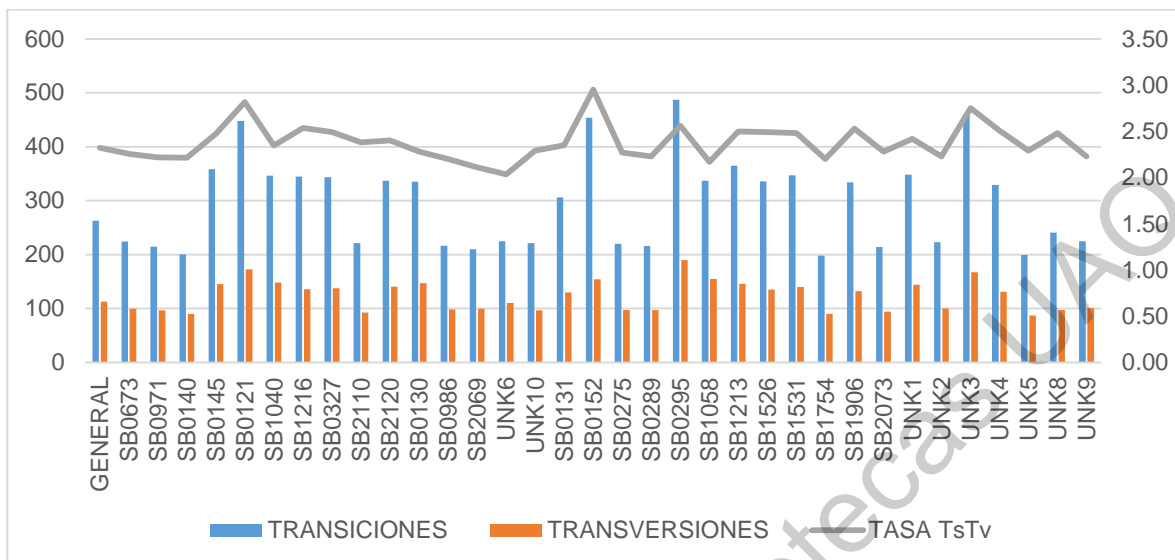


Figura 11. Tasa de transiciones y transversiones totales.

Existen algunos reportes donde las mutaciones encontradas en los genomas de las micobacterias tienen como consecuencia algún cambio en la expresión génica del microorganismo y con esto un cambio fenotípico del patógeno. Por ejemplo, Bigi et al. (2019) reportaron la existencia de polimorfismos (no sinónimos) en nueve proteínas de virulencia en cepas de *M. bovis* de Argentina que ocasionaron cambios en la síntesis de tioserol dimicoserostato (PDIM), lo cual disminuye la capacidad de la micobacteria para evadir al fagosoma. Modipane et al. (2019) encontraron mutaciones en el genoma de *M. bovis* BCG que sugieren una resistencia a antibióticos. Finalmente, Dutta et al. (2019) reportaron una mutación en el factor de transcripción sigF en el genoma de *M. bovis* que ocasiona cambios en los antígenos de superficie, disminuyendo su patogenicidad. En este estudio los resultados sobre el efecto de la función proteica están basados en las predicciones de algoritmos bioinformáticos, sería necesario llevar a cabo análisis más detallados con las cepas que tuvieron las mutaciones de más alto impacto para corroborar dichas predicciones.

6.5. Análisis del pangenoma

En total, se calcularon 3,577 genes que caen dentro del pangenoma, lo cual hace referencia al total de genes encontrados en las 318 cepas de *Mycobacterium bovis* incluidas en el análisis. En general, se ha reportado un total de secuencias codificantes para *M. bovis* de 3,951 (Garnier et al., 2003). Lasserre et al. (2018) reportaron 2,370 genes como su genoma central; mientras que Zimpel et al. (2017) reportaron un total de 4,216 genes. Esto significa que puede existir un rango de hasta 1,000 genes que pudieran considerarse como “accesorios”, y que la micobacteria utiliza para adaptarse al gran rango de hospedadores que posee (Smith et al., 2001). El pangenoma se puede clasificar en tres tipos de genes: 1) Genes core son los genes contenidos en todos los genomas, 2) Genes soft-core son los genes encontrados en 95% de los genomas incluidos, 3) Genes shell son aquellos que se encuentran entre 15 y 95% de los genomas (Cuadro 7).

Cuadro 7. Distribución de los genes que forman parte del pangenoma.

Genes core (99-100%)	3,577
Genes soft-core (95-99%)	209
Genes shell (15-95%)	412

Como genoma central (o genoma “core”), genes que se encuentran en todos y cada uno de los genomas incluidos en este estudio, se calcularon 2,188 genes. Es decir, como mínimo, todos los genomas contienen y comparten estos genes. En la Figura 12 se puede observar el número de genes “accesorios” (izquierda) y el número de genes que conforman el pangenoma (derecha) de esta colección de cepas de *M. bovis*.

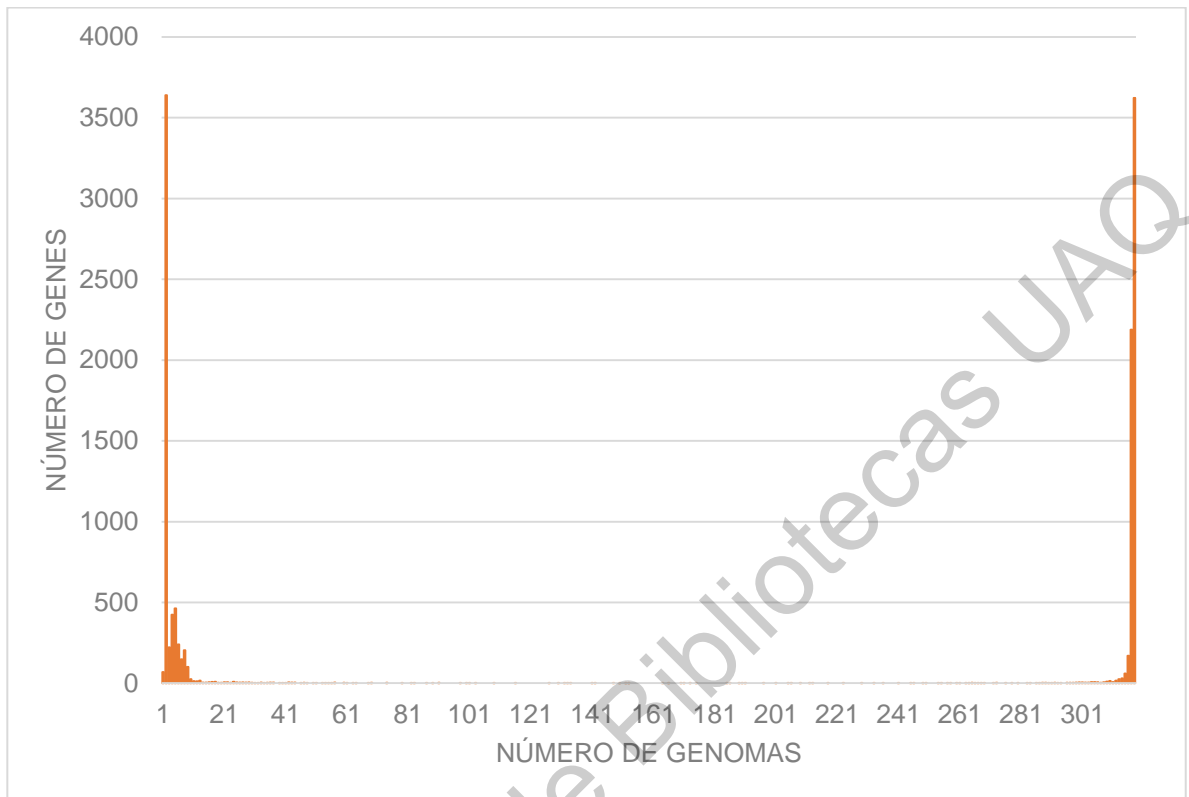


Figura 12. Frecuencia de genes y número de genomas.

En la Figura 13 se puede observar que el número de genes “nuevos” no es alto, sin embargo, su existencia sugiere que *M. bovis* es capaz de adquirir genes dependiendo del “estilo de vida” de la bacteria (Tettelin et al., 2005). Más evidencia que apoya la idea de que *M. bovis* tiene un pangenoma abierto la describe Rouli et al. (2015), que reportaron que para saber si un pangenoma es cerrado es necesario incluir tan solo 6 genomas en el análisis, mientras que para un pangenoma abierto 10 genomas son suficientes. Pero puede haber la posibilidad de pangenomas infinitos, donde la cantidad de genomas necesarios es difícil de determinar, como es el caso de *E. coli*.

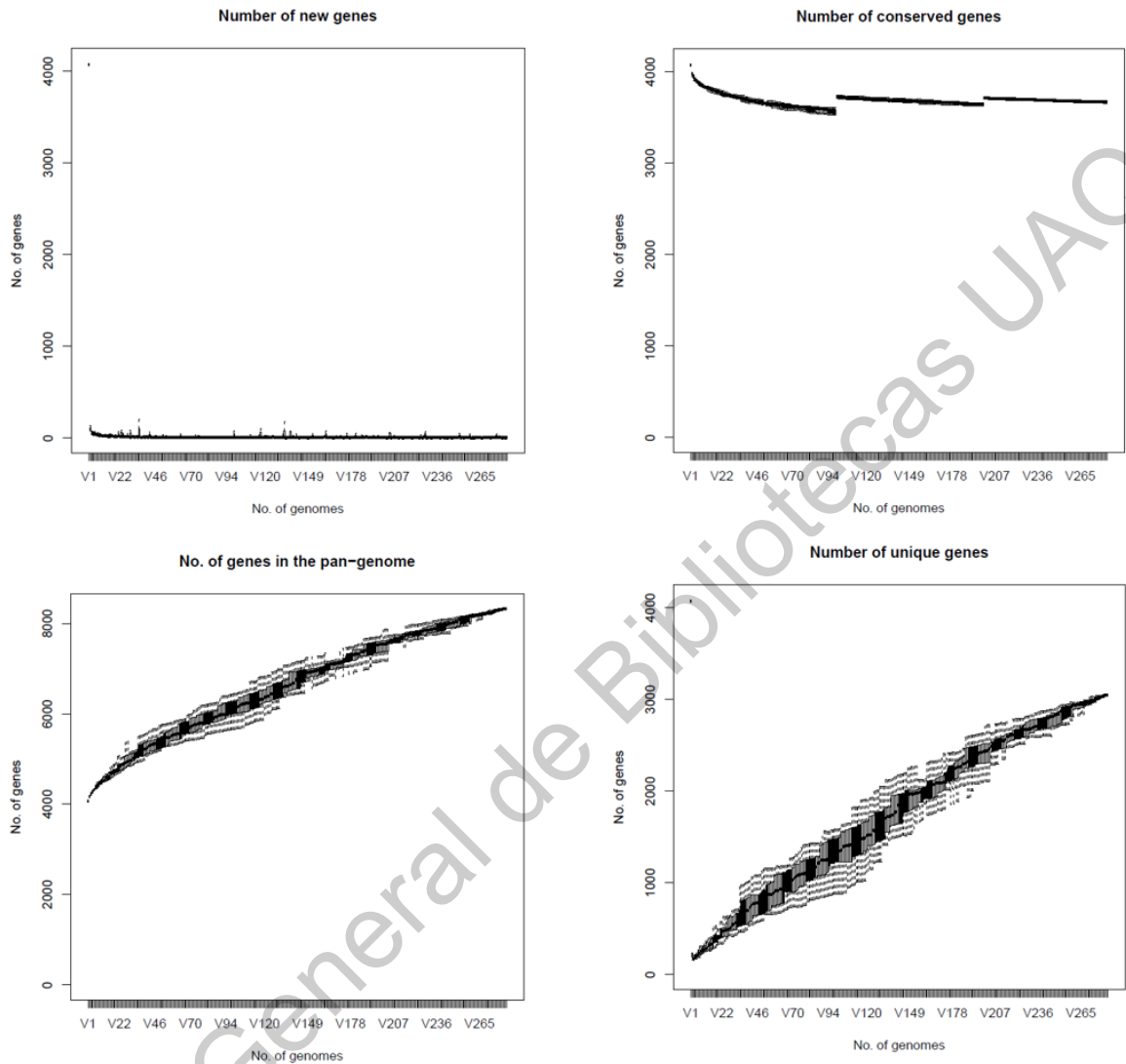


Figura 13. Esquina superior izquierda: Número de genes nuevos con respecto al pangenoma; Esquina superior derecha: Número de genes conservados con respecto al pangenoma; Esquina inferior izquierda: Número total de genes que conforman el pangenoma; Esquina inferior derecha: Número de genes únicos encontrados en el pangenoma.

6.6. Análisis estadístico

Se analizó la similitud de los genomas de las cepas de *M. bovis* en México por medio de un análisis de componentes principales con base a los SNP encontrados, así como por medio de un análisis de distancias de Jaccard (mapa de

calor), para evaluar el grado de diferenciación genética en general (Figura 14). Los resultados dan evidencia de tres grupos principales. Aunque en la filogenia se obtuvieron 12 grupos genéticos, estos se derivan de tres grandes clados (Figura A). Hasta ahora se conocen cuatro Complejos Clonales (CC) de *M. bovis* en el mundo (Smith, 2012; Rodrigues-Campos et al., 2012). Es posible deducir que esta agrupación se debe a los distintos CC a los que pertenecen las cepas, sin embargo, Bravo-Sales et al. (2019) reportaron haber encontrado cepas que no pertenecen a ninguno de los cuatro CC ya conocidos. Laserre et al. (2018) reportó algo similar con cepas de Uruguay, donde también demostraron pertenecer a tres grupos diferentes. Sería necesario realizar análisis adicionales para resolver esto con más detalle para las cepas mexicanas.

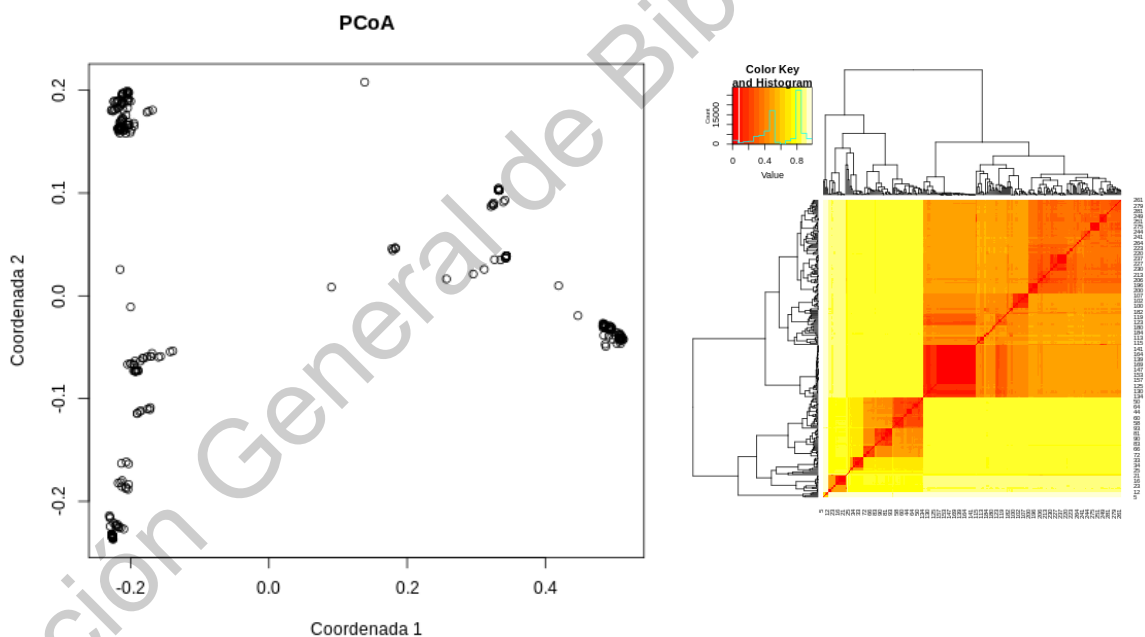


Figura 14. Análisis de componentes principales y mapa de calor representan la similitud genética de los genomas las cepas de *M. bovis* de México.

VII. CONCLUSIONES

La tuberculosis bovina en México es una enfermedad endémica del ganado bovino que también infecta a la población humana. *Mycobacterium bovis* es el agente causal y guarda relación genética muy estrecha con *Mycobacterium tuberculosis*, principal causante de la tuberculosis en el ser humano, y otras especies del género, en particular aquellas que pertenecen al Complejo *Mycobacterium tuberculosis*. Los métodos principales para su genotipificación y caracterización en México son la espiligotipificación, de manera general para realizar vigilancia epidemiológica, y la secuenciación genómica, como método auxiliar para aumentar la discriminación entre cepas.

En cuanto a la caracterización genética de *M. bovis* en México, no se encontró evidencia de con respecto a la existencia de cepas específicas de humano o de bovino, y en este último caso, tampoco se observaron cepas específicas de ganado lechero o cárnico. Sin embargo, sí se observó regionalización de las cepas, es decir, existen cepas que son exclusivas de Baja California. Las cepas de la zona centro del país no mostraron regionalización por estado. De manera general, el grado de diversidad genética observado aquí es alto, obteniendo 39 diferentes espiligotipos a partir de 320 cepas. Los espiligotipos más frecuentes fueron SB0971, SB0673, SB0145, SB0140 y SB0121. A partir de la genotipificación con base en los SNP obtenidos mediante secuenciación genómica, se establecieron 12 grupos genéticos, A a L, siendo los más frecuentes D y E, los cuales corresponden adecuadamente a cepas con espiligotipos SB0971 y SB0673.

Consecuentemente, el análisis por medio de espiligotipificación tuvo correspondencia con el análisis realizado a base de secuenciación genómica. Esto es importante puesto que da la pauta para mantener vigente la espiligotipificación, particularmente en países donde los recursos económicos no permiten el uso de secuenciación genómica de manera rutinaria. La secuenciación genómica sí dio mayor discriminación entre las cepas, con lo cual fue posible asociar los eventos de transmisión bovino-humano de manera directa, sin embargo; el tamaño de muestra

utilizado en este estudio no fue suficiente para demostrar posibles rutas de diseminación y focos de infección entre el ganado bovino.

La diversidad genómica de las cepas de *M. bovis* de México se logró observar mediante la caracterización de SNP y su pangenoma. La mayor proporción de los SNP se encontró dentro de zonas como exones, regiones río arriba y regiones río debajo de los genes, donde es posible que se vea comprometida la estructura proteica y por ende su función. Adicionalmente, la mayor proporción de los SNP se clasificó como antisentido, y con respecto al efecto predicho sobre la función proteica, la mayor proporción fueron de daño moderado. Además, este conjunto de cepas constituyeron un pangenoma abierto, con un conjunto mínimo de 3,687 genes y una cantidad variable de genes accesorios de hasta 5,000. Estos incluyeron genes con funciones involucradas en la síntesis de amino ácidos y sus derivados, metabolismo de ARN, virulencia y defensa.

Mycobacterium bovis es un patógeno que ha sido capaz de infectar una muy amplia variedad de mamíferos y diseminarse hacia diferentes regiones geográficas a nivel mundial. Es posible que dentro de las estrategias utilizadas por parte de este patógeno para adaptarse a tan variables ambientes, tanto de huésped como geográficos, la remodelación de su genoma sea importante. Identificar las alteraciones genéticas que se fijan en *M. bovis* puede ayudar a comprender su dinámica de transmisión. Consecuentemente, si se desea disminuir la prevalencia de la tuberculosis bovina es necesario evitar esta transmisión. Además, el riesgo que representa para la salud pública debe mantenerse relevante en los esfuerzos por controlar y eliminar esta enfermedad.

VIII. BIBLIOGRAFÍA

Abdelaal H.F.M., Spalink D., Amer A., Steinberg H., Hashish E.A., Nasr E.A., Talaat A.M. 2019. Genomic polymorphism associated with the emergence of virulent isolates of *Mycobacterium bovis* in the Nile Delta. *Sci. Rep.* 9: 11657.

Acosta-Salinas R., Estrada-Chávez C., Milián-Suazo F. 2009. Tipificación de cepas de *Mycobacterium bovis*. *Revisión. Téc. Pecu. Méx.* 47(4): 389-412.

Adzhubei I.A., Schmidt S., Peshkin L., Ramensky V.E., Gerasimova A., et al. 2010. A method and server for predicting damaging missense mutations. *Nat. Methods* 7: 248–249.

Aguilar-Bultet L. y Falquet L. 2015. Secuenciación y ensamblaje de novo de genomas bacterianos: una alternativa para el estudio de nuevos patógenos. *Rev. Salud Anim.* 37(2): 125 – 132.

Allen A.R., Dale J., McCormick C., Mallon T.R., Costello E., et al. 2013. The phylogeny and population structure of *Mycobacterium bovis* in the British Isles. *Infect. Genet. Evol.* 20: 8-15.

Álvarez A.H., Estrada-Chávez C., Flores-Valdez M.A. 2009. Molecular findings and approaches spotlighting *Mycobacterium bovis* persistence in cattle. *Vet. Res.* 40(22).

Arce M.A., Morales S.C.P., Solís S.J. 2008. Tuberculosis: mecanismos de defensa, inmunopatogénesis y biomarcadores de susceptibilidad y resistencia. *Ciencia UANL* 11(3).

Aziz R.K., Bartels D., Best A.A., DeJongh M., Disz T., et al. 2008. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics.* 9: 75.

Bankevich A., Nurk S., Antipov D., Gurevich A., Dvorkin M., et al. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* 19(5): 455 – 477.

Brosch R., Gordon S.V., Marmiesse M., Brodin P., Buchrieser C., et al. 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci. U.S.A.* 99: 3684-3689.

Brunton L.A., Nicholson R., Ashton A., Alexander N., Wint W., et al. 2015. A novel approach to mapping and calculating the rate of spread of endemic bovine tuberculosis in England and Wales. *Spatial Spatio-temporal Epidemiol.* 13:41-50.

Bryant J.M., Schurch A.C., van Deutekom H., Harris S.R., de Beer J.L., et al. 2013.

Inferring patient to patient transmission of *Mycobacterium tuberculosis* from whole genome sequencing data. BMC Infect. Dis. 13: 110.

Cadmus S.I.B., Gordon S.V., Hewinson R.G., Smith N.H. 2011. Exploring the use of molecular epidemiology to track bovine tuberculosis in Nigeria: an overview from 2002 to 2004. Vet. Microbiol. 151: 133-138.

Cingolani P., Platts A., Wang le L., Coon M., Nguyen T., Wang L., Land S.J., Lu X., Ruden D.M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly. 6(2): 80 – 92. Disponible en: http://snpeff.sourceforge.net/SnpEff_paper.pdf.

Deatherage D.E. y Barrick J.E. 2014. Identification of mutations in laboratory evolved microbes from next-generation sequencing data using *breseq*. Methods Mol. Biol. 1151: 165 – 188.

DePristo M.A., Banks E., Poplin R., Garimella K.V., Maguire J.R., et al. 2011. A framework for variation discovery and genotyping using nextgeneration DNA sequencing data. Nat. Genet. 43: 491–498.

Diamond J. 2002. Evolution, consequences and future of plant and animal domestication. Nature. 418: 700-707.

Faksri K., Xia E., Tan J.H., Teo Y.Y., Ong R.T.H. 2016. *In silico* region of difference (RD) analysis of *Mycobacterium tuberculosis* complex from sequence reads using RD-Analyzer. BMC Genomics. 17(847).

Fernandes-Ramos D., Tavares L., Almeida da Silva P.E., Dellagostin O.A. 2014. Molecular typing of *Mycobacterium bovis* isolates: A review. Brazilian J. Microbiol. 45(2): 365-372.

Ford C.B., Lin P.L., Chase M.R., Shah R.R., Iartchouk O., et al. 2011. Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. Nat. Genet. 43: 482–486.

Gagneux S. 2018. Ecology and evolution of *Mycobacterium tuberculosis*. Nat Rev Microbiol. 16: 202-213.

Garcia-Pelayo M.C., Uplejar S., Keniry A., Mendoza-Lopez P., Garnier T., et al. 2009. A comprehensive survey of single nucleotide polymorphisms (SNP) across *Mycobacterium bovis* strains and *M. bovis* BCG vaccine strains refines the genealogy and defines a minimal set of SNP that separate virulent *M. bovis* strains and *M. bovis* BCG strains. Infect. Immun. 77(5): 223-2238.

Garnier T., Eiglemeier K., Camus J.C., Medina N., Mansoor H., et al. 2003. The

- complete genome sequence of *Mycobacterium bovis*. PNAS. 100(13): 7877-7882.
- Golby P., Nunez J., Witney A., Hinds J., Quail M.A., et al. 2013. Genome level analyses of *Mycobacterium bovis* lineages reveal role of SNP and antisense transcription in differential gene expression. BMC Genomics 14: 710.
- Gormely E. y Corner L.A.L. 2018. Pathogenesis of *Mycobacterium bovis* infection: the badger model as a paradigm for understanding tuberculosis in animals. Front. Vet. Sci. 4: 247.
- Gurevich A., Saveliev V., Vyahhi N., Tesler G. 2013. QUASt: quality assessment tool for genome assemblies. Bioinformatics. 29(8): 1072 – 1075.
- Guta S., Casal J., Napp S., Saez J.L., Garcia-Saenz A., et al. 2014. Epidemiological investigation of bovine tuberculosis herd breakdowns in Spain 2009/2011. PLoS ONE. 9(8): e104383.
- Gutiérrez-Reyes J.A., García-Casanova L., Romero-Torres C., Sosa-Gallegos S., Cantó-Alarcón G.J., et al. 2012. Population structure of *Mycobacterium bovis* isolates from cattle in Mexico. Prev. Vet. Med. 106: 1-8.
- Hauer A., De Cruz K., Cochard T., Godreuil S., Karoui C., et al. 2015. Genetic evolution of *Mycobacterium bovis* causing tuberculosis in livestock and wildlife in France since 1978. PLoS ONE. 10(2).
- Ho T.B.L., Robertson B.D., Taylor G.M., Shaw R.J., Young D.B. 2000. Comparison of *Mycobacterium tuberculosis* genomes reveals frequent deletions in a 20 kb variable region in clinical isolates. Yeast. 17: 272-282.
- Joshi D., Harris N.B., Waters R., Thacker T., Mathema B., et al. 2012. Single nucleotide polymorphisms in the *Mycobacterium bovis* genome resolve phylogenetic relationships. J. Clin. Microbiol. 50 (12): 3853-3861.
- Joshi D.J. 2012. Population genetic frameworks and functional genomics of *Mycobacterium bovis* (Tesis de Doctorado). University of Minnesota.
- Kamerbeek J., Schouls L., Kolk A., van Agterveld M., van Soolingen D., et al. 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J. Clin. Microbiol. 35: 907-914.
- Kato-Maeda M., Ho C., Passarelli B., Banaei N., Grinsdale J., et al. 2013. Use of whole genome sequencing to determine the microevolution of *Mycobacterium tuberculosis* during an outbreak. PLoS ONE 8: e58235.
- Li H. y Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25: 1754–60.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25: 2078 – 9.

Ludannyy R., Figueroa-Alvarez M., Levi D., Markelov M., Dedkov V., et al. 2015. Whole-genome sequence of *Mycobacterium bovis* BCG-1 (Russia). *Genome Announcements*. 3(6).

Majlessi L., Prados-Rosales R., Casadevall A., Brosch R. 2015. Release of mycobacterial antigens. *Immun. Rev.* 264: 25-45.

McInerney J.O., McNally A., O'connell M.J. 2017. Why prokaryotes have pangenomes. *Nat. Microbiol.* 2:17040.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–1303.

Milián-Suazo F., García-Casanova L., Romero-Torres C., Cantó-Alarcón G.J., Gutiérrez-Reyes J.A., et al. 2012. Diversidad genética y distribución regional de cepas de *Mycobacterium bovis* del ganado en México. *Rev. Mex. Cien. Pecu.* 3(4): 459-471.

Milián-Suazo F., Rubio-Venegas Y., Cantó-Alarcón G.J., Olver-Ramírez M.A., Pizano-Martínez O., et al. 2013. La tuberculosis bovina en México: Las Bases. Primera Edición. INIFAP. Querétaro, México.

More S.J., Radunz B., Glanville R.J. 2015. Lessons learned during the successful eradication of bovine tuberculosis from Australia. *Vet. Rec.* 177: 224-232.

Mostowy S, Inwald J., Gordon S., Martin C., Warren R., Kremer K., Cousins D., Behr M.A. 2005. Revisiting the evolution of *Mycobacterium bovis*. *J. Bacteriol.* 187(18): 6386-6395.

Mostowy S., Cousins D., Brinkman J., Aranaz A., Behr M.A. 2002. Genomic deletions suggest a phylogeny for the *Mycobacterium tuberculosis* complex. *J. Infect. Dis.* 186: 74-80.

Muller B., Steiner B., Bonfoh B., Fane A., Smith N.H., et al. 2008. Molecular characterisation of *Mycobacterium bovis* isolated from cattle slaughtered at the Bamako abattoir in Mali. *BMC Vet. Res.* 4(26).

Naranjo V., Gortazar C., Villar M., de la Fuente J. 2007. Comparative genomics and proteomics to study tissue-specific response and function in natural *Mycobacterium bovis* infections. *Animal Health Res. Rev.* 8(1): 81-88.

Navarro Y., Romero B., Copano M.F., Bouza E., Domínguez L., et al. 2015. Multiple

sampling and discriminatory fingerprinting reveals clonally complex and compartmentalized infections by *M. bovis* in cattle. *Vet. Microbiol.* 175: 99-104.

Nava-Vargas A. 2015. Caracterización genética de aislados de *M. bovis* obtenidos a partir de ganado bovino de México. Tesis de Maestría. Facultad de Ciencias Naturales. Universidad Autónoma de Querétaro.

Oksanen J., Guillaume Blanchet F., Kindt R., Legendre P., Minchin P.R., et al. 2012. *vegan: Community Ecology Package*, R package version 2.0-5. Disponible en: <http://CRAN.Rproject.org/package=vegan>.

Organización Mundial de la Salud. 2015. Informe mundial sobre la tuberculosis. Disponible en: http://www.who.int/tb/publications/global_report/es/.

Organización Panamericana de la Salud. 2008. Manual para el diagnóstico bacteriológico de la tuberculosis, Parte 2. Disponible en: http://www.paho.org/hq/index.php?option=com_docman&task=doc_details&gid=17477&Itemid=270&lang=es. Accedido el 19 de mayo de 2016.

Page A.J., Cummins C.A., Hunt M., Wong V.K., Reuter S., et al. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics.* 31(22): 3691 – 3693.

Pontiroli A., Khera T. T., Oakley B.B., Mason S., Dowd S.E., Travis, et al. 2013. Prospecting environmental mycobacteria: combined molecular approaches reveal unprecedented diversity. *PLoS ONE* 8(7): e68648.

R Core Team. 2012. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Robbe-Austerman, S. 2014. *Mycobacterium bovis* whole genome sequencing SNP analysis. United States Department of Agriculture, National Veterinary Services Laboratories, Animal and Plant Health Inspection Service.

Robinson J.T., Thorvaldsdottir H., Winckler W., Guttman M., Lander E.S., et al. 2011. Integrative genomics viewer. *Nat. Biotechnol.* 29: 24–26.

Rodrigues-Campos S., Romero B., de Juan L., Bezos J., Alvarez J., et al. 2012. The population structure of *Mycobacterium bovis* in Spain. 33rd Annual Congress of the European Society of Mycobacteriology, European Society of Mycobacteriology, Brasov, Rumania, 4 de julio de 2012. Disponible en: <https://www.visavet.es/es/the-population-structure-of-mycobacterium-bovis-in-spain/34=1081/>.

Rodrigues-Campos, S., Smith, N. H., Boniotti, M. B., Aranaz, A. 2014. Overview and phylogeny of *Mycobacterium tuberculosis* complex organisms. *Res. Vet. Sci.* 97: S5-S19.

Rodríguez-Hernández E., Pizano-Martínez O.E., Milián-Suazo F., Cantó-Alarcón G.J., Anaya-Escalera A.M., et al. 2014. Manual de procedimientos para la identificación molecular y genotipificación de *M. bovis*. CENIDF y MA. Primera Edición. Querétaro, México.

Rouli L., Merhej V., Fournier P.E., Raoult D. 2015. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect.* 7: 72-85.

Sales É.B., de Alencar A.P., Hodon M.A., Filho P.M.S., de Souza-Filho A.F., Lage A.P., Heinemann M.B., Junior A.A.F. 2019. Identification of clonal complexes of *Mycobacterium bovis* in Brazil. *Arch. Microbiol.* 201: 1047–1051 (2019).

Sanou A., Tarnagda Z., Kanyala E., Zingué D., Nouctara M., et al. 2014. *Mycobacterium bovis* in Burkina Faso: Epidemiologic and genetic links between human and cattle isolates. *PLoS Negl. Trop. Dis.* 8(10): e3142.

Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 30(14): 2068 – 2069.

Servicio de Información Agroalimentaria y Pesquera. 2018. Resumen Nacional de Población Ganadera. Disponible en: https://www.gob.mx/cms/uploads/attachment/file/516347/Inventario_2018_Resumen.pdf.

Servicio Nacional de Sanidad, Inocuidad y Calidad Agroalimentaria. 2018. Reporte del 01 de enero al 31 de marzo de 2018. Disponible en: <https://www.gob.mx/pronabive/articulos/tuberculosis-bovina-en-mexico>.

Servicio Nacional de Sanidad, Inocuidad y Calidad Agroalimentaria. 2019. Acciones y programas, Campañas zoonosológicas, Tuberculosis bovina, Indicadores, 2017. Disponible en: https://www.gob.mx/cms/uploads/attachment/file/326595/3_TBPrevalencias_TB_diembre_2017.pdf.

Servicio Nacional de Sanidad, Inocuidad y Calidad Agroalimentaria. 2019. Documentos, Situación actual de la tuberculosis bovina. Disponible en: <https://www.gob.mx/senasica/documentos/situacion-actual-de-tuberculosis-bovina>.

Smith N.H. 2012. The global distribution and phylogeography of *Mycobacterium bovis* clonal complexes. *Infect. Genet. Evol.* 12: 857-865.

Smith N.H., Gordon S.V., de la Rúa-Domenech R., Clifton-Hadley R.S., Hewinson, R.G. 2006. Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. *Nat. Rev. Microbiol.* 4: 670-681.

Smith N.H., Kremer K., Inwald J., Dale J., Driscoll J.R., et al. 2006. Ecotypes of the

Mycobacterium tuberculosis complex. J. Theo. Biol. 239(2): 220-225.

Stamatakis A. 2014. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30(9): 1312 – 1313.

Stucki D. y Ganeux S. 2013. Single nucleotide polymorphisms in *Mycobacterium tuberculosis* and the need for a curated database. Tuberc. 93(1): 30-39.

Tettelin H., Riley D., Cattuto C., Medini D. 2008. Comparative genomics: the bacterial pan-genome. Curr. Opin. Microbiol. 11: 472–477.

Thacker T.C., Palmer M.V., Robber-Austerman S., Stuber T.P., Waters W.R. 2015. Anatomical distribution of *Mycobacterium bovis* genotypes in experimentally infected white-tailed deer. Vet. Microbiol. 180: 75-81.

Thorvaldsdottir H., Robinson J.T., Mesirov J.P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14: 178–192.

Tsao K., Robbe-Austerman S., Miller R.S., Portacci K., Grear D.A., et al. 2014. Sources of bovine tuberculosis in the United States. Infect. Genet. Evol. 28: 137-143.

Van der Auwera G.A., Carneiro M.O., Hartl C., Poplin R., del Angel G., et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinform. 11:11.10.1–11.10.33.

Wang T. y Chen F. 2013. The evolutionary landscape of the *Mycobacterium tuberculosis* genome. Gene. 518(1): 187-193.

Wright D.M., Allen A.R., Mallon T.R., McDowell S.W.J., Bishop S.C., et al. 2013. Field-isolated genotypes of *Mycobacterium bovis* vary in virulence and influence case pathology but do not affect outbreak size. PLoS ONE. 8(9): e74503.

Zenteno-Cuevas R., Mendoza-Damián F., Cansino-Muñoz I., Enciso-Moreno L., Pérez-Navarro L. M., et al. 2014. Description of the population structure and genetic diversity of tuberculosis in Estado de Mexico, a low prevalence setting from Mexico. APMIS. 123: 116-122.

Zerbino D.R. y Birney E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18(5): 821 – 829.